DAGSTUHL
REPORTS

Volume 12, Issue 7, July 2022

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at https://www.dagstuhl.de/dagpub/2192-5283

Publication date
February, 2023

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `https://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.12.7.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
https://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 22271

Algorithms for Participatory Democracy

Markus Brill[*1], **Jiehua Chen**[*2], **Andreas Darmann**[*3], **David Pennock**[*4], **and Matthias Greger**[†5]

1 **TU Berlin, DE.** brill@tu-berlin.de
2 **TU Wien, AT.** jiehua.chen@tuwien.ac.at
3 **Universität Graz, AT.** andreas.darmann@uni-graz.at
4 **Rutgers University – Piscataway, US.** dpennock17@gmail.com
5 **TU München, DE.** matthias.greger@tum.de

Abstract

Participatory democracy aims to make democratic processes more engaging and responsive by giving all citizens the opportunity to participate, and express their preferences, at many stages of decision-making processes beyond electing representatives. Recent years have witnessed an increasing interest in participatory democracy systems, enabled by modern information and communication technology. Participation at scale gives rise to a number of algorithmic challenges. In this seminar, we addressed these challenges by bringing together experts from *computational social choice* (COMSOC) and related fields. In particular, we studied algorithms for online decision-making platforms and for participatory budgeting processes. We also explored how innovations such as prediction markets, liquid democracy, quadratic voting, and blockchain can be employed to improve participatory decision-making systems.

Seminar July 3–8, 2022 – http://www.dagstuhl.de/22271
2012 ACM Subject Classification Applied computing → Law, social and behavioral sciences; Theory of computation → Algorithmic game theory and mechanism design
Keywords and phrases liquid democracy, participatory budgeting, social choice and currency, platforms for collective decision making
Digital Object Identifier 10.4230/DagRep.12.7.1

1 Executive Summary

Markus Brill (TU Berlin, DE)
Jiehua Chen (TU Wien, AT)
Andreas Darmann (Universität Graz, AT)
David Pennock (Rutgers University – Piscataway, US)

Participatory democracy aims at a broad and direct participation of citizens in policy decision making, enabling a large fraction of citizens to propose ideas, debate issues, and vote on decisions. Modern-day participatory democracy processes entail several kinds of *algorithmic* challenges. This seminar focused on the algorithms underlying three types of participatory democracy systems: (1) online decision-making platforms for governments and organizations (such as *LiquidFeedback* or *decidim*), (2) participatory budgeting processes that enable citizens to directly and collectively decide how to spend tax dollars, and (3) collective decision-making systems involving currency. We also had dedicated sessions discussing algorithmic challenges

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Algorithms for Participatory Democracy, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 1–18
Editors: Markus Brill, Jiehua Chen, Andreas Darmann, and David Pennock
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

related to *liquid democracy* and the relation between participatory democracy and blockchain technology. Working groups have been initiated discussing partial participation in voting mechanisms, the use of currency in social choice problems, participatory budgeting, and the impact of computational social choice.

The technical program was complemented by a demo session in which Jobst Heitzig demonstrated *vodle* (`http://www.vodle.it`) and Daniel Reeves demonstrated Beeminder (`https://www.beeminder.com/`) and other decision-making tools. Moreover, we organized a panel discussion on *The Past, Present, and Future of Computational Social Choice*, moderated by Piotr Faliszewski. In this panel discussion, Haris Aziz, Edith Elkind, Jérôme Lang, and Bill Zwicker gave their perspectives on the development of the field of COMSOC.

The organizers thank all participants for their interesting ideas and viewpoints presented in talks, discussions, and informal meetings. Moreover, we would like to express our gratitude towards Schloss Dagstuhl and its staff for all the support before and during the seminar, which contributed to making this seminar a successful one.

2 Table of Contents

3 Overview of Talks

3.1 Flexible Representative Democracy – An Introduction with Binary Issues

Ben Abramowitz (Tulane University – New Orleans, US), Nicholas Mattei (Tulane University – New Orleans, US)

Joint work of Ben Abramowitz, Nicholas Mattei
Main reference Ben Abramowitz, Nicholas Mattei: "Flexible Representative Democracy: An Introduction with Binary Issues", in Proc. of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, pp. 3–10, ijcai.org, 2019.
URL https://doi.org/10.24963/ijcai.2019/1

We introduce Flexible Representative Democracy (FRD), a novel hybrid of Representative Democracy (RD) and Direct Democracy (DD) in which voters can alter the issue-dependent weights of a set of elected representatives. FRD allows the voters to actively determine the degree to which the system is direct versus representative. We introduce and analyze FRD in the setting where issues are binary and symmeric and compare the outcomes of various voting systems using Direct Democracy with majority voting and full participation as an ideal baseline. First, we demonstrate the shortcomings of Representative Democracy in our model. We provide NP-Hardness results for electing an ideal set of representatives, discuss pathologies, and demonstrate empirically that common multi-winner election rules for selecting representatives do not perform well in expectation. To analyze the effects of adding flexibility, we begin by providing theoretical results on how issue-specific delegations determine outcomes. Finally, we provide empirical results comparing the outcomes of Representative Democracy, proxy voting with fixed sets of proxies across issues, and Flexible Representative Democracy with issue-specific delegations. Our results show that variants of Proxy Voting yield no discernible benefit over unweighted representatives and reveal the potential for Flexible Representative Democracy to improve outcomes as voter participation increases.

3.2 Proportionally Representative Participatory Budgeting with Ordinal Preferences

Haris Aziz (UNSW – Sydney, AU)

Joint work of Haris Aziz, Barton E. Lee
Main reference Haris Aziz, Barton E. Lee: "Proportionally Representative Participatory Budgeting with Ordinal Preferences", in Proc. of the Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, pp. 5110–5118, AAAI Press, 2021.
URL https://ojs.aaai.org/index.php/AAAI/article/view/16646

Participatory budgeting (PB) is a democratic paradigm whereby voters decide on a set of projects to fund with a limited budget. We consider PB in a setting where voters report ordinal preferences over projects and have (possibly) asymmetric weights. We propose proportional representation axioms and clarify how they fit into other preference aggregation settings, such as multi-winner voting and approval-based multi-winner voting. As a result of our study,

we also discover a new solution concept for approval-based multi-winner voting, which we call Inclusion PSC (IPSC). IPSC is stronger than proportional justified representation (PJR), incomparable to extended justified representation (EJR), and yet compatible with EJR. The well-studied Proportional Approval Voting (PAV) rule produces a committee that satisfies both EJR and IPSC; however, both these axioms can also be satisfied by an algorithm that runs in polynomial-time.

3.3 Comparing input formats for participatory budgeting

Gerdus Benadè (Boston University, US)

Joint work of Gerdus Benadè, Ariel Procaccia, Nisarg Shah, Swaprava Nath, Kobi Gal, Roy Fairstein
Main reference Gerdus Benadè, Swaprava Nath, Ariel D. Procaccia, Nisarg Shah: "Preference Elicitation for Participatory Budgeting", Manag. Sci., Vol. 67(5), pp. 2813–2827, 2021.
URL https://doi.org/10.1287/mnsc.2020.3666

We ask how a voter in a participatory budgeting process should be prompted to express their preferences. When assuming additive utilities, theoretical analysis finds clear separation in the worst case loss of common input formats, compared to the optimal social welfare under complete information. We complement this with user experiments that compare input formats based on the efficiency of their outcomes as well as how expressive and easy to use voters find the format.

3.4 Markets for Aggregation

Rupert Freeman (University of Virginia, US)

Joint work of Rupert Freeman, David Pennock, Dominik Peters, Jennifer W. Vaughan
Main reference Rupert Freeman, David M. Pennock, Dominik Peters, Jennifer Wortman Vaughan: "Truthful aggregation of budget proposals", J. Econ. Theory, Vol. 193, p. 105234, 2021.
URL https://doi.org/10.1016/j.jet.2021.105234

By the characterization of Moulin (1980), all anonymous and strategyproof voting rules on single-peaked domains take the form of generalized median mechanisms. Recent work has identified one particular generalized median mechanism known as the uniform phantom mechanism. In this talk I describe the uniform phantom mechanism and its properties, and propose possible generalizations and extensions.

3.5 Social Choice Around the Block: On the Computational Social Choice of Blockchain

Davide Grossi (University of Groningen, NL)

Main reference Davide Grossi: "Social Choice Around the Block: On the Computational Social Choice of Blockchain", in Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022, pp. 1788–1793, International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2022.
URL https://dl.acm.org/doi/10.5555/3535850.3536111

One of the most innovative aspects of blockchain technology consists in the introduction of an incentive layer to regulate the behavior of distributed protocols. The designer of a blockchain system faces therefore issues that are akin to those relevant for the design of economic mechanisms, and faces them in a computational setting. In this talk I argue for the importance of computational social choice in blockchain research and I identify a few challenges at the interface of the two fields that illustrate the strong potential for cross-fertilization between them.

3.6 Maximum Partial Consensus – a probabilistic, nonmajoritarian, single-winner voting method aiming at fairness and efficiency

Jobst Heitzig (Potsdam-Institut für Klimafolgenforschung (PIK), DE)

Joint work of Jobst Heitzig, Forest W. Simmons, Sara Constantino
Main reference Jobst Heitzig, Forest W. Simmons, Sara Constantino: "Fair Group Decisions via Non-deterministic Proportional Consensus". Available at SSRN, 2022.
URL http://dx.doi.org/10.2139/ssrn.3751225

Are there group decision methods which: (i) give everyone, including minorities, an equal share of effective power even when voters act strategically, (ii) promote consensus and equality, rather than polarization and inequality, and (iii) do not favour the status quo or rely too much on chance?

We describe two non-deterministic group decision methods that meet these criteria, one based on automatic bargaining over lotteries, the other on conditional commitments to approve compromise options.

Through theoretical analysis, agent-based simulations and a behavioral experiment, we show that these methods prevent majorities from consistently suppressing minorities, as with deterministic methods, and proponents of the status quo from blocking decisions as in other consensus-based approaches. These methods achieve aggregate welfare comparable to that of common methods, while employing chance judiciously.

In an experiment with naive participants, we find that a sizable fraction prefers to use a non-deterministic method over familiar Plurality Voting to allocate resources, though this depends on participants' position within the group.

Overall, we show that the welfare costs of fairness and consensus are small compared to the inequality costs of majoritarianism.

3.7 Quadratic Voting: An Overview

Anson Kahng (University of Rochester, US)

In this talk, I give an overview of quadratic voting (also called plural voting), which is a system of voting in which voters may spend real or artificial currency in order to purchase votes, where x votes costs x^2 currency units. In the binary, discrete case, the option with a majority of votes wins the election. I cover the intuition behind why quadratic voting leads to near-perfect efficiency in theory, assuming voters have quasi-linear utilities for saving credits. I also discuss implementations of quadratic voting in the real world, notably in the Colorado state legislature, as well as uses of quadratic voting for information elicitation (notably surveys, where it is used as an alternative to the Likert scale). Lastly, I discuss criticisms of quadratic voting and open theoretical and practical problems in the area.

3.8 Fairness in Long-Term Participatory Budgeting

Martin Lackner (TU Wien, AT)

Joint work of Martin Lackner, Jan Maly, Simon Rey
Main reference Martin Lackner, Jan Maly, Simon Rey: "Fairness in Long-Term Participatory Budgeting", in Proc. of the AAMAS '21: 20th International Conference on Autonomous Agents and Multiagent Systems, Virtual Event, United Kingdom, May 3-7, 2021, pp. 1566–1568, ACM, 2021.
URL https://dl.acm.org/doi/10.5555/3463952.3464161

Participatory Budgeting (PB) processes are often designed to span several years, with referenda for new budget allocations taking place regularly. In this talk, I will discuss a formal framework for long-term PB, based on a sequence of budgeting problems as main input. I introduce a theory of fairness for this setting, focusing on three main concepts that apply to types (groups) of voters: (i) achieving equal welfare for all types, (ii) minimizing inequality of welfare (as measured by the Gini coefficient), and (iii) achieving equal welfare in the long run. All three fairness criteria cannot be guaranteed in a single round of PB and thus necessitate a long-term perspective.

3.9 Upgrading Liquid Democracy: Multiagent Delegations and Interconnected Issues

Arianna Novaro (Université Paris I, FR) and Umberto Grandi (University Toulouse Capitole, FR)

Joint work of Rachael Colley, Umberto Grandi, Arianna Novaro
Main reference Rachael Colley, Umberto Grandi, Arianna Novaro: "Unravelling multi-agent ranked delegations", Auton. Agents Multi Agent Syst., Vol. 36(1), p. 9, 2022.
URL https://doi.org/10.1007/s10458-021-09538-2

In this talk, we have first given an overview of the (relatively recent) framework of liquid/delegative democracy. Then, in the first part of the talk Arianna presented the framework of multiagent ranked delegations (what we call "smart voting"), which generalizes standard

liquid democracy in two ways: (1) the agents can express more complex delegation functions to multiple other agents; (2) the agents can express a ranking over preferred delegations. We proposed four greedy unravelling procedures, and two optimal procedures, that transform the delegation profiles into "classical" voting profiles, and we studied both computational and axiomatic properties for them. Then, in the second part of the talk, Umberto focused on the case where multiple interconnected issues are to be decided upon, and the agents can delegate parts of the decision (i.e., the decision on a subset of the issues) to other agents. To this end, he presented some procedures to be used to restore or guarantee that the final ballots respect the underlying constraints. Finally, we discussed some open problems (e.g., a more refined axiomatic analysis, and the existing tension between anonymity and transparency in delegations), as well as some recent computational and experimental work that has been done for the smart voting model.

3.10 Social Choice with Currency: A Survey

David Pennock (Rutgers University – Piscataway, US)

This talk surveyed uses of currency for group decision making. Currency is useful for both preference aggregation and belief aggregation, the two main components of group decision making. Both virtual and real currency can be used for preference aggregation, including trading votes, storable votes, quadratic voting, and decision auctions. Virtual and real currency is also used for belief aggregation in the form of scoring rules, prediction markets, and wagering mechanisms.

3.11 Participatory Budgeting: A Survey

Dominik Peters (University Paris-Dauphine, FR)

URL https://dominik-peters.de/slides/dagstuhl-pb-survey.pdf

Using participatory budgeting (PB), cities give their residents an opportunity to influence how the city's budget is spent. This is often done by collecting project proposals, and then voting on those proposals. I give a survey on work in computational social choice on this voting problem. First, I give an overview of voting systems in use in major cities, which mostly involve a naive greedy algorithm based on approval scores. Then I formalize a model of voting under a knapsack constraint with additive valuations, and discuss how standard approval votes fit in that model. Then I discuss the goal of proportional representation formalized using axioms such as Extended Justified Representation, and introduce the Method of Equal Shares. I mention PB work on the core, strategyproofness, computational complexity, and several extensions which could form directions for future research: allowing negative votes, allowing for constraints, as well as analyzing input formats and the process of agenda setting, among others.

3.12 Condorcet Solutions in Frugal Models of Budget Allocation

Clemens Puppe (KIT – Karlsruher Institut für Technologie, DE)

Joint work of Klaus Nehring, Clemens Puppe
URL https://econpapers.wiwi.kit.edu/downloads/WP_156.pdf

We study a voting model with incomplete information in which the evaluation of social welfare must be based on information about agents' top choices plus general qualitative background conditions on preferences. The former is elicited individually, while the latter is not. We apply this "frugal aggregation" model to multi-dimensional budget allocation problems, relying on the specific assumptions of convexity and separability of preferences. We propose a solution concept of ex-ante Condorcet winners which flexibly incorporates the epistemic assumptions of particular frugal aggregation models. We show that for the case of convex preferences, the ex-ante Condorcet approach naturally leads to a refinement of the Tukey median. By contrast, in the case of separably convex preferences, the same approach leads to different solution, the 1-median, i.e. the minimization of the sum of the L1-distances to the agents' tops. An algorithmic characterization renders the latter solution analytically tractable and efficiently computable.

3.13 Homo Economicus Wannabees

Daniel Reeves (Beeminder – Portland, US)

Joint work of Daniel Reeves, Bethany Soule
URL https://www.beeminder.com

What if you were so enamored with formalizing social choice problems that you made all group decisions that way and even raised your kids to do so? I describe in this talk what happens. In particular, I describe the half dozen or so auction- and prediction-market-based decision mechanisms we use in our family. Additionally I describe new work at Beeminder implementing group commitment mechanisms.

3.14 Shortlisting Rules and Incentives in an End-to-End Model for Participatory Budgeting

Simon Rey (University of Amsterdam, NL)

Joint work of Simon Rey, Ulle Endriss, Ronald de Haan
Main reference Simon Rey, Ulle Endriss, Ronald de Haan: "Shortlisting Rules and Incentives in an End-to-End Model for Participatory Budgeting", in Proc. of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021, pp. 370–376, ijcai.org, 2021.
URL https://doi.org/10.24963/ijcai.2021/52

In this talk I introduced an end-to-end model for participatory budgeting grounded in social choice theory. Our model accounts for the interplay between the two stages commonly encountered in real-life participatory budgeting. In the first stage participants propose

projects to be shortlisted, while in the second stage they vote on which of the shortlisted projects should be funded. Prior work of a formal nature has focused on analysing the second stage only. We introduce several shortlisting rules for the first stage and analyse them in both normative and algorithmic terms. Our main focus is on the incentives of participants to engage in strategic behaviour during the first stage, in which they need to reason about how their proposals will impact the range of strategies available to everyone in the second stage. On top of the technical presentation, this talk was also a call for more realistic models for participatory budgeting, in an attempt to capture processes as they occur in real-life.

3.15 Liquid Democracy with Ranked Delegations

Ulrike Schmidt-Kraepelin (TU Berlin, DE)

Joint work of Markus Brill, Théo Delemazure, Anne-Marie George, Martin Lackner, Ulrike Schmidt-Kraepelin
Main reference Markus Brill, Théo Delemazure, Anne-Marie George, Martin Lackner, Ulrike Schmidt-Kraepelin: "Liquid Democracy with Ranked Delegations", in Proc. of the Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 – March 1, 2022, pp. 4884–4891, AAAI Press, 2022.
URL https://ojs.aaai.org/index.php/AAAI/article/view/20417

Liquid democracy is a novel paradigm for collective decision-making that gives agents the choice between casting a direct vote or delegating their vote to another agent. We consider a generalization of the standard liquid democracy setting by allowing agents to specify multiple potential delegates, together with a preference ranking among them. This generalization increases the number of possible delegation paths and enables higher participation rates because fewer votes are lost due to delegation cycles or abstaining agents. In order to implement this generalization of liquid democracy, we need to find a principled way of choosing between multiple delegation paths. We provide a thorough axiomatic analysis of the space of delegation rules, i.e., functions assigning a feasible delegation path to each delegating agent. In particular, we prove axiomatic characterizations as well as an impossibility result for delegation rules. We also analyze requirements on delegation rules that have been suggested by practitioners, and introduce novel rules with attractive properties. By performing an extensive experimental analysis on synthetic as well as real-world data, we compare delegation rules with respect to several quantitative criteria relating to the chosen paths and the resulting distribution of voting power. Our experiments reveal that delegation rules can be aligned on a spectrum reflecting an inherent trade-off between competing objectives.

3.16 Cordial Miners: The Tip of the Blocklace Consensus Protocol Stack

Ehud Shapiro (Weizmann Institute – Rehovot, IL)

Joint work of Idit Keidar, Oded Naor, and Ehud Shapiro
Main reference Idit Keidar, Oded Naor, Ehud Shapiro: "Cordial Miners: Blocklace-Based Ordering Consensus Protocols for Every Eventuality", arXiv, 2022.
URL https://doi.org/10.48550/ARXIV.2205.09174

Cordial Miners are a family of permissioned "blockchain consensus" protocols, with optimal instances for asynchrony and eventual synchrony. Their efficiency – almost half the latency of state-of-the-art DAG-based protocols – stems from their not using reliable broadcast as a building block. Rather, Cordial Miners use the blocklace – a partially-ordered generalization of the totally-ordered blockchain – for all algorithmic tasks required for ordering consensus: Dissemination, equivocation-exclusion, and ordering. We present the protocols within the broader context of the blocklace consensus protocol stack, offered as an alternative architectural foundation for the digital realm that is grassroots and egalitarian.

3.17 PB++

Nimrod Talmon (Ben Gurion University – Beer Sheva, IL)

Joint work of Nimrod Talmon, Krzysztof Sornat, Pallavi Jain, Meirav Zehavi, Martin Koutecký, Matthias Köppe
Main reference Matthias Köppe, Martin Koutecký, Krzysztof Sornat, Nimrod Talmon. *Fine-Grained Liquid Democracy for Cumulative Ballots*. arXiv:2208.14441, 2022.
URL https://arxiv.org/abs/2208.14441

The standard setting of participatory budgeting does not treat several aspects of the problem that are relevant for certain cases and can improve the usability of such processes. I will speak about algorithms for several generalizations of participatory budgeting, in particular: a setting with project interactions in which projects are not independent; a setting with project groups in which projects have (perhaps intersecting) types; and on the possibility of fine-grained liquid democracy for participatory budgeting.

3.18 Incentive-Compatible Forecasting Competitions

Jens Witkowski (Frankfurt School of Finance & Management, DE)

Joint work of Jens Witkowski, Rupert Freeman, Jennifer W. Vaughan, David Pennock, Andreas Krause
Main reference Jens Witkowski, Rupert Freeman, Jennifer Wortman Vaughan, David M. Pennock, Andreas Krause: "Incentive-Compatible Forecasting Competitions", Management Science, 2022.
URL https://doi.org/10.1287/mnsc.2022.4410

We initiate the study of incentive-compatible forecasting competitions in which multiple forecasters make predictions about one or more events and compete for a single prize. We have two objectives: (1) to incentivize forecasters to report truthfully and (2) to award the prize to the most accurate forecaster. Proper scoring rules incentivize truthful reporting if all

forecasters are paid according to their scores. However, incentives become distorted if only the best-scoring forecaster wins a prize, since forecasters can often increase their probability of having the highest score by reporting more extreme beliefs. In this paper, we introduce two novel forecasting competition mechanisms. Our first mechanism is incentive compatible and guaranteed to select the most accurate forecaster with probability higher than any other forecaster. Moreover, we show that in the standard single-event, two-forecaster setting and under mild technical conditions, no other incentive compatible mechanism selects the most accurate forecaster with higher probability. Our second mechanism is incentive compatible when forecasters' beliefs are such that information about one event does not lead to belief updates on other events, and it selects the best forecaster with probability approaching 1 as the number of events grows. Our notion of incentive compatibility is more general than previous definitions of dominant strategy incentive compatibility in that it allows for reports to be correlated with the event outcomes. Moreover, our mechanisms are easy to implement and can be generalized to the related problems of outputting a ranking over forecasters and hiring a forecaster with high accuracy on future events.

4 Working groups

4.1 Partial participation in participatory budgeting

Reshef Meir (Technion – Haifa, IL), Paul Gölz (Carnegie Mellon University – Pittsburgh, US), Umberto Grandi (University Toulouse Capitole, FR), Christian Klamler (Universität Graz, AT), Sonja Kraiczy (University of Oxford, GB), Stefan Napel (Universität Bayreuth, DE), and Sofia Simola (TU Wien, AT)

Much of the research around participatory budgeting (both theory and practice) makes an implicit assumption that only the preferences and votes of the active voters matter. Based on this assumption, there are many "guarantees" such as axiomatic properties, welfare and fairness bounds.

However in reality only a small fraction of the population actively votes, while everyone is affected by the outcome. This means that a crucial part of studying PB is about understanding and mitigating the effect of partial participation.

In the group discussion we briefly discussed two design ideas, and started to develop a model to evaluate the effect of partial participation.

The model itself involves a measure based on welfare loss, similar to the "price of anarchy" common in game theory literature. The difference is that we compare partial participation to full participation, rather than equilibrium to optimal behavior. Without further assumptions, the "cost of partial participation" may be quite high, but more nuanced results can be obtained under assumptions on the "distance" between the full and partial vote distribution.

The two suggestions we started to discuss relate to increasing the incentive to participate: Instead of using the full votes, sample a small set of voters to make the decisions. While this seems counter-productive, note that it makes every voter more pivotal and thus makes participation more attractive. Allowing the budget (which is usually assumed to be fixed) to depend on the participation rate in each region, thereby creating further reason to participate. One way to implement this indirectly is to set a minimum quorum for projects.

4.2 Impact of COMSOC

Arianna Novaro (Université Paris I, FR), Robert Bredereck (TU Clausthal, DE), Andreas Darmann (Universität Graz, AT), Théo Delemazure (University Paris-Dauphine, FR), Jobst Heitzig (Potsdam-Institut für Klimafolgenforschung (PIK), DE), Ayumi Igarashi (National Institute of Informatics – Tokyo, JP), Jérôme Lang (CNRS – Paris, FR), and William S. Zwicker (Istanbul Bilgi University, TR)

This working group focused on how to make the general public more aware of existing methods and platforms for collective decision-making, in order to increase the impact of COMSOC. The discussion was triggered by an app that Ayumi developed for fair division of chores in couples, which was promoted on the Japanese National TV. Indeed, some specific apps, such as Spliddit, have been quite successful (for instance, in helping with rent division), but this success has not necessarily spread to other COMSOC apps: our goal was to investigate why and to find possible avenues to change this.

Concerning the why, we first asked ourselves what was the main motivation for people to use a popular platform like Doodle: is it to make a decision for them, or just to more easily collect the information? Is it important for the users to just feel like their voice is being heard, or do they also want the experience of using the app to be "fun"? Which kind of obstacles do they perceive with existing COMSOC tools? (The latter could be an interesting question to investigate experimentally).

After some discussion, the main proposal that emerged was to create a general tool, such as a chatbot or an interactive "meta-website", which would either be integrated to existing platforms (like on Slack or Facebook) or be independent, to guide and help the general public towards other COMSOC tools tailored to their specific problem at hand. Then, via some small-scale case studies, we could test the effectiveness of such a tool, before advertising it more broadly. A natural location for the case study would be an university, where many instances of collective decision-making arise, involving different types of agents (e.g., rent or chore division among students, creation of diverse committees, choice of journal subscriptions), and where we, as researchers, have more opportunity to have an impact.

In terms of concrete support in the development and diffusion of such a tool, we discussed the possibility for Dagstuhl itself (i.e., the Leibniz-Zentrum für Informatik) to provide help, either directly or by pointing to the right partners; or to investigate grant opportunities. To promote the tool, some options we thought of would be to write in the local university/student newspaper, and to advertise it on classical and social media (possibly via Youtube videos).

4.3 Social choice and currency

David Pennock (Rutgers University – Piscataway, US), Ben Abramowitz (Tulane University – New Orleans, US), Robert Bredereck (TU Clausthal, DE), Markus Brill (TU Berlin, DE), Rupert Freeman (University of Virginia, US), Davide Grossi (University of Groningen, NL), Anson Kahng (University of Rochester, US), Nicholas Mattei (Tulane University – New Orleans, US), Reshef Meir (Technion – Haifa, IL), Marcus Pivato (University of Cergy-Pontoise, FR), Daniel Reeves (Beeminder – Portland, US), Ehud Shapiro (Weizmann Institute – Rehovot, IL), Nimrod Talmon (Ben Gurion University – Beer Sheva, IL), and Jens Witkowski (Frankfurt School of Finance & Management, DE)

This working group explored the use of currency to improve mechanisms for collective decision making.

The group discussed a number of wide-ranging ideas, including voting using auctions, plutocracy, getting away from quasi-linear preferences, Switzerland as an unheralded but good model of democracy, sovereign coins as a grass-roots economy on blockchain, peer evaluation, peer prediction, price of anarchy for quadratic voting, use of currency to improve on envy-freeness up to one vote, a theory of investing in opposite-party candidates, allowing people to privately fund participatory-budgeting projects, allocating the public good of grant funds using lightweight methods, and whether mega-concentration of capital are important for innovation. Among these ideas, the group produced four main concrete outcomes:

First, the group discussed how to address income inequality and unfairness when real currency is used to vote. Redistribution must be part of the model. The right way to model currency in voting is in the context of the optimal taxation literature in economics. A first step would be to add quadratic voting (QV) with redistribution into the model and recompute optimal taxation. We conjecture that optimal taxes will be less due to the redistributive nature of QV.

Second, one member proposed a problem for blocklace. It's a cooperative blockchain mining procedure called DAG-writer, instead of a competitive mining procedure, so it's much more efficient. A good analogy is a group of runners running around a track who need to: circle the track as quickly as possible, yet stay together as a group. Another member proposed a possible solution with a preliminary analysis that it might work.

Third, a subgroup explored how to understand Uniswap, a common automated market maker (AMM) protocol responsible for trillions of dollars of transactions in cryptocurrencies, as a prediction market AMM. (This is the AMM that Manifold Markets uses.) There is a large literature on prediction market AMMs. A natural question is whether Uniswap is equivalent to a known AMM. This subgroup made some progress in rewriting the Uniswap price function as a cost function in the AMM literature. The Uniswap price function has some advantages for combinatorial prediction markets that no other known price function has.

Fourth, the group considered a liquid democracy model where voting records are public and voters delegate (probabilistically) to more historically accurate voters on binary issues. Society enacts a weighted average of the votes on each issue. Does society act as a no-regret learning algorithm? That is, are the decisions of society competitive with the single best voter?

4.4 What should we focus on when considering participatory budgeting?

Simon Rey (University of Amsterdam, NL), Haris Aziz (UNSW – Sydney, AU), Dorothea Baumeister (Heinrich-Heine-Universität Düsseldorf, DE), Gerdus Benadè (Boston University, US), Edith Elkind (University of Oxford, GB), Piotr Faliszewski (AGH University of Science & Technology – Krakow, PL), Matthias Greger (TU München, DE), Martin Lackner (TU Wien, AT), and Dominik Peters (University Paris-Dauphine, FR)

The idea of this working group was to have a general discussion about participatory budgeting (PB). It started from some considerations on how fairness should be defined in the context of PB. It evolved from there into a more general exchange about the more fundamental focus of PB. In the present report, we elaborate on the different points that have been discussed.

The standard model used to describe PB elections can actually be seen as the more general task of collective selection of costly alternatives. With that in mind, several real-life scenarios can be captured through the study of "PB", for which the goals will be quite different. Let us present three typical ones.

- The usual PB election: A set of voters express their preferences over a set of costly alternatives in order to decide which ones to implement, under some budget constraint. Voter turnout is a crucial issue here (at least for the decision-makers). In that view, trying to fairly represent the diversity of voters is an important goal.
- Grant selection: A committee (the voters, actually) has to decide on behalf of a funding agency which research proposals to fund subject to a budget limit. As opposed to the above, voter participation is definitely not an issue. Similarly, a property such as exhaustiveness[1] – considered highly important in PB elections – needs not be relevant here.
- Selecting the catering options for an event: A group of organizers has to decide on the catering options offered for an event. The typical goal here would be to find cheap solutions that cover a large number of guests. In that sense, exhaustiveness is probably not desirable, but proportionality-like requirements might be.

These three examples already display a wide variety of scenarios. In the following we will only focus on PB elections, but the other examples also deserve to be studied.

One of the most fundamental questions we touched upon is that of the actual goal of a PB election. Two general approaches emerged from our discussion.

The first one is to consider PB processes as a way to inform and help decision makers with selecting which projects to fund. The idea is that PB processes help reveal the societal value of the alternatives that can then be used to make an informed decision. In this view, concepts relating to proportionality enforce selecting high societal value projects (cohesive groups can be seen as indication of high societal value). This approach also calls for an epistemic analysis where there exists a ground truth societal value for the projects, that we are trying to approximate.

Another goal for PB is that of educating citizens to the democratic process. One motivation for this is to consider PB as a way to attract citizens to democratic instances (if you engage in the PB process, then you might also engage in other democratic activities).

[1] Exhaustiveness states that it should not be possible to still fund an extra project with the leftover budget.

Another aspect here could be to force citizens to think about budget constraints in general, to give them a taste of decision-making with scarce resources. This approach seems harder to account for within the social choice literature.

Given all the above, it is not surprising that fairness requirements such as proportionality are the ones that received the most attention in the current social choice literature about PB. However, it is still unclear what exactly fairness should be about. A large part of our discussion focused on this issue.

The current literature focuses mainly on satisfaction-based fairness. It can be argued that satisfaction cannot really be captured, especially when using approval ballots, and that we would thus need different kinds of fairness requirements. Several directions are opened here. One such idea could be to study equity of resources in PB, via the concept of share, see for instance [2]. A requirement that already exists in the literature and that could be explored further is IPSC [1], which is not a satisfaction-based criterion. Finally, another approach that is worth pursuing is that of market-based fairness, which has been operationalized via the idea of priceability, see for instance [3].

The objections about satisfaction-based fairness we made above are all within the context of approval voting. Another point of discussion was to challenge the use of approval ballots. Indeed, it could be that they are just too simplistic to allow for convincing fairness definitions. In that regard, exploring other types of ballots would be interesting. On the one hand, we could investigate more complex ballots to get closer to the voters' preferences. On the other hand, we could also look into very simple ballots with clear semantics (the semantics of approval ballots is ambiguous), such as 1-approval ballots: voters are only allowed to approve of one alternative.

Studying 1-approval ballots could teach us about the fundamentals of fairness in PB. For the same reason, it would be worth studying the other end of the spectrum: weak orders over feasible subsets of alternatives. This is, of course, not a practical model, but developing a study of fairness with perfect information about the voters' preferences could lead to a deeper understanding of what is happening in PB.

Reaching a deeper understanding of fairness in PB could also lead to new appealing mechanisms. At the moment, the method of equal shares [3] stands out as *the* uncontested mechanism to go for (at least when it comes to fairness). This fact is somehow surprising – for multi-winner voting, we know of several mechanisms providing strong fairness guarantees – and probably indicates that there is more to look for here.

References

1 Haris Aziz, Barton E. Lee. *Proportionally representative participatory budgeting with ordinal preferences*. Proceedings of the 35th AAAI Conference on Artificial Intelligence, 5110–5118, 2021.

2 Jan Maly, Simon Rey, Ulle Endriss, Martin Lackner. *Effort-based fairness for participatory budgeting*. arXiv preprint arXiv:2205.07517, 2022.

3 Dominik Peters, Grzegorz Pierczynski, Piotr Skowron. *Proportional participatory budgeting with additive utilities*. Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS), 12726–12737, 2021.

Participants

- Ben Abramowitz
Tulane University – New Orleans, US
- Haris Aziz
UNSW – Sydney, AU
- Dorothea Baumeister
Heinrich-Heine-Universität Düsseldorf, DE
- Gerdus Benadè
Boston University, US
- Robert Bredereck
TU Clausthal, DE
- Markus Brill
TU Berlin, DE
- Alfonso Cevallos
Web3 – Zug, CH
- Andreas Darmann
Universität Graz, AT
- Théo Delemazure
University Paris-Dauphine, FR
- Edith Elkind
University of Oxford, GB
- Piotr Faliszewski
AGH University of Science & Technology – Krakow, PL
- Rupert Freeman
University of Virginia, US
- Ashish Goel
Stanford University, US
- Paul Gölz
Carnegie Mellon University – Pittsburgh, US
- Umberto Grandi
University Toulouse Capitole, FR
- Matthias Greger
TU München, DE
- Davide Grossi
University of Groningen, NL
- Jobst Heitzig
Potsdam-Institut für Klimafolgenforschung (PIK), DE
- Ayumi Igarashi
National Institute of Informatics – Tokyo, JP
- Anson Kahng
University of Rochester, US
- Christian Klamler
Universität Graz, AT
- Sonja Kraiczy
University of Oxford, GB
- Martin Lackner
TU Wien, AT
- Jérôme Lang
CNRS – Paris, FR
- Nicholas Mattei
Tulane University – New Orleans, US
- Reshef Meir
Technion – Haifa, IL
- Stefan Napel
Universität Bayreuth, DE
- Arianna Novaro
Université Paris I, FR
- David Pennock
Rutgers University – Piscataway, US
- Dominik Peters
University Paris-Dauphine, FR
- Marcus Pivato
University of Cergy-Pontoise, FR
- Clemens Puppe
KIT – Karlsruher Institut für Technologie, DE
- Daniel Reeves
Beeminder – Portland, US
- Simon Rey
University of Amsterdam, NL
- Ulrike Schmidt-Kraepelin
TU Berlin, DE
- Ehud Shapiro
Weizmann Institute – Rehovot, IL
- Sofia Simola
TU Wien, AT
- Nimrod Talmon
Ben Gurion University – Beer Sheva, IL
- Jens Witkowski
Frankfurt School of Finance & Management, DE
- William S. Zwicker
Istanbul Bilgi University, TR

Report from Dagstuhl Seminar 22272

Eat-IT: Towards Understanding Interactive Technology and Food

Florian 'Floyd' Mueller[*1], Marianna Obrist[*2], Soh Kim[*3], Masahiko Inami[*4], and Jialin Deng[†5]

1 **Exertion Games Lab, Monash University – Clayton, AU.** floyd@exertiongameslab.org
2 **Department of Computer Science, University College London, UK.** m.obrist@ucl.ac.uk
3 **Civil and Environmental Engineering, Stanford University, US.** sohkim@stanford.edu
4 **Information Somatics Lab, The University of Tokyo, JP.** drinami@star.rcast.u-tokyo.ac.jp
5 **Exertion Games Lab, Monash University – Clayton, AU.** jialin.deng@monash.edu

Abstract

Eating is a basic human need while technology is transforming the way we cook and eat food. For example, see the internet-connected Thermomix cooking appliance, desserts using virtual reality headsets, projection mapping on dinner plates and 3D-printed food in Michelin-star restaurants. Especially within the field of Human-Computer Interaction (HCI), there is a growing interest in understanding the design of technology to support the eating experience. There is a realization that technology can both be instrumentally beneficial (e.g. improving health through better food choices) as well as experientially beneficial (e.g. enriching eating experiences). Computational technology can make a significant contribution here, as it allows to, for example, present digital data through food (drawing from visualization techniques and fabrication advances such as 3D-food printing); facilitate technology-augmented behaviour change to promote healthier eating choices; employ big data across suppliers to help choose more sustainable produce (drawing on IoT kitchen appliances); use machine learning to predictively model eating behaviour; employ mixed-reality to facilitate novel eating experiences; and turn eating into a spectacle through robots that support cooking and serving actions. The aim of this Dagstuhl seminar called "Eat-IT" was to discuss these opportunities and challenges by bringing experts and stakeholders with different backgrounds from academia and industry together to formulate actionable strategies on how interactive food can benefit from computational technology yet not distract from the eating experience itself. With this seminar, we wanted to enable a healthy and inclusive debate on the interwoven future of food and computational technology.

Seminar July 3–8, 2022 – https://www.dagstuhl.de/22272
2012 ACM Subject Classification Human-centered computing → Interaction design
Keywords and phrases Human-Food Interaction, FoodHCI
Digital Object Identifier 10.4230/DagRep.12.7.19

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Eat-IT: Towards Understanding Interactive Technology and Food, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 19–40
Editors: Florian 'Floyd' Mueller, Marianna Obrist, Soh Kim, Masahiko Inami, and Jialin Deng

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Florian 'Floyd' Mueller
Marianna Obrist
Soh Kim
Masahiko Inami

In July 2022, 21 researchers and academics from Europe, Australasia and the USA gathered for a week to discuss the future of the coming together of food and information technology (IT), shortly called eat-IT.

Eating is a basic human need, and there is a growing interest in the field of Human-Computer Interaction (HCI) in designing new interactive food experiences, for example, to promote healthier food practices (e.g., [4, 3, 12, 17]), to make eating a more enjoyable experience (e.g., [18, 19, 21, 10, 6]), and to design multisensory eating experiences (e.g., [13, 15, 16]). Theoretical work around the design of interactive food also emerged, for example, Grimes and Harper [5] proposed that a new view on human-food interactions (HFI) is required and introduced the concept of "celebratory technology" that emphasizes the positive aspects of eating in everyday life. Computational technology can make a significant contribution towards such celebratory technology, for example, Khot et al. [7] presented a system called TastyBeats that instead of presenting physical activity data on a screen, it offers users personalised sports drinks where the quantity and flavour is based on the amount of exercise a user has done. In a similar vein, EdiPulse [8] was introduced as a system that creates activity treats (chocolate creations) using a food printer. The shape and quality of the prints were based on the person's physical activities on that day, allowing for personal and shared reflections through consuming chocolate instead of looking at graphs on a screen. Computer science and in particular the information visualization community can, therefore, regard food (and drinks) as a medium to make data more approachable for people, communicating complex information in an easy-to-digest format [10]. Parametric design approaches have also influenced the way food is produced. For example, Wang et al. [20] developed the concept of shape-changing and programmable food that transforms during the cooking process. Through a material-based interaction design approach, the authors demonstrated the transformation of 2D into 3D food (i.e. pasta). They proposed these transformations for new dining experiences that can surprise users, but this can also be used for outer space, where food comes as a flat design and only transforms into a 3D form through the cooking process.

Furthermore, technological advancements in acoustic levitation have led to the design of taste-delivery technology that transports and manipulates food in mid-air [18], allowing for novel interactions between diners and food that is of interest to HCI researchers as it allows to study augmented food experiences without the use of cutlery. This work further extends taste stimulation towards a multisensory experience of levitating food due to the integration of smell, directional sound, lights, and touch [19]. Furthermore, robots are now in use to serve ice cream to the general public [2]. Lastly, laser-cutters have already been used to embed data into cookies through the engraving of QR codes [14]. Taken together, these examples suggest that computing technology can play a major role in the way we engage with food, in particular, there is a realization that technology can both facilitate instrumental benefits in regards to food (such as improved health through better food choices) as well as experiential benefits (such as enriched social experiences). In summary, computational technology has the potential to influence how people experience eating.

However, this notion of what we call "interactive food" also raises significant concerns. Will computing technology distract from the pure pleasures of eating? Will people accept meals that are optimized through data-driven approaches? Will people enjoy food that is served by robots? Will people understand and act on data that is embedded in food? Questions such as these and, of course, their answers are important for the future of the field, and the seminar tried to investigate them.

The seminar was based on the belief that computer scientists, designers, developers, researchers, chefs, restaurateurs, producers, canteen managers, etc. can learn from each other to positively influence the future of interactive food. Working together allows for the identification of new opportunities the field offers but will also highlight the challenges that the community will need to overcome. In particular, it is still unknown what theory to use to design such computational systems in which the interaction is very multisensorial, contrasting the traditional mouse, keyboard and screen interactions. Furthermore, it is unclear how interacting with food is benefiting from, and also challenged by, our mostly three-times-a-day engagement with it (breakfast, lunch and dinner), again different to our interactions with mobile phones that occur at any time.

Furthermore, how do we create and evaluate interactions with computationally-augmented food that needs preparation time, again very different from our usually immediate interactions with interactive technology? What interaction design theory can guide us in answering these questions in order to extend computer science also to include food interactions? If such theoretical questions could be answered, as a flow-on effect, more insights could be generated on how to evaluate the success of such interactions. The result will be not only more engaging eating experiences, but also the potential to influence when and how and what people eat. This can have major health implications, possibly address major issues such as overeating that results in obesity and then a higher risk of diabetes, heart disease, stroke, bone and joint diseases, sleep apnea, cancer, and overall reduced life expectancy and quality of life [1].

Interrogating such topics is important, as otherwise industry advances will drive the field forward that can easily dismiss or oversee negative consequences when it comes to combining computational technology and food. It is imperative to get ahead of the curve and steer the field in the right direction through an interdisciplinary approach involving a set of experts brought together through the seminar.

Although there is an increasing number of systems emerging, there is limited knowledge about how to design them in a structured way, evaluate their effectiveness and associated user experiences as well as how to derive theory from them to confirm, extend or reject an existing theory. The seminar, therefore, examined these in order to drive a more positive future around interactive food.

Understanding the role of computational technology in this area is a way to make a positive contribution and guide the field in a positive way. There are a couple of areas of imminent importance, and we highlight these here:

- With advances in ubiquitous sensing systems, such as wearables, personal data becomes available in abundance. Such personal data can be embedded into food in order to either communicate it to users in engaging ways or as a way to personalize the food, such as when presenting meals that contain only those calories previously expended. If users understand such data visualizations or want to eat size-controlled portions based on personal data is an ongoing question. Issues of privacy and sharing of such data with chefs and kitchens are also open questions to be investigated.
- With advances in persuasive technology (as already utilized in the form of mobile apps that aim to persuade people to eat more healthy food), new opportunities arise to combine

multiple sensor data from an IoT infrastructure to develop more persuasive systems. How people adhere to such approaches and change their behaviour to the better is still an underdeveloped area that needs to be investigated.

- Big data already used individually by producers, manufacturers and kitchens will increasingly converge, allowing to monitor and influence the supply chain from the farm to the diner's plate. This can help to optimize the sourcing of local produce, reducing the environmental impact through reduced transport distances and a reduction of food waste. How to make sense of such data and use machine learning and other AI advances to utilize this data, so it makes a difference to every part of this supply chain, is an important area for future work.
- Advanced sensor systems can now sense eating actions, such as through jaw movement [9]. By using machine learning, we can now gain an increased understanding of how people eat. This can inform the design of interactive systems that help people make better future eating choices. For example, people might stop overeating if a system could tell them that their stomach will produce a "full" feeling earlier than the 20 minutes it usually takes.
- With the advances of mixed-reality systems like VR headsets and augmented reality on mobile phones, new opportunities arise on how to augment food. For example, prior work has shown that people who perceive cookies through the use of augmented reality to be bigger than they actually are will change how much they eat [11]. There is therefore a significant opportunity to employ mixed-reality to change and offer new opportunities on what and how we eat. How to draw from and incorporate multimodal interaction design theory already established in computer science is an open question for the community to investigate.
- Robotic systems allow to prepare and serve food in novel and interesting ways, for example, robotic arms can already be purchased to be installed in personal kitchens and robots already serve ice cream in public ice cream shops. How to design such interactions so that they are engaging and safe, while still considering the joy and benefit from being engaged in cooking activities, is an interesting area for future work.
- Multisensory integration research has allowed to better understand how humans integrate sensory information to produce a unitary experience of the external world. It is reasonable to expect that technology will keep advancing and sensory delivery will become more accurate. In addition, our understanding of the human senses and perception will become more precise through large scale data from HCI and integration research. As such, there is the potential to systematize our definition of multisensory experiences, through adaptive, computational design. This is exciting but at the same time carries big questions on the implications of multisensory experiences as well as our responsibility when developing them.

The seminar began with talks by all attendees, in which they presented their work in the area and what they thought the biggest challenges are from their perspective the field is facing. After the presentations concluded, no more slides were used for the remainder of the week, with all activities being conducted either as a townhall meeting or in breakout groups. This was supplemented by optional morning and evening activities, such as jogging, beach volleyball, foosball, or cycling.

The structure of the seminar was based around theory, design and their intersection.

References

1 The world health report 2010.

2 Home – niska retail retail, Oct 2021.

3 Rob Comber, Eva Ganglbauer, Jaz Hee-jeong Choi, Jettie Hoonhout, Yvonne Rogers, Kenton O'hara, and Julie Maitland. Food and interaction design: designing for food in everyday life. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 2767–2770. 2012.

4 Rob Comber, Jaz Hee jeong Choi, Jettie Hoonhout, and Kenton O'Hara. Designing for human – food interaction: An introduction to the special issue on "food and interaction design". *International Journal of Human-Computer Studies*, 72(2):181–184, 2014.

5 Andrea Grimes and Richard Harper. Celebratory technology: new directions for food research in hci. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 467–476, 2008.

6 Rohit Ashok Khot, Deepti Aggarwal, Ryan Pennings, Larissa Hjorth, and Florian 'Floyd' Mueller. Edipulse: investigating a playful approach to self-monitoring through 3d printed chocolate treats. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6593–6607, 2017.

7 Rohit Ashok Khot, Jeewon Lee, Deepti Aggarwal, Larissa Hjorth, and Florian 'Floyd' Mueller. Tastybeats: Designing palatable representations of physical activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2933–2942, 2015.

8 Rohit Ashok Khot, Ryan Pennings, and Florian 'Floyd' Mueller. Edipulse: supporting physical activity with chocolate printed messages. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1391–1396, 2015.

9 Naoya Koizumi, Hidekazu Tanaka, Yuji Uema, and Masahiko Inami. Chewing jockey: augmented food texture by using sound based on the cross-modal effect. In *Proceedings of the 8th international conference on advances in computer entertainment technology*, pages 1–4, 2011.

10 Florian 'Floyd' Mueller, Tim Dwyer, Sarah Goodwin, Kim Marriott, Jialin Deng, Han D. Phan, Jionghao Lin, Kun-Ting Chen, Yan Wang, and Rohit Ashok Khot. Data as delight: Eating data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

11 Takuji Narumi, Yuki Ban, Takashi Kajinami, Tomohiro Tanikawa, and Michitaka Hirose. Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 109–118, 2012.

12 Marianna Obrist, Rob Comber, Sriram Subramanian, Betina Piqueras-Fiszman, Carlos Velasco, and Charles Spence. Temporal, affective, and embodied characteristics of taste experiences: A framework for design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2853–2862, 2014.

13 Marianna Obrist, Yunwen Tu, Lining Yao, and Carlos Velasco. Space food experiences: Designing passenger's eating experiences for future space travel scenarios. *Frontiers in Computer Science*, 1, 2019.

14 Johannes Schoning, Yvonne Rogers, and Antonio Kruger. Digitally enhanced food. *IEEE pervasive computing*, 11(3):4–6, 2012.

15 Carlos Velasco and Marianna Obrist. Multisensory experiences: A primer. *Frontiers in Computer Science*, 3, 2021.

16 Carlos Velasco, Marianna Obrist, Gijs Huisman, Anton Nijholt, Charles Spence, Kosuke Motoki, and Takuji Narumi. Editorial: Perspectives on multisensory human-food interaction. *Frontiers in Computer Science*, 3, 2021.

17 Carlos Velasco, Marianna Obrist, Olivia Petit, and Charles Spence. Multisensory technology for flavor augmentation: a mini review. *Frontiers in psychology*, 9:26, 2018.

18 Chi Thanh Vi, Asier Marzo, Damien Ablart, Gianluca Memoli, Sriram Subramanian, Bruce Drinkwater, and Marianna Obrist. Tastyfloats: A contactless food delivery system. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pages 161–170, 2017.

19 Chi Thanh Vi, Asier Marzo, Gianluca Memoli, Emanuela Maggioni, Damien Ablart, Martin Yeomans, and Marianna Obrist. Levisense: A platform for the multisensory integration in levitating food and insights into its effect on flavour perception. *International Journal of Human-Computer Studies*, 139:102428, 2020.

20 Wen Wang, Lining Yao, Teng Zhang, Chin-Yi Cheng, Daniel Levine, and Hiroshi Ishii. Transformative appetite: shape-changing food transforms from 2d to 3d by water interaction through cooking. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6123–6132, 2017.

21 Yan Wang, Zhuying Li, Robert Jarvis, Rohit Ashok Khot, and Florian 'Floyd' Mueller. iscream! towards the design of playful gustosonic experiences with ice cream. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2019.

2 Table of Contents

3 Introduction

3.1 PechaKutchas

The seminar began with an embodied introduction round, where everyone introduced their name including a bodily gesture, repeated by everyone, in order to quickly learn everyone's names.

Then, everyone gave their introductory talk, which participants had prepared beforehand. These talks were delivered in a PechaKucha format that was time-constrained to 5 minutes and favored visual material and aimed to present a personal account of participant's experience with the coming together of interactive technology and food. The preparation instructions to participants asked them to present key readings on the topic of HFI that we also shared beforehand, so to have a common ground understanding of the existing literature. This collection of important writings in HFI can now serve as universal library as curated by experts in order to guide new entrants to the field (such as junior postgraduate students) that supervisions can guide them to, instead of curating such collections themselves again and again from scratch.

We had also asked participants to present their personal "Grand Challenges" they had encountered in their work so far. During the presentations, we asked everyone to use the whiteboard in the room to note down any Grand Challenges that they could identify, which we loosely grouped on the whiteboard based on the headings of "technology", "users", "society", "design" and "other" inspired by the groupings previously identified in other "Grand Challenge" publications in HCI [1, 2, 3]. The "other" category, our wildcard, was considered to be either "theory", "responsible design", or "experience" at this stage.

The disadvantage of this format was that not everything that participants did around the topic could have been presented, however, the timing allowed to discuss all questions that arose, of which they were plenty. The associated discussions helped to identify additonal challenges that we added to the whiteboard as well as helped to refine existing ones. Discussions also helped clarify whether some of them can indeed be considered "grant" or whether they are important, but not "key" to be solved for the future of HFI.

3.2 Grouping activity

We then formed teams in order to work on refining and grouping the challenges that were identified on the whiteboard. During a breakout group session, the teams were working individually to try to identify groupings and refining the Grand Challenges. Each team started off with one of the existing headings to see if this grouping does make sense as a starting point and applies to HFI. Results were documented and collated and then discussed together again with the entire group in order to resolve any instances where one Grand Challenge was put into different groups by the teams. Such multi-assignments were easily resolved, however, the groupings, in particular their namings, were heavily debated. Everyone agreed on the "technology" grouping, but the user grouping was relatively light in content in comparison. The society grouping was discussed in terms of whether it had the right name. Even more contentions was the "design" grouping, as it quickly turned into an "other" grouping where everything went that did not fit in anywhere else. The most debated grouping was the "other" grouping, as it was now discussed whether it should be called "sustainable design" or could be merged with the "design" grouping altogether. It was decided to resolve this the next day.

However, critical voices were also raised, questioning whether some of the Grand Challenges are not specific enough for HFI but rather apply for HCI projects more generally. Furthermore, it was questioned whether some of the proposed Grand Challenges are so big that they are outside the scope of HFI or cannot be solved by HFI, in particular those that were concerned with sustainability and associated supply chains. This was illustrated through examples where interactive technology could help address major issues around food, yet it is not a challenge for the field of HFI that hinders the field from flourishing, such as the use of interactive technology to optimize supply chains to avoid food getting off during transport due to inefficient routing: certainly an important area of concern where technology could be useful, however, not a challenge that hinders the development of the HFI field significantly.

3.3 Interactivity session

An interactivity-style session involved participants trying out interactive systems concerning food. We had a 3D animated display that showed a person's heart in a mobile format, which was connected to a heart rate sensor worn by the user. Such a system could display how a person's interior body [11] responds to a particular food intake.

We also tried the biosensors used in the interactive system by Kunze et al. [4] that captured electrodermal activity (EDA) and heart rate by wrist-worn devices reflecting participants' physiological reactions in real-time. This data is then wirelessly streamed over Wifi into a bespoke application that allows to pass it onto an OSC capable software, like touchdesigner, to visualize the data. This could be used to measure people's responses during dining experiences even in large social groups, such as function settings like birthdays and weddings.

We also tried out sensor-equipped eyeglasses that can detect when a user touches their nose, when they blink and whether they look left or right. Most relevant, the captured data can also be used to detect chewing activity, which can be important when considering any eating behavior change-type of system.

Prior research showed that the weight distribution of a cup can result in a different flavor perception of the drink [5]. Based on this finding, Hirose and Inami [6] developed a system that allows to distribute the weight when a person drinks out of a cup: a motor shifts the weight in the drinking vessel contraption based on the angle of the cup. When the user is about to complete two-thirds of a drink, the more towards the top of the cup. User reports included comments such as that the apple juice that was used tasted more sweet after the weight distribution occurred. The system is still in early development, so several suggestions were made for further iterations, such as reducing the noise of the motor that might also influence the taste perception, and replace the linear motor actuation with a more complex algorithm that produces a more smooth weight distribution transition.

We also tried an "analog" food experiment where we ate different jelly beans while holding our nose in order to demonstrate the extent to which smell informs our overall taste experience.

Experiencing these systems first-hand not only generated further discussions around the opportunity of interactive technology to enhance or enrich the eating experience, it also inspired ideas of what other systems could be developed and how some of the Grand Challenges could be solved.

The interactivity session also brought to light the discussion of what the vision of HFI is. It was discussed whether the Grand Challenges themselves can help us identify a vision for the field, representing a bottom-up approach, or whether we should construct a vision first in

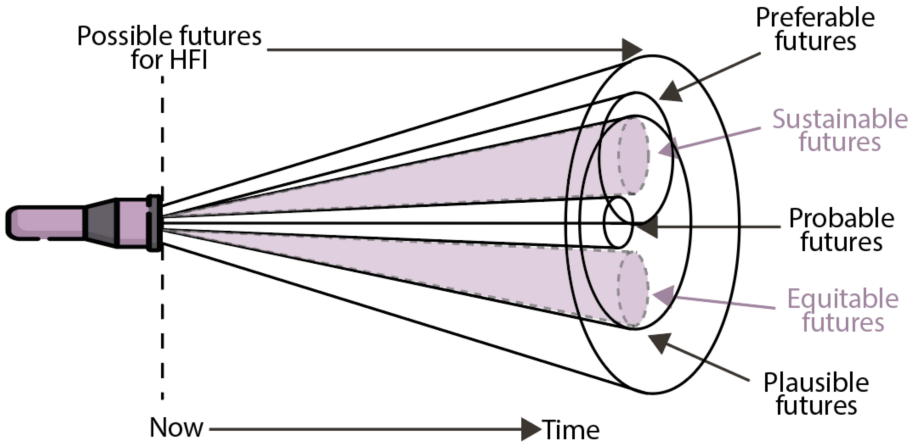


Figure 1 A "flashlight" visualization of futures.

order to guide the development of the Grand Challenges in a top-down fashion. The "use of interactive technology to enhance or enrich the food experience" was complemented by also considering the wider systems around it, like cooking or social interactions and deemed to be sufficient for now, although it was considered to be more of a mission rather than a vision.

3.4 Peer-review

In order to refine the Grand Challenges, we conducted a peer-review session in which the different teams from the previous grouping exercise peer-reviewed the work from one other group, respectively. They were also encouraged to comment, add and edit any content that was shared in an online document as they saw fit. This resulted in a more structured set of grant challenges that were more conform in the way they were articulated, bringing them more "in line" with one another. This activity helped to refine, in particular, the "technology", "user" and "design" groupings, however, the "society" and now called "meta" groupings were not making much progress and it was decided that they needed further work.

3.5 Speculating session

Based on the prior activities, participants then engaged in a speculation exercise in order to refine the Grand Challenges through imagining what the field of HFI would look like in 2052 – a 30 years timeframe – if the Grand Challenges would not be solved. For this, the speculative design approach [7] was used to collectively reimagine possible, probable, plausible, and preferable futures. We encourage participants to also consider sustainable and equitable futures using a "flashlight" visualization (Figure 1).

Based on this, new team formations were asked to speculate about direct and indirect futures if a particular challenge (based on their own choosing) will not have been addressed. Additionally, participants were instructed to imagine what a newspaper article would look like that responds to this at that point in the future, which was supplemented by a speculative Tweet/news item. Together, these speculative media snippets were encouraged to develop as a way to help thinking more concretely about the Grand Challenges and their implications, cementing the idea that they are indeed "grand", that is, key for the positive development of the field as a whole.

The resulting speculations were very useful in helping to cement the idea that the Grand Challenges we identified can indeed be considered to be key for the future of the field. In particular, the exercise helped to underline for everyone how important the work on the Grand Challenges is to help the HFI field as a whole.

In addition, the session also sparked additional discussions not previously had around what are desirable food futures and how HFI could help with that. Maybe frameworks from other, but related areas, could help with that? For example, the "SPRUCE" framework [8], developed to create a desirable future for autonomous vehicles through providing a practically-oriented structure, was considered as exemplar to what HFI could benefit from. The "SPRUCE" framework asks for "safe, predictable, reasonable, uniform, comfortable, and explainable" autonomous vehicles, raising the thought that HFI should similarly aim for "safe" (as in food safe), "healthy-in the long run" (facilitating healthy eating), individuality-appreciating (respond to the highly individual nature of food experiences, see especially the need to consider dietary choices and food allergies), local (responding to the highly local aspect of where food grows and is produced), time-sensitive (responding to the fact that food generally has a very limited shelf-life, especially when compared to other material HCI is concerned with, like plastic in personal fabrication HCI), and social (responding to the highly social nature of HFI experiences that produced its own term: commensality [9]).

Speculating about the future revealed the complexity about the topic but also brought to the fore the responsibilities we have to consider when designing future HFIs. Inspired by the SPRUCE framework, the team embraced responsible design principles, developing a dedicated SPROUT framework and vision for HFI to promote safe, personalized, responsible, original, uniform, and transformational interactive food futures.

3.6 2nd peer review session

We conducted another peer-review session in which we critically examined the refined Grand Challenges as expressed in the previous round. In particular, with this session, the goal was to arrive at a more textual representation of a Grand Challenge. For this, a template was provided based on prior structures as expressed in previous Grand Challenge publications in HCI [1, 2, 3]. This structure was provided in order to ensure that each group does not forget about particular aspects that need to be fleshed out. For example, every Grand Challenge should begin with a definition or explanation that ensures that the reader has a clear understanding of what the challenge is. The structure also aimed to remind authors to explain why the challenge is "grand" enough to find its way into the document. Furthermore, the structure asked authors to consider and express what happens if the Grand Challenge will be resolved and what happens if not. The results from this session were again documented in an online document and shared with the group that elicited feedback that helped refine the Grand Challenges.

3.7 Writing session

Now that participants received feedback on their individual work both within their own group, from other groups, and the entire cohort, it was time to start dedicated writing sessions where participants worked in small teams on fleshing out the text as it will go into the article for each Grand Challenge. Participants were encouraged to look back at the

structure provided above again in order to follow a common structure when expressing a Grand Challenge. In particular, participants were encouraged to work on a shared document that was provided by the organizers so that everyone could see everyone else's work on the Grand Challenge while it was emerging, giving participants the opportunity to quickly ask questions as they emerged or avoid duplication of points being made. For this, it was useful that the Dagstuhl location had enough individual rooms and breakout spaces that allowed concentrated team work, yet these rooms and breakout spaces were very close-by so that participants could quickly stick their head into another space to ask any clarifying questions.

3.8 Career hike

We also organized a hike through the local landscape in order to nurture not just the mind, but also the body, making use of the tight interlink between a healthy body and a healthy mind. We instructed participants to consider using the opportunity of walking next to each other, in contrast to sitting opposite each other, facing each other, like in a meeting scenario, especially over videoconferencing, to have different kinds of topics of conversations. In particular, we suggested to discuss each other's careers and, speaking to the wide range of career spans represented in our participant list, provide career advice and also ask for career advice. This could make use of the opportunity to connect with people outside the topic of the seminar, yet relevant to their academic or industry career.

3.9 Personas

In order to illustrate the dangers that lurk if the Grand Challenges will not be solved, a session was conducted in which participants were encouraged to develop personas. The teams were encouraged to present the personas they developed to the other teams in an engaging and entertaining way, going beyond presentation slides. Most teams decided to role-play the scenarios around their personas, speaking to the embodied nature of food. For example, one team conducted a play around a single mother who relies too much on her cooking robot, while another team presented food made out of play-do that could nourish people at the most optimal level but requires social approval in order to offer access. These personas not only offered an entertaining perspective on the Grand Challenges, but also highlighted that solving them is not only an abstract exercise but affects real people.

3.10 Thinking through writing

Having engaged more acutely with the consequences if not solving the Grand Challenges, the participants were encouraged to further refine the Grand Challenges text document. In particular, the idea was to do thinking through writing, in which participants, in teams, collaboratively develop the text for each Grand Challenge and peer review it while it is emerging. This "thinking through writing" session resulted in several refinements of particular Grand Challenges. It also identified the need to spend more time on the categorization labels, as it was identified that they were too generic and not particular to HFI. To address this, a different visualization for the presentation of the groupings of the Grand Challenges was developed, based on prior work that also preferred an illustration [3] over a table in

order to present the results [2]. With this new illustration, a new editing session was started where participants were instructed to further refine the descriptions of the Grand Challenges while also considering which ones are key for the article and which ones might need to go for brevity purposes. As part of the thinking through writing process, one group decided to develop the vision section of the article further, building on the SPRUCE framework that has by now become the SPROUT framework that aims to guide HFI practitioners new to the field but are unsure what has already been done and what one needs to consider if wanting to conduct research.

3.11 Collating and reflection activity

The final activity involved collating all the work that has been done so far and discussing it in a townhall-style environment. This was done to ensure nothing important had been missed while trying to attempt to give everyone an equal voice across the different Grand Challenges. Reflection across the entire group was then facilitated through the presentation of three final questions everyone should think about and then were free to articulate, which were suggested to start with "I like ...", "I wish ...", and "I wonder ...". These questions were aimed to facilitate reflection on what was achieved and could be improved upon for further seminars like this.

References

1 Juliet Norton, Ankita Raturi, Bonnie Nardi, Sebastian Prost, Samantha McDonald, Daniel Pargman, Oliver Bates, Maria Normark, Bill Tomlinson, Nico Herbig, et al. A grand challenge for hci: Food+ sustainability. *interactions*, 24(6):50–55, 2017.

2 Jason Alexander, Anne Roudaut, Jürgen Steimle, Kasper Hornbæk, Miguel Bruns Alonso, Sean Follmer, and Timothy Merritt. Grand challenges in shape-changing interface research. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–14, 2018.

3 Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, et al. Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

4 Moe Sugawa, Taichi Furukawa, George Chernyshov, Danny Hynds, Jiawen Han, Marcelo Padovani, Dingding Zheng, Karola Marky, Kai Kunze, and Kouta Minamizawa. Boiling mind: Amplifying the audience-performer connection through sonification and visualization of heart and electrodermal activities. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–10, 2021.

5 Betina Piqueras-Fiszman, Vanessa Harrar, Jorge Alcaide, and Charles Spence. Does the weight of the dish influence our perception of food? *Food Quality and Preference*, 22(8):753–756, 2011.

6 Masaharu Hirose and Masahiko Inami. Balanced glass design: A flavor perception changing system by controlling the center-of-gravity. In *ACM SIGGRAPH 2021 Emerging Technologies*, pages 1–4. 2021.

7 Anthony Dunne and Fiona Raby. *Speculative everything: design, fiction, and social dreaming.* MIT press, 2013.

8 Julian De Freitas, Andrea Censi, Bryant Walker Smith, Luigi Di Lillo, Sam E. Anthony, and Emilio Frazzoli. From driverless dilemmas to more practical commonsense tests for automated vehicles. *Proceedings of the National Academy of Sciences*, 118(11):e2010202118, 2021.

9 Radoslaw Niewiadomski, Eleonora Ceccaldi, Gijs Huisman, Gualtiero Volpe, and Maurizio Mancini. Computational commensality: from theories to computational models for social food preparation and consumption in hci. *Frontiers in Robotics and AI*, 6:119, 2019.
10 Zhuying Li, Rakesh Patibanda, Felix Brandmueller, Wei Wang, Kyle Berean, Stefan Greuter, and Florian 'Floyd' Mueller. The guts game: towards designing ingestible games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pages 271–283, 2018.
11 Zhuying Li, Yan Wang, Stefan Greuter, and Florian 'Floyd' Mueller. Playing with the interior body. *Interactions*, 28(5):44–49, 2021.

4 Overview of Talks

4.1 Multi-Sensory (tasty) Experience in Virtual Reality

Christopher Dawes (University College London, UK, c.dawes@ucl.ac.uk)

Human-Food Interaction is a completely novel area to me. Coming from a background in experimental psychology and computer science, I am interested in applications of Virtual Reality and user experience. In my recent transition to multi-sensory experience research, our team have been investigating how factors such as food shape and ambient lighting may affect taste experiences in Virtual Reality. As I near the end of my PhD, I am looking to explore new potential research areas such as HFI. Specifically, I am interested in how we can integrate senses other than vision and sound into VR, such as taste experiences.

4.2 "Good food ends with good talk": investigating social interaction through the lens of Computational Commensality

Eleonora Ceccaldi (University of Genoa, IT, eleonoraceccaldi@gmail.com)

Food is a social fact: commensality, is a far more complex concept than just sitting together at the dining table. It has been linked to improved health and well-being, food enjoyment, and positive affect among co-diner. Commensality is a multifaceted phenomenon that requires a multidisciplinary approach to be fully understood and thoroughly investigated. Computational Commensality is an approach that grounds on affective computing and social signal processing, bringing together models and theories from social and cognitive sciences with techniques and methods used in Artificial Intelligence and Human-Computer Interaction. The talk will start with the social side of eating to then illustrate a study investigating remote eating experiences through the lenses of Computational Commensality. The results of the study can help shed more light on how technology can foster commensality and how Computational Commensality could be leveraged to better understand commensal, and hence social, interactions.

4.3 Co-designing to realize the playful potential of food practices

Ferran Altarriba Bertran (Escola Universitària ERAM, Universitat de Girona – Salt, ES & Gamification Group, Tampere University, FI, ferranaltarriba@gmail.com)

Ferran Altarriba Bertran is an interaction designer and researcher whose work explores the design of technologies and experiences that add an element of playfulness to people's mundane activities. In this talk, Ferran will discuss some of his recent work in the space of Playful Human-Food Interaction research. His presentation will discuss the intersection between play, food, and technology from three angles: conceptual, i.e. why are play and playfulness needed in our interactions with and around food?; methodological, i.e. how do we design play-food experiences that are contextually meaningful?; and design-oriented, i.e. what are specific actions designers can take to playfully reframe our food systems and practices? Overall, Ferran's talk will provoke designers and researchers to embrace fun and joy as foundational values in the design of food-related technology, and will provide inspirational starting points to facilitate that move.

4.4 Computer-Food Integration

Florian 'Floyd' Mueller (Monash University – Melbourne, AU, floyd@exertiongameslab.org)

To date the coming together of food and technology is mostly one similar where computers and the human body were at the beginning of wearable computing: computers were initially developed to be lighter so that they could be worn. Only more recently we realized that there are more affordances besides weight that need to be considered (such as always-available, context-aware, always-on, etc.) as they allow to meaningfully integrate the human body and the computer into an assemblage. In this talk, I ask the question what the equivalent is for computer-food integration by drawing on recent experiments in our lab on the coming together of food and technology that also considers experiential, not just instrumental, perspectives.

4.5 SMARTMOBILITY: A toolbox for behavior change in the field of mobile wellness addressing physical activity and normal eating

Harald Reiterer (University of Konstanz, DE, harald.reiterer@uni-konstanz.de)

SMARTMOBILITY is part of the interdisciplinary project SMARTACT. The main aim of SMARTACT is to develop and empirically test the efficacy of a toolbox for mobile, real-time interventions targeting normal eating and physical activity using mobile technology (smartphones, body monitoring). The SMARTACT toolbox encompasses different mobile and in-person intervention tools based on "what people do" (behavioral pattern), "why people

do what they do" (psychosocial and contextual triggers of behavior), and "when people do what they do" (timing of behavior and triggers). A mobile app – the wellness diary – combines several tools to promote physical activity and normal eating: The physical activity tracking tool keeps track of users' physical activities (e.g., step count), the food journal lets users document their food intake, the questionnaire tool gathers eating motives, and the feedback tool provides the user interactive multidimensional visualization of gathered data. The toolbox provides high quality and depth of collected data. A user-centered design approach minimizes the burden of manual data entry and maximizes ease of use. For the researcher, an export tool provides export options for data stored on the server. This data can then be used for collaborative immersive visual data analysis.

4.6 CyberFood: Food-Computation Integration

Jialin Deng (Monash University – Melbourne, AU, jialin@exertiongameslab.org)

Contemporary human-food interaction design is predominantly a technology-driven endeavor, which has highlighted the functionality and novelty of computing technology. Such a technology-centric approach might outweigh the exploration of inherent affordances of food, such as the food's material properties emphasizing its aesthetic, affective, sensual, and sociocultural qualities. Here, I introduce a new approach to food-computation integration that employs food as a primary material to realize computation. I present a "Research through Design" exploration of designing food as computational artifact through a case study called the "Logic Bonbon", which is a liquid-centered dessert that can regulate its own flavor and visual presentation. Through the design-led exploration, I hope to unpack how computational qualities of food could be leveraged in the development of novel human-food interactions and shape the future of food innovation.

4.7 Food-Grade Sensors Made of Biomaterials

Jürgen Steimle (Saarland University – Saarbrücken, DE, steimle@cs.uni-saarland.de)

Designers and makers are increasingly interested in leveraging bio-based and bio-degradable 'do-it-yourself' (DIY) materials for sustainable prototyping. Self-produced bioplastics possess compelling properties such as self-adhesion but have so far not been functionalized to create soft interactive devices, due to a lack of DIY techniques for the fabrication of functional electronic circuits and sensors. I present a DIY approach for creating Interactive Bioplastics that is accessible to a wide audience, making use of easy-to-obtain bio-based raw materials and familiar tools. It enables additive and subtractive fabrication of soft circuits and sensors. Our biomaterials possess attractive properties for edible interfaces. I present an application case that realizes functional capacitive touch sensors which can be embedded within food.

4.8 Food for the Mind

Kai Kunze (Keio University – Tokyo, JP, kai@kmd.keio.ac.jp)

One important characteristic of the human mind is that it has significant fluctuations in productivity and capacity. Our mind has ebb and flow, and is affected by various factors, some of which we do not even realize. The types of food we eat has a large impact on the state of our mind (cognitive performance, sleepiness, alertness etc.). These cognitive fluctuations manifest in patterns in human behavior and physiological signals (body temperature, eye movements, galvanic skin response etc.) related to the Autonomous Nervous System and can be captured with unobtrusive sensors embedded in glasses, granments, clothes etc.I want to develop tools that enable people to be aware on the impact of food intake on their everyday cognitive functioning and to live more healthily.

4.9 Thinking Big: The Five Senses

Marianna Obrist (University College London, UK, c.dawes@ucl.ac.uk)

We recognise the rich potential of the human senses (vision, hearing, touch, taste and smell) when designing human-food interactions. This talk will engage the audience in a reflection on how to design multisensory food experiences driven by key questions, namely, the why (the rationale/reason), what (the impression), when (the event), how (the sensory elements), who (the someone), and whom (the receiver), associated with a given multisensory experience design, exemplified through food experiences. In other words, we need to think of the context of eating (e.g. home, restaurant, other spaces), the technology we use/design (e.g. augmented cutlery/utensils, VR), and the user and the desired experience. All that is further embedded in relevant responsible innovation framework, asking uncomfortable questions, anticipating unintended consequences and possible negative but also positive outcomes for future human-food interaction designs.

4.10 Fill the World with Emotion

Matti Schwalk (Sony Europe B.V. | R&D Center – Stuttgart, DE, Matti.Schwalk@sony.com)

User experience design should always aim to engage users on different levels of perception, emotion, and cognition. Vision and sound are still the predominant interaction channels, mainly because they can be used to transmit large amounts of information in a short time, also digitally. Meanwhile, haptic/tactile cues have been established as the third modality, e. g. to generate more immersive entertainment scenarios. For future interactive experiences, though, and to evoke even deeper emotions, it will be crucial to include smell and taste sensations, too. Possible use cases are increased realism and sense of presence in virtual

environments or shared traveling and eating. By means of multisensory integration and interdisciplinary research, we are following Sony's purpose which is to "fill the world with emotion, through the power of creativity and technology."

4.11 Introducing Artificial Commensal Companions

Maurizio Mancini (University of Rome "Sapienza", IT, m.mancini@di.uniroma1.it)

How could it be possible to "synthesize" the experience of commensal eating in a HCI setting? That is one of the questions we recently started to investigate through a new type of interface that we called Artificial Commensal Companion (ACC). ACCs could bring the benefits of commensal eating to, for example, people who voluntarily choose or are constrained to eat alone (e.g., in a hospital or rural setting, or during sanitary lockdown). In the presentation, we introduce an interactive system implementing an ACC in the form of a robot with non-verbal socio-affective capabilities. Future tests are already planned to evaluate its influence on the eating experience of human participants.

4.12 Taste & Flavor Integration as Source of Extended User Experience

Nadejda Krasteva (Sony Europe B.V. | R&D Center, Stuttgart, DE, nadejda.krasteva@sony.com)

Human perception of the world occurs through several interaction channels which provide both factologic information about the individual's surrounding and affect the behaviour and emotional state of the human in particular context. Taste and flavour sensations are deeply encoded in our brain and as such contribute to a large extent to emotions, memory and cognition, and to the user experience in the physical world. We are interested in designing possibilities to transfer and even extend experience related to taste and flavor to the virtual space. This transfer requires solving the fundamental problem of de- and re-materialization of the chemical interactions governing taste and flavor perception from real into the cyber space.

4.13 Designing Playful Technologies to Nurture Human-Microbial Interaction and Its Understanding

Nandini Pasumarthy (RMIT University – Melbourne, AU, nandini.pasumarthy@rmit.edu.au)

Our interaction with microbes is crucial and dictates not just a momentary gastronomic gratification but also sets the stage for future generations' gastronomic delight and their ability to metabolise food. To secure these experiences for future generations, understanding

how to nurture gut microbial diversity is important. The biggest challenge though is not just the translation of this rapidly advancing science, or its multifaceted influence on gut microbial diversity, but also its transformation into novel experiences to generate public understanding. For this, we propose the design of interactive play-based technologies to develop alternative framings that foster understanding and reflection on human-microbial interaction. Prior works emphasised the importance and the need for more-than-human perspectives for sustainable food futures. We contribute to this through the design and development of interactive playful tools. Our learnings and strategies can aid future playful design explorations towards nurturing human-microbial relationships to promote sustainable food behaviours.

4.14 Food for informing a new foundation of human-technology relationship

Patrizia Marti (University of Siena, IT, patrizia.marti@unisi.it)

After long focusing on food as an object, taking on its shape, materials, naming, packaging and rituals of use, food design has opened the door to the design of enhanced food interaction that combines the embodied sensory stimulation of food with multi-sensory or cross-modal interactions possibilities offered by technology. I'm interested in exploring the possibilities of food as a design material that can afford cultural and transformations, and experimenting with prototypes for instantiating a more general theory of interaction which can inform the design of technology enhanced experiences. Through designing for human-food interaction we necessarily need to consider all senses, how they are solicited by food, the subtleties which make the experience of eating memorable and worthy of sharing with others. Food offers us the possibility to be engaged somatically, enabling us to connect feeling, thinking, movement, and expression into one subjectivity. Learning how to acquire a somatic and experiential sensibility can help designing richer and more engaging future technologies.

4.15 Rethinking the "T" in HFI

Rohit Ashok Khot (RMIT University – Melbourne, AU, rohit.a.khot@gmail.com)

The field of Human-Food Interaction (HFI) has garnered considerable currency in recent years in the field of HCI with researchers exploring how technologies can be designed to support different aspects of food practices. However, the majority of the exploration is still centered around digital or is computer/arduino-mediated, here we see an opportunity to look beyond the existing technologies and rethink food as the technology itself. In this talk, I will cover some of HAFP Research Lab's works that address this theme.

4.16 Design Thinking, Design Doing: Co-Designing Future Food Experiences

Soh Kim (Stanford University, US, sohkim@stanford.edu)
Sahej Claire (Stanford University, US, saclaire@stanford.edu)
Kyung Seo (Kay) Jung (Stanford University, US, kjungk@stanford.edu)
Neharika Makam (Stanford University, US, npmakam@ucsd.edu)

Food is a form of media that allows us to experience creativity, pleasure, connectedness, trend-seeking, and relaxation. Eaters, as users, are motivated to consume food by many different factors: community, familiarity, convenience, culture, hedonism, functionality, health, morality, novelty, and more. Through the Design Thinking, Design Doing workshop, we aim to speculate on the consequences that may arise from the implementation of food technologies in the future. By considering the broader future of food technology, we aim to provoke discussion about potential utopian outcomes and the considerations required for the field of Human-Food Interaction in the present.

4.17 Understanding the design of Playful Gusotsonic Experiences

Yan Wang (Monash University – Melbourne, AU, wangyandesign@gmail.com)

Prior human-computer interaction research shows that overlaying sounds to eating can change how people experience food, affecting how and what we eat. However, how to design such gustosonic experiences – referring to multisensory interactions between sounds and the act of eating/drinking – is not well understood. In this talk, I introduce my three gustosonic systems to arrive at the understanding of the design of playful gustosonic experiences, in particular the design of "sonic straws" which is straws that can play personalized notes. This work advances interaction design theory by contributing to the enrichment of eating/drinking through playful design, furthering the future of the food experience.

Participants

- Ferran Altarriba Bertran
ERAM University School – Salt, ES
- Sahej Claire
Stanford University, US
- Christopher Dawes
University College London, GB
- Jialin Deng
Monash University – Clayton, AU
- Masahiko Inami
University of Tokyo, JP
- Kyung seo Jung
Stanford University, US
- Sohyeong Kim
Stanford University, US
- Nadejda Krasteva
Sony – Stuttgart, DE
- Kai Kunze
Keio University – Yokohama, JP
- Neharika Makam
Stanford University, US
- Florian 'Floyd' Mueller
Monash University – Clayton, AU
- Marianna Obrist
University College London, GB
- Harald Reiterer
Universität Konstanz, DE
- Matti Schwalk
Sony – Stuttgart, DE
- Jürgen Steimle
Universität des Saarlandes, DE

Remote Participants

- Eleonora Ceccaldi
University of Genova, IT
- Simon Henning
Sony Design – Lund, SE
- Rohit Khot
RMIT University – Melbourne, AU
- Maurizio Mancini
Sapienza University of Rome, IT
- Patrizia Marti
Università di Siena, IT
- Nandini Pasumarthy
RMIT University – Melbourne, AU
- Yan Wang
Monash University – Clayton, AU

Report from Dagstuhl Seminar 22281

Security of Machine Learning

Battista Biggio[*1], Nicholas Carlini[*2], Pavel Laskov[*3], Konrad Rieck[*4], and Antonio Emanuele Cinà[†5]

1 **University of Cagliary, IT.** battista.biggio@unica.it
2 **Google – Mountain View, US.** nicholas@carlini.com
3 **University of Liechtenstein – Vaduz, LI.** pavel.laskov@uni.li
4 **TU Braunschweig, DE.** k.rieck@tu-bs.de
5 **University of Venice, IT.** antonioemanuele.cina@unive.it

Abstract

Machine learning techniques, especially deep neural networks inspired by mathematical models of human intelligence, have reached an unprecedented success on a variety of data analysis tasks. The reliance of critical modern technologies on machine learning, however, raises concerns on their security, especially since powerful attacks against mainstream learning algorithms have been demonstrated since the early 2010s. Despite a substantial body of related research, no comprehensive theory and design methodology is currently known for the security of machine learning. The proposed seminar aims at identifying potential research directions that could lead to building the scientific foundation for the security of machine learning. By bringing together researchers from machine learning and information security communities, the seminar is expected to generate new ideas for security assessment and design in the field of machine learning.

Seminar July 19–15, 2022 – http://www.dagstuhl.de/22281
2012 ACM Subject Classification Computer systems organization → Real-time operating systems; Computing methodologies → Machine learning
Keywords and phrases adversarial machine learning, machine learning security
Digital Object Identifier 10.4230/DagRep.12.7.41

1 Executive Summary

Battista Biggio
Nicholas Carlini
Pavel Laskov
Konrad Rieck

Overview

Modern technologies based on machine learning, including deep neural networks trained on massive amounts of labeled data, have reported impressive performances on a variety of application domains. These range from classical pattern recognition tasks, for example, speech and object recognition for self-driving cars and robots, to more recent cybersecurity tasks, such as attack and malware detection. Despite the unprecedented success of technologies based on machine learning, it has been shown that they suffer from vulnerabilities and data leaks. For example, several machine-learning algorithms can be easily fooled by adversarial examples, that is, carefully-perturbed input samples aimed to thwart a correct prediction.

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Security of Machine Learning, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 41–61
Editors: Battista Biggio, Nicholas Carlini, Pavel Laskov, Konrad Rieck, and Antonio Emanuele Cinà
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

These insecurities pose a severe threat in a variety of applications: the object recognition systems used by robots and self-driving cars can be misled into seeing things that are not there, audio signals can be modified to confound automated speech-to-text transcriptions, and personal data may be extracted from learning models of medical diagnosis systems.

In response to these threats, the research community has investigated various defensive methods that can be used to strengthen current machine learning approaches. Evasion attacks can be mitigated by the use of robust optimization and game-theoretical learning frameworks, to explicitly account for the presence of adversarial data manipulations during the learning process. Rejection or explicit detection of adversarial attacks also provides an interesting research direction to mitigate this threat. Poisoning attacks can be countered by applying robust learning algorithms that natively account for the presence of poisoning samples in the training data as well as by using ad-hoc data-sanitization techniques. Nevertheless, most of the proposed defenses are based on heuristics and lack formal guarantees about their performance when deployed in the real world.

Another related issue is that it becomes increasingly hard to understand whether a complex system learns meaningful patterns from data or just spurious correlations. To facilitate trust in predictions of learning systems, the explainability of machine learning becomes a highly desirable property. Despite recent progress in development of explanation techniques for machine learning, understanding how such explanations can be used to assess the security properties of learning algorithms still remains an open and challenging problem.

This Dagstuhl Seminar aims to bring together researchers from a diverse set of backgrounds to discuss research directions that could lead to the scientific foundation for the security of machine learning.

Goal of the Seminar

The seminar focused on four main themes of discussion, consistently with the research directions reported in the previous section:

- Attacks against machine learning: What attacks are most likely to be seen in practice? How do existing attacks fail to meet those requirements? In what other domains (i.e., not images) will attack be seen?
- Defenses for machine learning: Can machine learning be secure in all settings? What threat models are most likely to occur in practice? Can defenses be designed to be practically useful in these settings?
- Foundations of secure learning: Can we formalize “adversarial robustness”? How should theoretical foundations of security of machine learning be built? What kind of theoretical guarantees can be expected and how do they differ from traditional theoretical instruments of machine learning?
- Explainability of machine learning: What is the relationship between attacks and explanations? Can interpretation be trusted?

Overall Organization and Schedule

The seminar intends to combine the advantages of conventional conference formats with the peculiarities and specific traditions of Dagstuhl events. The seminar activities were scheduled as follows:

Schedule	Activities
Day 1	Workshop presentation, short self-introductions by participants, one keynote presentation
Day 2	One keynote presentation on participant results, contributed presentations
Day 3	One keynote presentation on negative results, organization of working groups, one breakout session
Day 4	One breakout session, social event
Day 5	Keynote presentation, reporting from breakout sessions, summary of results

2 Table of Contents

3 Overview of Talks

3.1 Concept Drift in Machine Learning-based Detection Systems

Fabio Pierazzi (King's College London, UK)

Joint work of Federico Barbero, Zeliang Kan, Feargus Pendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, Lorenzo Cavallars
Main reference Feargus Pendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, Lorenzo Cavallaro: "TESSERACT: Eliminating Experimental Bias in Malware Classification across Space and Time", in Proc. of the 28th USENIX Security Symposium, USENIX Security 2019, Santa Clara, CA, USA, August 14-16, 2019, pp. 729–746, USENIX Association, 2019.
URL https://www.usenix.org/conference/usenixsecurity19/presentation/pendlebury

Distribution shifts causing violations of the i.i.d. assumptions are prevalent in the security domain (e.g., malware detection), causing performance decay over time, and making it challenging for Machine Learning (ML) models. The cybersecurity community initially adopted best practices such as k-fold cross validation without a deep understanding of the implications of dataset shift and temporal concept drift (e.g., malware evolution over time).

In this context, we proposed TESSERACT [1] as a framework for proper evaluations in the presence of concept drift, and showed the impact of concept drift in the Android malware domain. When a proper time-aware train/test split is conducted, even performance of state-of-the-art classifiers quickly decay over time; in presence of drift, the k-fold cross validation provides an upper bound of detection performance assuming absence of drift.

This shows that this research area is still open. To mitigate drift, we may adopt a variety of approaches: *retraining* (e.g., with active learning) is one possibility, although "labeling" in the security domain is extremely costly; *classification by rejection* quarantines samples with low confidence (in this context, we propose a conformal prediction-inspired approach for rejecting drifting samples [2]). These approaches show that it is not anymore just about detection performance, instead systems need to be evaluated as a trade-off between accuracy, labeling and quarantine costs.

After identifying that online learning is more challenging than we originally thought [3], we advocate for more research into machine learning approaches robust to drift.

References

1 Feargus Pendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, Lorenzo Cavallaro, *TESSERACT: Eliminating experimental bias in malware classification across space and time*, USENIX Security Symposium, 2019.
2 Federico Barbero, Feargus Pendlebury, Fabio Pierazzi, Lorenzo Cavallaro. *Transcending Transcend: Revisiting Malware Classification in the Presence of Concept Drift.* IEEE Symposium on Security & Privacy, 2022.
3 Zeliang Kan, Feargus Pendlebury, Fabio Pierazzi, Lorenzo Cavallaro, *Investigating Labelless Drift Adaptation for Malware Detection*, AISec Workshop (co-located with ACM CCS), 2021.

3.2 When too good is bad: On the re-use of Datasets in ML Security

Giovanni Apruzzese (PostDoc Researcher at the University of Liechtenstein – Vaduz, LI)

In this (very informal!) talk, I will talk about two problems that (I believe) affect the whole community of ML security. Namely: the peer-review process, which sometimes leads to superficial reviews; and the constant re-use of benchmark datasets (sometimes of domains unrelated to security) which aggravates the previous problem, because reviewers will inevitably tend to familiarize with "benchmarks" (which can be flawed), and are hence more critical to experimental evaluations carried out on datasets they are not familiar with.

To describe such twofold problems, I will present some "stories", derived from my own experience as a "young researcher" in this domain. Specifically, I will highlight that many papers (accepted in top-venues) which rely on "benchmark" datasets are provided with very little information describing the collection and preprocessing of such data [2, 3, 4, 5, 6, 7]. Therefore, neither the authors nor the reviewers believe it necessary to include such information, due to the high "familiarity" of researchers with such datasets. Then, I narrate the backstory of a recently accepted paper that I authored [1], which was rejected 4 times at top security conferences. Among the reasons for such "rejections", one was always related to "missing details about the datasets". Despite being a legitimate observation, as none of the datasets used in that paper were "benchmarks" in the ML security research domain (although some were popular in other domains [8]), all such details were always included in the paper – but in the Appendix, which apparently no reviewer read. Finally, I show that even "benchmark" datasets [10] have flaws (documented by reputable works [9]), thereby showing that relying on benchmarks – despite having some advantages – can be detrimental for our research. Simply put, this talk aims to inspire a change in the current evaluation protocol adopted in our research (from the perspective of both reviewers and authors).

References

1 Giovanni Apruzzese, Rodion Vladimirov, Aliya Tastemirova, and Pavel Laskov. "*Wild Networks: Exposure of 5G Network Infrastructures to Adversarial Examples*." IEEE Transactions on Network and Service Management, 2022.
2 Chulin Xie, Keli Huang, Pin-Yu Chen, and Bo Li. "DBA: Distributed backdoor attacks against federated learning." In International Conference on Learning Representations, 2019.
3 Francesco Croce and Matthias Hein. "Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks." In International Conference on Machine Learning, 2020.
4 Qi Tian, Kun Kuang, Kelu Jiang, Fei Wu, and Yisen Wang. "Analysis and applications of class-wise robustness in adversarial training." In Conference on Knowledge Discovery and Data Mining, 2021.
5 Kaifa Zhao, Hao Zhou, Yulin Zhu, Xian Zhan, Kai Zhou, Jianfeng Li, Le Yu, Wei Yuan, and Xiapu Luo. "Structural attack against graph based android malware detection." In Conference on Computer and Communications Security, 2021.
6 Mani Malek Esmaeili, Ilya Mironov, Karthik Prasad, Igor Shilov, and Florian Tramer. "Antipodes of label differential privacy: Pate and alibi." In Advances in Neural Information Processing Systems, 2021.
7 Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov. "How to backdoor federated learning." In International Conference on Artificial Intelligence and Statistics, 2020.

8 Timothy J. O'shea and Nathan West. "Radio machine learning dataset generation with GNU radio." In Proceedings of the GNU Radio Conference, 2016.
9 Gints Engelen, Vera Rimmer, and Wouter Joosen. "Troubleshooting an intrusion detection dataset: the CICIDS2017 case study." In IEEE Security and Privacy Workshops, 2021.
10 Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization." In International Conference on Information Systems Security and Privacy, 2018

3.3 Where ML Security Is Broken and How to Fix It

Antonio Emanuele Cinà (Ph.D. candidate at Ca' Foscari University of Venice, IT)
Maura Pintor (Postdoc at University of Cagliari, IT)

Machine learning security is attracting considerable interest from the research community and industries due to its practical influence on machine learning data services that today are becoming the cornerstone of applications. However, this interest is not sometimes repaid by a real advancement of knowledge in the field. Partly because of the pressure exerted by the sword of Damocles that haunts researchers today, namely "publish or perish," research in this area focuses more on numbers than on the actual quality of the proposed work. In this talk, we want to show the results of our research and highlight some aspects of machine learning security that we believe are broken and deserve special consideration. In addition, for each issue, a new way to address the problem will be proposed, or the real obstacle that needs to be overcome to bring further knowledge to machine learning security will be highlighted.

The first part of our talk addresses attacks at training time, namely poisoning attacks. Poisoning attacks are staged at training time by manipulating the training data or compromising the learning process to degrade the model's performance at test time. Among the two scenarios, the case where data are influenced by the attacker, namely data poisoning, has attracted increasing attention from ML stakeholders, perhaps after the incident of Tay [9], to the point that now it is considered the largest concern for ML applications [7, 8]. We identified three main categories of data poisoning attacks [6, 5], namely indiscriminate, targeted, and backdoor poisoning attacks. *Indiscriminate* poisoning attacks are staged to maximize the classification error of the model on the (clean) test samples. The attacker aims to reduce the system's availability to legitimate users who can not trust the output of the poisoned model. *Targeted* poisoning attacks influence the model to cause misclassification only for a specific set of (clean) test samples. In *backdoor* poisoning attacks, the training data is manipulated by adding poisoning samples containing a specific pattern, referred to as the backdoor trigger, and labeled with an attacker-chosen class label. This typically induces the model to learn a strong correlation between the backdoor trigger and the attacker-chosen class label. Accordingly, the input samples that embed the trigger are misclassified at test time as samples of the attacker-chosen class, while the pristine samples remain correctly classified. Although multiple poisoning attacks have been suggested to attack or test the robustness of ML models, we observed that state-of-the-art works rely on unrealistic assumptions or do not scale against real production systems.

The second part of our talk is dedicated to attacks at testing time [3, 4]. Rigorous testing against such perturbations requires enumerating all possible outputs for all possible inputs, and despite impressive results in this field, these methods remain still difficult to scale to

modern deep learning systems. For these reasons, empirical methods are often used. These adversarial perturbations are optimized via gradient descent, minimizing a loss function that aims to increase the probability of misleading the model's predictions. To understand the sensitivity of the model to such attacks, and to counter the effects, machine-learning model designers craft worst-case adversarial perturbations and test them against the model they are evaluating. However, many of the proposed defenses have been shown to provide a false sense of robustness due to failures of the attacks, rather than actual improvements in the machine-learning models' robustness. They have been broken indeed under more rigorous evaluations [2, 1]. Although guidelines and best practices have been suggested to improve current adversarial robustness evaluations, the lack of automatic testing and debugging tools makes it difficult to apply these recommendations in a systematic and automated manner.

References

1 Tramèr, F., Carlini, N., Brendel, W. & Madry, A. On Adaptive Attacks to Adversarial Example Defenses. *Advances In Neural Information Processing Systems 33: Annual Conference On Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, Virtual*. (2020)

2 Athalye, A., Carlini, N. & Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. *Proceedings Of The 35th International Conference On Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. **80** pp. 274-283 (2018)

3 Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G. & Roli, F. Evasion Attacks against Machine Learning at Test Time. *Machine Learning And Knowledge Discovery In Databases – European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III*. **8190** pp. 387-402 (2013), https://doi.org/10.1007/978-3-642-40994-3

4 Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. & Fergus, R. Intriguing properties of neural networks. *2nd International Conference On Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. (2014)

5 Cinà, Antonio Emanuele, Kathrin Grosse, Ambra Demontis, Sebastiano Vascon, Werner Zellinger, Bernhard Alois Moser, Alina Oprea, Battista Biggio, Marcello Pelillo and Fabio Roli. Wild Patterns Reloaded: A Survey of Machine Learning Security against Training Data Poisoning. *ArXiv* (2022).

6 Cinà Antonio Emanuele, Kathrin Grosse, Ambra Demontis, Battista Biggio, Fabio Roli and Marcello Pelillo. Machine Learning Security against Data Poisoning: Are We There Yet? *ArXiv* (2022).

7 Grosse, Kathrin, Lukas Bieringer, Tarek R. Besold, Battista Biggio and Katharina Krombholz. "Why do so?" – A Practical Perspective on Machine Learning Security. *ArXiv* (2022).

8 Kumar, Ram Shankar Siva, Magnus Nyström, John Lambert, Andrew Marshall, Mario Goertzel, Andi Comissoneru, Matt Swann and Sharon Xia. Adversarial Machine Learning – Industry Perspectives. *CompSciRN: Other Cybersecurity* (2020).

9 Learning from Tay, Learning from Tay's introduction – The Official Microsoft Blog, https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/ (2016).

3.4 Adversarial Machine Learning in Practice

Kathrin Grosse (Postdoc at University of Cagliari, IT)

Beyond media coverage, few works have attempted to understand the impact of machine learning (ML) security in practice scientifically [3, 4]. While prior work does not exclusively study ML security but also privacy [4] or studies ML security on a company level [3], we present two studies that deepen our knowledge about ML security in practice: One is a study with 15 participants in semi-structured interviews with a drawing task [1], which provides great detail on how industrial practitioners think and approach ML security. Furthermore, our questionnaire-based survey with 139 participants [2] enables us to get a broader understanding of threat concern and exposure.

We find that real word ML pipelines are often more complex than their academic counterparts [1], raising questions about the applicability of current results in practice. This is particularly relevant in case of attack mitigations, in particular since our studies support occurrences of direct attacks on AI systems in the wild: there are instances of both poisoning and evasion in practice [1, 2]. While neither attack seems frequent (yet), a third of the participants in our interview based study [1] expresses concern—yet at the same time ML security is often marginalized, in contrast to other security measures like access control or usage of cryptography. In terms of perception, we find that participants conflate concepts such as ML security and security that stems from other components of the system, for example, a circumvention of access control [1]. Another often conflated concepts are safety and security of a system, e.g. distinguishing whether a failure was caused by an attacker with malicious intent (security) or a benign failure, for example, due to incomplete training data (safety) [2]. Our work [2] also highlights the influence of (self-reported) prior knowledge in ML security, which leads to higher concern in all studied attacks. Finally, concern related to individual attacks is highly motivated by economic, performance, and even ethical factors [2]. Another concern that was often uttered is that ML is used to support decision making within the company, and attacks can thus alter decisions taken within a company [2].

References

1 Lukas Bieringer, Kathrin Grosse, Michael Backes, Battista Biggio and Katharina Krombholz, *Industrial practitioners' mental models of adversarial machine learning*. Eighteenth Symposium on Usable Privacy and Security (SOUPS), 97–116, 2022.

2 Kathrin Grosse, Lukas Bieringer, Tarek Richard Besold, Battista Biggio, Katharina Krombholz: "Why do so?" – A Practical Perspective on Machine Learning Security. CoRR abs/2207.05164, 2022.

3 Ram Shankar Siva Kumar, Magnus Nyström, John Lambert, Andrew Marshall, Mario Goertzel, Andi Comissoneru, Matt Swann, Sharon Xia, *Adversarial Machine Learning-Industry Perspectives. Security & Privacy Workshops, 69-75, 2020.*

4 Franziska Boenisch, Verena Battis, Nicolas Buchmann, Maija Poikela. *"I Never Thought About Securing My Machine Learning Systems": A Study of Security and Privacy Awareness of Machine Learning Practitioners*. Mensch und Computer, 520-546, 2021.

3.5 From Wild Patterns to Wild Networks: A New Threat Model for Adversarial Examples in 5G Networks

Pavel Laskov (Professor at the University of Liechtenstein – Vaduz, LI)

Joint work of Giovanni Apruzzese, Rodion Vladimirov, Aliya Tastemirova, Pavel Laskov
Main reference Giovanni Apruzzese, Rodion Vladimirov, Aliya Tastemirova, Pavel Laskov: "Wild Networks: Exposure of 5G Network Infrastructures to Adversarial Examples", IEEE Transactions on Network and Service Management, pp. 1–1, 2022.
URL https://doi.org/10.1109/TNSM.2022.3188930

5G networks must support billions of heterogeneous devices while guaranteeing op Quality of Service. Such requirements are impossible to meet with human effort alone, and Machine Learning (ML) represents a core asset in future 5G infrastructures. ML is known to be vulnerable to adversarial examples; however, classical threat models for adversarial attacks are not suitable for the complex 5G ecosystem. To illustrate the potential vulnerability of ML components in 5G infrastructures to adversarial example, we present a new threat model [1] specifically addressing the interplay between ML and network management tasks in the 5G infrastructures. The proposed "myopic" threat model outlines the attacker's goals, knowledge, capability and strategy which are specific to the application of ML in the 5G context. The model assumes, for example, that the attacker has unlimited capability to change the behavior of the user equipment (UE) at her disposal but limited visibility into the archtecture of ML componets in the overall 5G infrastructure. Furthmore, a myopic attacker does not even have an oracle access to the predictions of a ML system. The only feedback about ML system's decision an attacker may receive is the change in the behavior of infrastructural components which her UE communicates, e.g., signals received over the radio interface. We evaluate the attacks conceivable under the myopic threat model on 6 applications of ML envisioned in 5G. Such attacks may affect both the training and the inference stages, can degrade the performance of state-of-the-art ML systems in terms of traditional metrics, such as accuracy or mean squared error, as well as in terms of physical quality metrics, for example, spectral efficiency. Given that myopic attacks have a lower entry barrier in comparison with attacks using previous threat models, further investigation of technical and operational impact of such attacks is indispensable.

References

1 Giovanni Apruzzese, Rodion Vladimirov, Aliya Tastemirova, and Pavel Laskov. "*Wild Networks: Exposure of 5G Network Infrastructures to Adversarial Examples*." IEEE Transactions on Network and Service Management, 2022.

3.6 Security and Privacy in Federated learning: Challenges and Possible Solutions

Mitrokotsa Aikaterini (Professor at the University of St. Gallen, CH)

Joint work of Georgia Tsaloli, Bei Liang, Carlo Brunetta, Gustavo Banegas

Mobile phones, wearables, autonomous vehicles and in general Internet of Things (IoT) devices are just some examples of distributed networks that create a wealth of data every day. This data is subsequently used as input in centralised machine learning models in order to

achieve reliable user modeling and personalisation. The growing storage and computational power of mobile devices as well as increased privacy concerns have led to an increased interest in federated learning, which allows multiple clients to collaboratively train learning models under the orchestration of a central server, while the data remain located on the sources.

Distributed machine learning has many significant advantages compared to centralised machine learning, mainly regarding efficiency and privacy. However, some serious challenges remain:

- Privacy: Although only updates are sent to the server, research has shown that these updates may still leak sensitive information, thus, providing no formal guarantee of privacy. For instance, by having access to a gradient update and the previous model, it might be possible to infer a training example.
- Security: The central server represents a single point of failure or even a bottleneck. How can a client be sure that the server has performed the aggregation correctly? A "lazy" server might use a simpler model to reduce its computational load, or modify the aggregation result to bias the model.
- Heterogeneity:The federated learning process is massively parallel involving multiple clients (up to 10^10) with different resources/capabilities. Many of these devices (5% or more) will fail or drop (being controlled by different clients), creating thus, a highly stateless environment.

In this talk, we discuss the main security and privacy challenges in federated learning as well as how we may guarantee and secure and private dynamic aggregation of data [1, 2] which can be employed in the federated learning setting. More precisely, we discuss how by relying on verifiable homomorphic secret sharing, we can achieve secure and verifiable aggregation of multiple users' secret data (e.g., parameters of the learning model), while employing multiple untrusted servers. The proposed solutions compute the sum of the users' input and provides public verifiability, i.e., anyone can be convinced about the correctness of the aggregated sum computed from a threshold amount of servers, while no communication between the users occurs.

References

1 Georgia Tsaloli, Bei Liang, Carlo Brunetta, Gustavo Banegas and Aikaterini Mitrokotsa. *DEVA: Decentralized, Verifiable Secure Aggregation for Privacy-Preserving Learning*. Proceedings of the 24th International Conference on Information Security (ISC) 2021, Nov. 10-12, 2021.

2 Carlo Brunetta, Georgia Tsaloli, Bei Liang, Gustavo Banegas and Aikaterini Mitrokotsa. *Non-interactive, Secure Verifiable Aggregation for Decentralized, Privacy-Preserving Learning*. Proceedings of the 26th Australasian Conference on Information Security and Privacy (ACISP 2021), Dec. 1-3, 2021.

3.7 Entering the Cursed World of Explainable Machine Learning

Konrad Rieck (Professor at TU Braunschweig, DE)

Machine learning is increasingly used as a building block of security systems. Unfortunately, most learning models are hard to interpret and typically opaque to practitioners. The machine learning community has started to address this problem by developing methods for

explaining the predictions of learning models, often coined "explainable AI". While several of these approaches have been successfully applied in computer vision, their application in security has received little attention so far.

In principle, explanation methods for machine learning promise to open the black box of learning models. They are considered one of the key enablers for using learning in security systems, as they allow to make the decisions of learning models transparent to practitioners. Rumor has it, however, that explanation methods are cursed and bring dangerous spells upon its users in computer security. In this talk, we learn that explanations are plagued by three problems: inconsistency, infidelity, and insecurity.

- First, various concepts exist for explaining learning models so that the same prediction can be described through different, often conflicting parts of the input.
- Second, several explanation methods tend to focus on properties of the data rather than the learning models. As a result, even random models may attain seemingly reasonable explanations.
- Third, explanations open a new attack surface for adversaries. Instead of gaining trust in a decision, we thus may be fooled twice – by a manipulated prediction and a manipulated explanation.

In the end, the security users may know less about a learning model than they did before. Lifting these spells is a journey yet to make and the basis for discussion at the seminar.

3.8 A semantic gap in malware analysis

Nedim Šrndić (Researcher at Huawei Technologies – München, DE)

Applications of machine learning begin long before model training. Being based on data, they start where the data starts: when a physical phenomenon is observed. A well-known example are the surroundings of an autonomous robot. They can be observed using physical sensors that capture images, audio, depth, touch, gravity, acceleration, smell, etc. In security, specifically in malware analysis, a common scenario is the execution of programs on end-point devices. This phenomenon is usually observed using software sensors which capture properties of programs or their behavior, e.g., executable files or behavior traces.

Once the target phenomenon is observed, the collected data is optionally pre-processed and then the machine learning model is applied. After optional post-processing the final prediction is made.

But what if the *observed phenomenon* is just a proxy for the *target phenomenon*, i.e., the one we are interested in understanding? In this talk I describe just such a situation in the field of malware analysis. In malware detection, we are interested if a program is malware or not. A program may be considered malware if it intentionally performs a harmful behavior in order to take advantage of its host system. Thus, to detect malware means to observe this harmful behavior. The harmful behavior is the target phenomenon.

In static malware analysis the observed phenomenon is the executable file. Results in program analysis show us that the behavior of a program cannot be accurately and comprehensively deduced from its executable file. Thus we are faced with a *semantic gap* between the observation and real-world effect. Similarly, in dynamic malware analysis on a

sandbox, we observe the behavior of the program *as influenced by our sandbox* at run-time. Under such circumstances, the program may conceal its true intentions. Executed on a legitimate endpoint, the same program may behave differently, and perform its harmful behavior.

In the talk, I underline that the security of the entire machine learning application crucially builds upon its first stage – data pre-processing. I show examples of the semantic gap and compare to the computer vision domain where the gap appears much smaller.

4 Working Groups

During a match-making session taking place in day 3, all interests expressed by the participants were consolidated into a set of working groups, addressing the following six areas:

- Machine learning security in the real world;
- Non-forgetting classifiers;
- Explainability and security;

The topics to be discussed in the break-out sessions shall be tailored to the interests of specific participants and will be chosen in an informal topic selection session during the seminar. The following topics, for example, are conceivable:

- Theoretical foundations. How should theoretical foundations of security of machine learning be built? What kind of theoretical guarantees can be expected and how do they differ from traditional theoretical instruments of machine learning?
- Machine learning as a methodical instrument of security. What requirements should be met for machine learning methods to be accepted as a reliable instrument in security operations?
- New applications. Most of research addressing security of machine learning uses computer vision and audio signal processing tasks as underlying applications and data. What other applications may be expected to face similar security challenges?
- Benchmark datasets. The existence of large benchmark tasks (e.g., ImageNet) is in many ways responsible for the success of deep learning. How can new benchmark datasets be created for further development of learning methods?
- Practical applications of secure learning. While there are a set of well-known examples where robust machine learning is important (e.g., autonomous vehicles), are there other non-security domains where robust machine learning would be useful?

4.1 Machine Learning Security in the Real World

Giovanni Apruzzese (PostDoc Researcher at the University of Liechtenstein – Vaduz, LI)
Antonio Emanuele Cinà (Ph.D. candidate at Ca' Foscari University of Venice, IT)
Katerina Mitrokotsa (Professor at the University of St. Gallen, CH)
Vitaly Shmatikov (Professor at Cornell University – Ithaca and Cornell Tech – New York, US)

This working group was tasked to identify some open challenges related to the real-world impact of research on ML security.

4.1.1 Discussed Problems

We began our activity by identifying some areas in which ML methods are (known to be) integrated into real-world applications. Then, we observed that most current research on ML security focuses only on a small subset of such areas. This is a significant shortcoming of our research: some areas (e.g., Computer Vision) are "inflated" with papers showing very effective attacks – thereby potentially over-emphasizing the problem; whereas other areas (which may be even more attractive for attackers, e.g., finance [1]) are under-investigated – thereby leaving dangerous blind spots.

Then, we reasoned why our research might not have a significant impact on the real world: indeed, several recent surveys revealed that practitioners seem not to care about the security of their ML models (e.g., [3, 4]). Our conclusion is that this is due to research papers making simplistic assumptions, as the evaluations are carried out mostly on "benchmarks" (e.g., CIFAR, MNIST), which are far different from real-world deployments of ML.

4.1.2 Possible root-causes

We attempted to find explanations as to why research on ML security only focuses on (sometimes decades-old) benchmarks. We conjectured that this is mainly due to two causes.

First, most ML systems deployed in the real world are **not open**, and researchers can hardly use them – at least to the extent necessary to derive scientific publications. The direct consequence is that researchers themselves are oblivious as to what the "attacked" ML system is actually doing (e.g., if the attack "works", is it because of a security issue of the ML model or because of another faulty component of the overall system?).

Second, performing experiments on real systems may have an **unfavorable "cost/benefit"** ratio – from a research perspective. Indeed:

- (*high cost*) using real systems for research purposes without the explicit consent of the developers of such systems may lead to legal problems when the researchers disseminate their findings; or it may lead to any progress being suddenly "voided" (i.e., before the researchers can conclude their experiments) because the ML system is naturally updated by its developers. Conversely, acquiring permission to use such systems is incredibly hard for researchers, as it may take months of communication with the owners of such systems (if they respond) and signing of NDA.
- (*low benefit*) a paper announcing a security vulnerability of a real system may not get much attention in research (i.e., few citations, e.g., [2, 5, 6, 7]—all having less than 60 citations as of July 2022). This is because the system will be patched, thereby preventing future works from reproducing the experiments and attempting to "outperform" the previous attack (unless the "vulnerable" ML system is made publicly accessible).

4.1.3 Conclusions and Recommendation

Most research on the security of Machine Learning overlooks many real-world applications of ML, and the corresponding evaluations mostly entail benchmarks. Such tunnel-visioning leads to practitioners in ML security to be confused about the real-world value of our research.

To improve the real-world impact of research on ML security, we advocate better cooperation between researchers and practitioners. Despite both sides ultimately having the same goal (i.e., improving the security of ML systems), such a goal is difficult to achieve if they keep working independently.

References

1 Dixon, Matthew F., Igor Halperin, and Paul Bilokon. "*Machine learning in Finance*." Vol. 1406. New York, NY, USA: Springer International Publishing, 2020.

2 Liang, Bin, Miaoqiang Su, Wei You, Wenchang Shi, and Gang Yang. "*Cracking classifiers for evasion: A case study on the Google's phishing pages filter*." In Proceedings of the 25th International Conference on World Wide Web, pp. 345-356. 2016.

3 Kumar, Ram Shankar Siva, Magnus Nyström, John Lambert, Andrew Marshall, Mario Goertzel, Andi Comissoneru, Matt Swann, and Sharon Xia. "*Adversarial machine learning-industry perspectives*." In IEEE Security and Privacy Workshops (SPW), 2020.

4 Boenisch, Franziska, Verena Battis, Nicolas Buchmann, and Maija Poikela. *"I Never Thought About Securing My Machine Learning Systems": A Study of Security and Privacy Awareness of Machine Learning Practitioners*. In Mensch und Computer, 2021.

5 Hosseini, Hossein, Baicen Xiao, Andrew Clark, and Radha Poovendran. "*Attacking automatic video analysis algorithms: A case study of Google Cloud video intelligence API*." In Proceedings of the Workshop on Multimedia Privacy and Security (CCS Workshop), 2017.

6 Li, Juncheng, Shuhui Qu, Xinjian Li, Joseph Szurley, J. Zico Kolter, and Florian Metze. "*Adversarial music: Real world audio adversary against wake-word detection system*." Advances in Neural Information Processing Systems 32 (2019).

7 Pajola, Luca, and Mauro Conti. "*Fall of Giants: How popular text-based MLaaS fall against a simple evasion attack*." In IEEE European Symposium on Security and Privacy, 2021.

4.2 Non-forgetting Classifiers

Lea Schönherr (CISPA Helmholtz Center for Information Security – Saarbrücken, DE)
Thorsten Eisenhofer (Ruhr University Bochum, DE)
Maura Pintor (University of Cagliari, IT)
Battista Biggio (University of Cagliari, IT)

4.2.1 Discussed Problems

For enhanced malware detection, machine-learning-based approaches are often applied to learn from the distribution of a multitude of samples. In the everlasting cat-and-mouse game, attackers might use blind spots of the classifier's learned distribution to camouflage their malware as benign. These artificial modifications cause a distribution shift of malicious samples, causing, over time, a drop in the models' performances. The classifier has thus to be updated to perform well on malicious samples taken from unseen distributions. Retraining with standard techniques is costly and requires saving all past and future data points for an indefinite time. Fine-tuning on new data samples has been shown to make the classifier forget about older training samples, also known as catastrophic forgetting [2]. The computational overhead and the forgetting of old data make both alternatives not applicable for data with a constant distribution shift [2].

An alternative approach would be continual learning, which enables learning from sequential data without forgetting samples from the past [1]. Here the explicit target is to retain good performances on the old data and achieve good performances on the new test data without storing all the training data points.

Existing approaches can be divided into three main categories: structural, functional, and architectural methonds. Structural methods utilize regularization terms that ensure the learned distribution does not shift arbitrarily while retaining knowledge on the old

distribution [2]. In the case of functional methods, new distributions are added as new outputs along the classifier's lifetime and can be added at any training step without changing the parameters of the old model [3]. Finally, architectural methods modify a model's architecture to learn new feature representations and outputs simultaneously [4].

4.2.2 Possible Approaches

All the proposed methods try to mitigate the issue with heuristics and are only evaluated empirically. Additionally, the methods are more focused on the occurrence of new classes and not on the adversarial nature of the problem, especially when focusing on applications such as malware detection, where there might be malicious parties exploiting blind spots in the learned distribution. All these problems open the question of whether the proposed techniques are near-optimal when applied to tractable cases. An optimal continual learning approach would perform equally with respect to an oracle-learning algorithm if they produce the same model, if trained incrementally, as when learning from the full training set.

Another issue is that current continual-learning approaches only care about maximizing or retaining average accuracy. However, this aggregate metric does not care how individual predictions are treated. While average accuracy may improve over time, even on previous tasks, some of the previously correctly-classified samples may be misclassified by the updated model, introducing the problem of model regression [5]. In addition, in the case of malware detection, it is more important to remember malicious samples, as a false negative sample can cause much more harm than a false positive. This opens up a potential research direction that tries to understand whether methods that mitigate model regression and methods that avoid catastrophic forgetting might be compatible or help each other in satisfying both conditions.

4.2.3 Conclusions

Lifetime learning for security-critical applications like malware detection requires considering a malicious party actively trying to circumvent a trained classifier by leveraging blind spots. Existing methods only focus on empirical methods for general classification tasks or architectural changes that enable the classification of new classes during the model's lifetime. Specific security-related properties, like prioritization to prevent false negative results, are in general not considered. Also, comparing the results with a model trained from scratch is neglected and would enable defining an upper bound for the performances of these methods. A continual learning system for malware classification tasks has, therefore, to be tailored to such properties to build systems that remain reliable for future distribution shifts and attack vectors.

References

1 Raia Hadsell, Dushyant Rao, Andrei A Rusu, and Razvan Pascanu. Embracing change: Continual learning in deep neural networks. Trends in cognitive sciences, 24(12):1028–1040, 2020.

2 James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 114(13):3521–3526, 2017.

3 Zhizhong Li and Derek Hoiem. Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence, 40(12):2935–2947, 2017.

4 Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. arXiv preprint arXiv:1606.04671, 2016.
5 Sijie Yan, Yuanjun Xiong, Kaustav Kundu, Shuo Yang, Siqi Deng, Meng Wang, Wei Xia, and Stefano Soatto. Positive-congruent training: Towards regression-free model updates. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14299–14308, 2021.

4.3 Explainability and Security

Asia Fischer (Ruhr-Universität Bochum, DE)
Kathrin Grosse (University of Cagliari, IT)
Nicola Paoletti (King's College London, UK)
Fabio Pierazzi (King's College London, UK)
Konrad Rieck (TU Braunschweig, DE)

This working group focused on the relationship between literature on explainable machine learning and cybersecurity. The emphasis was on local explainability methods, a popular class of methods that are concerned with explaining a model's predictions on individual inputs rather than the entire model.

4.3.1 Discussed Problems

In the following, we outline three major research gaps that complicate the applicability of explanations in security. Firstly, there is a semantic gap between what a human expects, and what the machine learns. While for images it is relatively clear for a human which pixels should be highlighted for a particular task, in the security context this becomes much less clear. A human may expect high-level relationships instead of low-level importance of which bytes/features contributed more to the classification. Also, there are cases when aligning explanations to human's expectations is not desirable: if the model uses shortcuts (e.g., spurious features) for its prediction, then such shortcuts should be exposed by the explanation, thereby identifying features that do not match human's intuition.

Secondly, it remains unclear how an explanation should behave in presence of malicious inputs (e.g., arising with evasion attacks on machine learning). On one hand, one would expect small changes in the explanation under adversarial perturbations that lead to small changes to the predicted class likelihoods. On the other hand, such small perturbations are often sufficient to flip the predicted class (especially if the input is close to the decision boundary), and hence, at it might be desirable for the explanation to change more drastically to "expose" the successful attack. However, some classes of explainability methods have been shown to be vulnerable to adversarial attacks as well [4] (i.e., adversarial inputs that change the model's decision but leave the explanation unchanged), which further complicates the understanding of their trustworthiness.

Finally, in security and privacy, the domain has a strong influence on the desired explanations. For example, assessing the trustworthiness of medical imaging tasks has very different implications and requirements than in malware detection.

4.3.2 Possible Approaches

To address the previously discussed issues and enable explainability for security, we suggest the following properties which should be addressed by newly designed methods in the area.

- **Completeness**: The method must produce an explanation for any given input and model. Blind spots need to be eliminated.
- **Stability**: For a given input and a given model, the method always produces the same explanation[1].
- **Descriptive Accuracy**: It answers to the question whether the identified features are actually important for the given task. If the important features are dropped, we should expect also a large accuracy drop.
- **Class-specificity**: In the security domain, benign and malicious inputs may depend on different sets of features or explanations.
- **Baseline**: Random data/models should generate "uniform" explanations. This property is inspired by results from Adebayo et al. [1], who show that some explanation methods act similarly to "edge detection" methods even in presence of untrained (i.e., randomized) models.
- **Smoothness**: Most work on smoothness focuses on inputs: small/large changes in inputs has small/large change on explanations. Our idea is to have smooth explanations *with respect to the output*: small/large changes in output have small/large changes on explanations. In this way, adversarial manipulations of prediction/confidence should become visible.

4.3.3 Conclusions

Many explainability methods are not designed with security in mind. This entails a possible semantic gap for explanations in the security domain, possible security vulnerabilities (e.g., adversarial examples), but also specifics of the applications domain. Hence, we advocate for rethinking explainability properties specifically desired in the context of security and robustness. To this end, we proposed six requirements that should be fullfilled by explainability in security.

References

1 Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. Advances in neural information processing systems, 31, 2018.
2 Emanuele La Malfa, Agnieszka Zbrzezny, Rhiannon Michelmore, Nicola Paoletti, and Marta Kwiatkowska. On guaranteed optimal robust explanations for NLP models. In International Joint Conference on Artificial Intelligence (IJCAI 2021), pages 2658–2665, 2021.
3 Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?" explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 1135–1144, 2016.
4 Xinyang Zhang, Ningfei Wang, Hua Shen, Shouling Ji, Xiapu Luo, and Ting Wang. Interpretable deep learning under fire. In 29th USENIX Security Symposium (USENIX Security 20), 2020.

[1] This seems an obvious requirement but there are state-of-the-art techniques relying on Monte-Carlo sampling [3], resulting in random explanations. Also, other methods admit multiple admissible explanation for the same model-input pair [2].

4.4 Letting attackers pay for the beer

Pavel Laskov (Universität Liechtenstein – Vaduz, LI)
Nicholas Carlini (Google, Mountain View, US)
David Freeman (Facebook, Menlo Park, US)
Kevin Alejandro Roundy (NortonLifeLock, Culver City, US)
Wieland Brendel (Max Planck Institute for Intelligent Systems, Tübingen, DE)

This working group focused on the potential and problems for complexity measures of adversarial attacks in robust machine learning.

4.4.1 Background

There are only a handful of model defenses that are able to evade attacks in a white-box setting substantially better than undefended models. Even attacks against defended models, however, still often find perturbations that are adversarial from a human point of view.

In many practical cybersecurity scenarios, however, the more relevant question is not whether an attacker can succeed in principal if the attacker gets infinite time and access to the model internals. Instead, in real-world settings the more relevant question is whether an attacker can succeed with only black-box access to the system and within a certain finite time window or cost budget.

Without a cost budget, the problem condenses down to model stealing, i.e., using black-box queries to reconstruct the white-box model. Then the attacker can just use (typically cheap and efficient) white-box attacks to craft adversarial examples. However, so far the number of samples needed to reconstruct a model is very high.

The group is unaware of a closer analysis as to whether there exist defenses that are broken in the white-box setting but are still costly to fool in a black-box setting. The general assumption, however, was that it should be relatively straight-forward to evade decision-based attacks by adding intrinsic noise into existing models.

In part, the question of attack complexity is yet mostly unexplored because no metric for attack complexity exists that is commonly agreed on. There are several directions in which costs can be measured, starting from perturbation budget, compute costs and others (see next subsection). To provide a suitable measure of progress, however, the community needs to agree on a common metric and setting to measure how much a certain intervention raises the attack costs.

4.4.2 How to quantify attack complexity

There are various dimensions and considerations to take into account for measures of attack complexity, some of which very much depend on the boundary conditions of the cyber security setting in consideration. A major goal of this working group was to collect and shed light on these different dimensions:

- **Number of queries**: The average or median number of queries needed to fool a model could serve as a simple baseline measure for attack complexity.
- **Cost to attack**: An attacker might be able to trade queries with more local processing (e.g., computing surrogate gradients in a local white-box model). Taking this cost into account would be interesting to consider the overall attack costs, but there were concerns how to quantify the attacker costs reliably.

- **Cost to evade**: Likewise, a defender might be able to evade attacks by certain measures, e.g., by averaging across several noisy inputs, employing additional detector methods, tracking inputs to evade finite-gradient methods, and others. These costs could be taken into account, especially if they are incurred only under attack and not vanilla inputs.
- **Model knowledge**: While in standard attack evaluations one typically only considers white-box access, we are here interested in the more relevant black-box scenarios where the attacker only receives partial information about the model like the final decision or the top-5 confidence scores, but not the actual gradients. The exact amount of information can greatly influence, e.g., the number of queries need for a successful attack, but are subject to the specific cybersecurity scenario to be considered.
- **Perturbation bound**: Just as in standard attack evaluations, it is important to specify perturbation bounds under which an attacker is allowed to operate (like L-infinity bounds for image classifiers). Typically, these bounds are chosen such that a certain function is achieved (like keeping human classification).
- **Query Access**: If accounts get rate-limited and/or if accounts are banned if too many slightly perturbed versions of the same input are submitted (hinting to a finite-difference attack), then the attacker might be forced to open many accounts to receive the necessary amount of information to attack a model. In this case, the defence problem reduces down to a standard fake account problem.
- **Attack goal**: Another important consideration is the attack goal, namely whether the attacker wants to evade the classifier for a given sample (e.g., copyright for a certain movie file), or whether the attacker just needs to evade on some samples (e.g., to post nude images in NSFW channels). In the first case, the mean or median complexity across a range of given samples is relevant, while in the latter case only the complexity of a lower percentile is the relevant metric.
- **Theoretical bounds**: Finally, there is an open question whether theoretical bounds for defenses can be derived with respect to the minimal costs (e.g., number of queries) for the attacker to craft an adversarial example.

4.4.3 Conclusions

There is no universally relevant complexity measure for adversarial attacks. However, the discussion converged to the general consensus that the number of queries an attack needs on average/median on a given sample will be a useful measure in most scenarios. Other potential components like the cost to evade or attack of typically hard to quantify and should thus be avoided unless absolutely needed in a certain real-world criterion.

A suitable topic for future research is whether existing attacks with small modifications (like some intrinsic noise, only decision-based access, maybe some memorization for nearest neighbour inputs in the past) are already enough to make existing black-box attacks close to impossible. The group agreed that this direction needs further investigation.

Participants

- Hyrum Anderson
Robust Intelligence – San Francisco, US
- Giovanni Apruzzese
Universität Liechtenstein – Vaduz, LI
- Verena Battis
Fraunhofer SIT – Darmstadt, DE
- Battista Biggio
University of Cagliari, IT
- Wieland Brendel
Universität Tübingen, DE
- Nicholas Carlini
Google – Mountain View, US
- Antonio Emanuele Cinà
University of Venice, IT
- Thorsten Eisenhofer
Ruhr-Universität Bochum, DE
- Asja Fischer
Ruhr-Universität Bochum, DE
- Marc Fischer
ETH Zürich, CH
- David Freeman
Facebook – Menlo Park, US
- Kathrin Grosse
University of Cagliari, IT
- Pavel Laskov
Universität Liechtenstein – Vaduz, LI
- Aikaterini Mitrokotsa
Universität St. Gallen, CH
- Seyed Mohsen Moosavi-Dezfooli
Imperial College London, GB
- Nicola Paoletti
Royal Holloway, University of London, GB
- Giancarlo Pellegrino
CISPA – Saarbrücken, DE
- Fabio Pierazzi
King's College London, GB
- Maura Pintor
University of Cagliari, IT
- Konrad Rieck
TU Braunschweig, DE
- Kevin Alejandro Roundy
NortonLifeLock – Culver City, US
- Lea Schönherr
CISPA – Saarbrücken, DE
- Vitaly Shmatikov
Cornell Tech – New York, US
- Nedim Srndic
Huawei Technologies – München, DE

Report from Dagstuhl Seminar 22282

Current and Future Challenges in Knowledge Representation and Reasoning

James P. Delgrande*[1], Birte Glimm*[2], Thomas Meyer*[3], Miroslaw Truszczynski*[4], Milene Santos Teixeira†[5], and Frank Wolter*[6]

1 **Simon Fraser University – Burnaby, CA.** jim@cs.sfu.ca
2 **Universität Ulm, DE.** birte.glimm@uni-ulm.de
3 **University of Cape Town, ZA.** tmeyer@cs.uct.ac.za
4 **University of Kentucky – Lexington, US.** mirek@cs.uky.edu
5 **Universität Ulm, DE.** milene.santos-teixeira@uni-ulm.de
6 **University of Liverpool, GB.** wolter@liverpool.ac.uk

Abstract

The area of Knowledge Representation and Reasoning (KR) is a central area in Artificial Intelligence that deals with the explicit, declarative representation of knowledge along with inference procedures for deriving further, implicit information from this knowledge. The goal of this Perspectives Seminar was to assess the area of KR, including its history, current state, and future prospects, and from this assessment to provide suggestions and recommendations for advancing the field, increasing participation in the area, and furthering links with related areas. Over the course of 5 days, 25 participants from a cross-section of subareas in KR and areas adjacent to KR met to discuss these topics. The workshop was composed of a number of invited talks and panels for reviewing the history and state of the art of KR, along with several working groups and general open discussions. In common with other Perspectives Workshops, a Manifesto will be produced; as well, recommendations contained in the manifesto will be also forwarded to the steering committee of the Principles of Knowledge Representation and Reasoning conference series for their consideration.

Seminar July 10–15, 2022 – http://www.dagstuhl.de/22282
2012 ACM Subject Classification Theory of computation → Complexity theory and logic; Computing methodologies → Knowledge representation and reasoning
Keywords and phrases applications of logics, declarative representations, formal logic, knowledge representation and reasoning
Digital Object Identifier 10.4230/DagRep.12.7.62

* Editor / Organizer
† Editorial Assistant / Collector

Current and Future Challenges in Knowledge Representation and Reasoning, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 62–79
Editors: James P. Delgrande, Birte Glimm, Thomas Meyer, Miroslaw Truszczynski, Milene S. Teixeira, and Frank Wolter

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

James P. Delgrande (Simon Fraser University – Burnaby, CA)
Birte Glimm (Universität Ulm, DE)
Thomas Meyer (University of Cape Town, ZA)
Mirek Truszczynski (University of Kentucky – Lexington, US)
Milene Santos Teixeira (Universität Ulm, DE)
Frank Wolter (University of Liverpool, GB)

Knowledge Representation and Reasoning (KR) is the field of Artificial Intelligence (AI) that deals with explicit, declarative representations of knowledge along with inference procedures for deriving further, implicit information from these symbolic representations. Research in KR as a mature area of AI is commonly taken as being marked by an Artificial Intelligence Journal Special Issue on Nonmonotonic Reasoning in 1980. In 1989 the Principles of Knowledge Representation and Reasoning Conference was founded, providing a dedicated, specialised forum for research in the area. While KR is one of the oldest and best-established areas of AI, it has continued to grow and thrive over the years. Most of the original research areas have evolved significantly, and have matured from the discovery and exploration of foundations, to the development and analysis of systems for emerging or established applications. Yet other areas, such as argumentation, arose much more recently, and are now thriving areas of KR.

While progress in KR has been steady and often impressive, it has not kept pace with the recent significant successes in AI in the use of statistical techniques and machine learning (ML). As a result, much of the work in AI, and much of the public perception of AI, centres on machine learning and on statistical applications. Nonetheless, we take it as given that KR is a vital, essential area of AI, and that research and development in KR remains necessary. Indeed, despite the unquestionable successes in machine learning and statistical techniques, limitations of these approaches are now emerging that, we believe, can only be overcome with advances in KR. Indicative of this is the recent interest in "Explainable AI", which requires a reference to declarative structures and reasoning over such structures. Furthermore, and in common with the majority opinion in AI, cognitive science, and philosophy, we take it as given that symbolic, declarative representations of knowledge are essential for any ultimate, general theory of intelligence.

For all of these reasons, a reassessment of the area of Knowledge Representation was a very timely undertaking of the Dagstuhl Perspectives Workshop 22282 "Current and Future Challenges in Knowledge Representation and Reasoning". During the seminar, the participants assessed the current state of KR along with future trends and developments. A questionnaire, which had been earlier distributed to the participants, helped in this assessment. Alltogether, the seminar served as a basis for developing an innovative agenda for the next 10-20 years of KR research. Key findings are measures to support a synergistic relationship with other subareas of the rapidly-changing field of AI and of computer science as a whole, e.g, through tutorials at the major KR conference, through new conference tracks and updated reviewing guidelines. The seminar further identified research areas for emphasis, assessed prospects for practical application of techniques, and considered how KR may address limitations of statistical techniques and machine learning.

The program comprised invited talks, panel discussions, working groups, and general dicussions. While the invited talks were agreed upon beforehand, the topics of the working groups (apart from Day 1) were decided interactively with all participants to allow for flexibility and reacting to the talks and the triggered discussions. Day 1 started with a short welcome and participant introduction session, followed by an assessment of *the past and present of KR* in the form of two invited talks by Anthony Cohn and Thomas Eiter. The remainder of Day 1 was dedicated to presenting the questionnaire results, which also prepared for the first working group on rethinking the call for papers (CfP) for the main KR conference, which not only served as rethinking the CfP, but also steered the working groups into thinking about the definition of KR as an area. The day closed with a report from the four working groups and indeed identified changes for the CfP, but also for the track structure and the recruitment and instructions for reviewers.

Day 2 focussed on the relationships of KR with four neighboring areas. For each sub-areas we began with a short invited talk (20 min) followed by a commentary (5 min), also invited, and a short general discussion (5 min). The function of the commentator was to look at the area from a different angle or give another perspective to avoid a too personal or narrow a perspective. The four talks addressed ""KR and AI" (Ian Horrocks, commentator: Sébastien Konieczny), "KR and ML" (Francesca Toni, commentator: Ana Ozaki), "KR and Information Systems" (Diego Calvanese, commentator: Meghyn Bienvenu), and "KR and Robotics" (Gerhard Lakemeyer, commentator: Michael Beetz). Working groups on research challenges for these subareas concluded the day.

The third day began with a short talk on "Handling Uncertainty" (Jean Christoph Jung), for initiating a panel discussion on this topic. The morning concluded with a continuation of the working groups on sub-areas of KR from the previous day. The afternoon was dedicated to hiking and biking in smaller groups.

Day 4 started with short invited talks on "Applications of KR" (Esra Erdem, Thorsten Schaub, Michael Tielscher). The remainder of the day was dedicated to working groups on assessing the state of the art in sub-areas of KR and to expanding KR. For this latter group, we discussed the fact that geographically KR is stronger in Europe than in other parts of the world. As well, we considered how to attract new talent and how to reach out to disadvantaged groups, along with thinking of new forms of events such as hybrid conferences or virtual seminar series.

The final day of the seminar looked at strengthening the interaction between sub-areas of KR and wrapped up with statements of the participants regarding their personal impressions and "take-home" messages. This has, for example, already led to the creation of a novel KR discussion channel (on a Discord server). Key findings include that KR applications are very important to make the field visible and that applications are to be made more visible, e.g., through a journal special issue. Another outcome includes measures to reach out to other areas of AI, in particular machine learning and statistical techniques, where symbolic approaches can make contributions, e.g., for general intelligent agents. A separate Manifesto will provide an assessment of the area, and will give a set of recommendations regarding the future of KR and its promotion.

2 Table of Contents

3 Overview of Talks

3.1 An (Abbreviated, Partial) History of Knowledge Representation: A Personal Perspective

Anthony Cohn (University of Leeds, GB)

In this invited talk I give a brief, and selective history of KR, starting with Aristotle and focusing mostly on work between 1950 and 2000. I talk about Newell's Knowledge Level, the Physical Symbol Hypothesis and early theorem provers, and key advances such as the resolution rule of inference which then led to Prolog and also Datalog. I talk about Frames, and the "scruffy vs neat" debate in the 1970s. I mention the rise of description logics, the semantic web, constraint reasoning and the origins of non monotonic logic. I discuss the representation of particular kinds of knowledge, including taxonomic knowledge and qualitative representation and reasoning methods focusing on spatial knowledge. Finally I note the founding of KR Inc in 1993 and the series of very successful KR conferences it has run since 1989.

3.2 History of Knowledge Representation: A Personal View

Thomas Eiter (TU Wien, AT)

In this talk, I give a brief account of the history of KR, which however is done in a selective manner and in a subjective perspective in which some developments and results are cherry-picked. The presentation will go by decades, covering the time from the beginnings of AI in the 1950s up to the late decade, even if this view is not well-suited as there are overlaps and interesting developments may start or end a bit earlier, and similarly done or start a bit later. Special emphasis is given to the KR conference and the co-development of other venues and communities such as for logic programming, constraint satisfaction, and planning. The talk ends with a status-quo assessment and an outlook on future challenges.

3.3 Knowledge Representation and Artificial Intelligence

Ian Horrocks (University of Oxford, GB)

Main reference Yavor Nenov, Robert Piro, Boris Motik, Ian Horrocks, Zhe Wu, Jay Banerjee: "RDFox: A Highly-Scalable RDF Store", in Proc. of the The Semantic Web – ISWC 2015 – 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 9367, pp. 3–20, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-25010-6_1

KR was central to early work on AI, e.g., at the famous Dartmouth Conference, where several of the well-known protagonists made important contributions to KR. Expert systems, a subsequent high-profile development in AI, is also closely linked to KR. Although there have

been ups and downs in the intervening years, we now have practical KR systems that are used in important applications, particularly in bio-health. An interesting recent (ish) phenomenon is the development of large-scale knowledge graphs (KGs) such as the Google KG. Such KGs are now pervasive in search, e-commerce, and personal assistants (Alexa, Siri, etc.). This success has encouraged the development of general-purpose KG systems and these are now being successfully applied in areas such as industrial design and configuration. So why are KR systems "suddenly" so successful? At least in part due to (i) increasing processing power; (ii) advances in theory and algorithms; (iii) availability of data and digitisation. Of course, many challenges remain, not least knowledge creation and curation – these are still hard problems!

3.4 Knowledge Representation and Machine Learning

Francesca Toni (Imperial College London, GB)

Machine Learning (ML) has grown massively in the last 10 years or so, predominantly due to increased processing power availability, big data, and powerful statistical and probabilistic models. ML is predominantly data-driven these days, with the ambition to automate reasoning as a vector, rather than a manipulator. In this talk, I have expressed the role that KR may have in ML, as well as (to a lower extent) the role that ML may have to KR. I have identified the need for ML to be verified and explained, so as to identify any artifacts and biases that may be present in ML models. KR can (and does) play an important role to support ML. Also, KR can contribute to 'hybrid ML models, integrating reasoning components with statistical/neural ML, as in inductive logic programming. KR-based verification/explanation methods and hybrid systems are all examples of how KR can support ML. On the other hand, ML can help KR with knowledge elicitation (e.g. for argumentation or knowledge graphs). Overall, ML is an important area of AI research and the KR community can gain lots from joining forces with the ML community.

3.5 Knowledge Representation and Information Systems

Diego Calvanese (Free University of Bozen-Bolzano, IT)

URL http://www.inf.unibz.it/ calvanese/presentations/2022-KR-IS-calvanese.pdf

An Information System (IS) is an integration of components for collecting, storing, and processing data, where the data is used to provide information and contribute to knowledge and digital products that facilitate decision making. This definition illustrates that the key elements that an IS has to deal with are data, knowledge, and processes operating over them. These are exactly the elements Knowledge Representation and Reasoning (KRR) has been concerned with, by studying semantic and computational aspects, developing techniques, and building tools. In this presentation we start by providing a brief overview of the main formalisms, techniques, and tools that have been proposed in KRR (notably, based on lightweight Description Logics) to address static and structural aspects of data

and knowledge as they are encountered in ISs. We then move to dynamic and temporal aspects, which are related to the evolution and change over time of data and knowledge due to processes that operate over them. We discuss proposals and results on the integrated modeling and reasoning over data, knowledge, and processes, which is a major concern in ISs. We conclude by presenting some challenges that this poses for KRR.

3.6 Knowledge Representation and Robotics

Gerhard Lakemeyer (RWTH Aachen, DE)

While KR was central to building autonomous robots initially, as demonstrated by the robot Shakey in the late sixties, the field diverged, and KR techniques have started to play a role starting in the late nineties, with good examples being the museum tour guides Rhino and Minerva. One of the most addressed KR frameworks for robots is the KnowRob system developed by Tenorth and Beetz, combining rich ontologies with specialized reasoning. Planning techniques also play a major role, starting with domain-dependent planners like IxTeT and TAL, and later, domain-independent planners like FF and TFD, followed by a combination of action programming languages like Golog and PDDL planners, task and motion planners as well as conditional planners. Task planning needs to be complemented by action execution and maintenance. For the diagnosis of failures model-based techniques with strong KR foundations have been developed. After touching on these issues, my talk ended by proposing the Robocup Logistic League, where a team of robots needs to dynamically assemble products with the help of machines, as a rich benchmark for research in KR and robotics.

3.7 Knowledge Representation and Uncertainty

Jean Christoph Jung (Universität Hildesheim, DE)

Uncertainty arises in many applications: it may come from incomplete information in games like Poker (or other adversarial scenarios), in the presence of random events (like dice rolls), and is a result of extracting information from text, images, audio, video. KR has early on recognized the need of dealing with uncertainty and proposed a range of solutions to various problems. In my talk, I survey the most important concepts that have been identified during the last 40 years (of course the choice is highly subjective): possible world semantics, probabilities, independence, graphical models, updates, and probabilistic and epistemic logic.

3.8 Applications of Knowledge Representation and Reasoning

Esra Erdem (Sabanci University – Istanbul, TR)

As the definition of AI changes towards building rational agents that are provably beneficial for humans, KR&R plays an important role in addressing the user-oriented challenges in applications that come along during this shift, such as generality, flexibility, provability, hybridity, bi-directional interactions, and explainability. In this talk, I present three applications underlining how KR&R addresses these challenges: to solve combinatorial search problems in cladistics, to address knowledge-intensive problems in bioinformatics, and to solve hybrid reasoning problems in robotics. I also discuss how evaluations of human-centric KR&R applications could be extended to include subjective quantitative/qualitative measures.

3.9 Knowledge-driven Artificial Intelligence

Torsten Schaub (Universität Potsdam, DE)

Knowledge plays a vital role in modern companies and organizations, be it in production, storage, or workforce management. Most crucial is the knowledge needed to accomplish creative processes, such as designing a product, an assembly line, a shift schedule or timetable, or planning a trip, routing vehicles, or diagnosing remote systems. All of these tasks involve taking knowledgeable decisions while respecting constraints and preferences. Answer Set Programming (ASP) has become a popular approach for modeling and solving such knowledge-intense combinatorial (optimization) problems. What makes ASP attractive is its combination of a declarative modeling language with highly effective solving engines. This allows us to concentrate on specifying – rather than programming the algorithm for solving – a problem at hand. The talk highlighted some current research topics, and concluded with an outlook on the ASP's potential impact as a knowledge-driven AI tool.

3.10 Application of Knowledge Representation and Reasoning

Michael Thielscher (UNSW – Sydney, AU)

Three systems with a clear KR&R component were presented: (1) A two-armed Blocksworld problem-solving robot with a high-level, symbolic (re-)planning component connected to a low-level robotic controller. (2) An interactive artwork with a BDI-based high-level agent programming component connected to a unity-based controller for virtual characters interacting with human users. (3) A general game-playing system that understands logic-based rule descriptions of new games and uses a propositional logic interface for reasoning connected to a neural network that, with the help of Monte Carlo tree search, learns to play any new game without human intervention.

4 Working groups

4.1 Rethinking the KR Call For Papers

James P. Delgrande (Simon Fraser University – Burnaby, CA), Birte Glimm (Universität Ulm, DE), Thomas Meyer (University of Cape Town, ZA), and Frank Wolter (University of Liverpool, GB)

Participants were divided in four different groups to discuss possible updates of the call for papers of KRR. Each group had 45 minutes to discuss between themselves and their conclusions were shared afterwards. The main points elicited were:

(i) Title of the conference: participants discussed the possibility of updating the conference title (e.g., replacing it by the more general "The international conference on knowledge representation and reasoning"). However, changing the name also has disadvantages since information might be lost (on google search engine, for example). An alternative is to keep the current official name and add the "new title" as a subtitle (as done by other conferences).

(ii) Definition of what KRR is: participants agreed that the main text introducing KRR needs updates. Some updates discussed at the Workshop have already been implemented in the Call for Papers for KR 2023.

(iii) Topics of interest: there are too many topics listed. An alternative is to elicit about 10 topics and add examples of subtopics (e.g. KRR and cognition (e.g. cognitive systems, cognitive reasoning...)). Just hiding all subtopics, however, might not be a good idea since less knowledgeable authors might not understand that their work does not fit the conference.

4.2 Research Challenges: Knowledge Representation and Robotics

Michael Beetz (Universität Bremen, DE), Anthony Cohn (University of Leeds, GB), Esra Erdem (Sabanci University – Istanbul, TR), Andreas Herzig (Paul Sabatier University – Toulouse, FR), Gerhard Lakemeyer (RWTH Aachen, DE), and Michael Thielscher (UNSW – Sydney, AU)

The working group emphasized that there were few submissions to KR and robotics 2022, there being only one paper accepted. The discussion on how it can be improved elicited a few alternatives: (i) besides a special track, introduce a workshop or summer school on the topic; (ii) inclusion of short papers; (iii) linking to a recently created European conference that integrates KR and robotics. Overall, the three main challenges identified by the participants were the following: (1) continuously learning about the world that is changing, i.e. the need for the identification of what KR issues are raised for a robot that works for long periods of time (not just setting a table, but a whole day task, for example); (ii) cognition: learn from demonstration, exchange information and knowledge given by human or available in ontologies; (iii) multi-agent path finding: a complete representation of everything going on robot to understand why they fail is necessary; there is still a lack of good solutions for this problem.

4.3 Research Challenges: Foundations of Knowledge Representation

James P. Delgrande (Simon Fraser University – Burnaby, CA), Marc Denecker (KU Leuven, BE), Sebastien Konieczny (University of Artois/CNRS – Lens, FR), and Thomas Meyer (University of Cape Town, ZA)

The participants of the work group agreed that KR is based on a sound foundational footing, but that many foundational questions remain open. An example that was mentioned is the case of a particular form of non-monotonic reasoning known as "rational closure" (or system Z). The question was raised whether this form of reasoning can really be described as rational. One possibility that was mooted is that of collaboration with psychologists to help with analyzing these theories. The participants also discussed the issue that, although there are many success stories of foundational KR being turned into applications, more ought to be done when it comes to being sufficiently successful with solving real-world problems. In summary, three main challenges were identified: (i) obtaining more precise models of different types of knowledge; (ii) the well-known bottleneck on knowledge acquisition; (iii) differentiating quantitative approaches to reasoning from qualitative approaches.

4.4 Research Challenges: Knowledge Representation and Machine Learning

Birte Glimm (Universität Ulm, DE), Ian Horrocks (University of Oxford, GB), Jean Christoph Jung (Universität Hildesheim, DE), Ana Ozaki (University of Bergen, NO), Steven Schockaert (Cardiff University, GB), and Francesca Toni (Imperial College London, GB)

The working group discussed the challenges regarding the integration of KR and ML. The participants highlighted that KR can contribute to ML in different ways; examples are through the manipulation of rewards or the injection of knowledge to speed up the learning process. However, it was brought to attention that it is still necessary to represent uncertainty on reasoning processes (probabilistic reasoning and conclusions) and insisting on "crisp" knowledge limits collaborations with other fields like ML. Participants agreed that, as a community, we need to be more open to these intersections and even promote them (e.g. special tracks, tutorials, summer schools, promoting a competition within KR&R). Finally, the three main challenges identified by the group were: (1) neuro-symbolic integration: dealing with uncertainty that comes from learning; (2) defining a benchmark/resource for a competition; (3) knowledge compilation: energy consumption, interpretability, identify which rule formalisms are feasible for each model.

4.5 Research Challenges: Knowledge Representation and Information Systems

Magdalena Ortiz (TU Wien, AT), Meghyn Bienvenu (University of Bordeaux, FR), Piero Andrea Bonatti (University of Naples, IT), Diego Calvanese (Free University of Bozen-Bolzano, IT), and Frank Wolter (University of Liverpool, GB)

The working group identified three concrete challenges within KR and information systems: (1) high-level descriptions and abstractions of dynamic data-centric systems and processes, and reasoning about them; (2) support for more comprehensive data management tasks in knowledge-enriched systems, such as analytical queries (numeric values, aggregation), updates, security and privacy, customization, and efficient system design; (3) better collaboration with communities like KGs, the semantic Web, and graph DBs. In particular, it is necessary to be aware and understand the standards of these communities so that we can leverage them to (a) make our techniques readily usable in practice, (b) identify and address their research challenges, and (c) increase the confidence in our solutions in different contexts. Regarding success stories in the field, we can highlight data integration (although several challenges still remain) and ontologies in biomedical domains. Medical ontologies are purpose-specific and hard to reuse, making it challenging their use in practice. One typically needs to extract some "relevant" knowledge, combine ontologies, match terms from different sources, etc., and the tools for such tasks are still underdeveloped. We lack easy-to-use tools that would allow developers to quickly use the existing knowledge. When it comes to processes with data, there are some theoretical results on temporal verification of dynamic data-centric systems, but still far from what is needed in practice. An example is the existence of a model-checker based on SMT, which still plays with toy examples, far from being deployed into real-world business processes. There are also some open challenges regarding privacy and security of information systems, namely (i) confidentiality: keeping the information secure and private (existing approaches are not secure and have many vulnerabilities); (ii) integrity: KBs should be correct (who inserts it? Is it trusted? Did I manage my ontology?); and (iii) availability: the KB should be reliable and not easy to crash.

4.6 Current Trends and Challenges in Knowledge Acquisition

Birte Glimm (Universität Ulm, DE), Steven Schockaert (Cardiff University, GB), and Francesca Toni (Imperial College London, GB)

Obtaining formalized knowledge is seen as a challenge that was discussed in this working group. The participants identified three main challenges regarding knowledge acquisition: (i) learning: it is possible to learn axioms, for example, but it still presents high uncertainty, (ii) maintenance: there are methods to develop ontologies, but not to maintain them, (iii) reasoning on (medical) guidelines, extracting and reasoning on rules is still challenging. Regarding success from last 5 years, the participants highlight: (i) a long running competition, (ii) learned graphs, (iii) availability of huge knowledge bases (fact bases, including certain levels of taxonomy), (iv) wide use of Protege, which offers a plugin architecture with successful built-in reasoners. Protege is not restricted to only the last 5 years, but it is still strongly relevant for the community.

4.7 Current Trends and Challenges in Description Logics

Magdalena Ortiz (TU Wien, AT), Meghyn Bienvenu (University of Bordeaux, FR), Piero Andrea Bonatti (University of Naples, IT), Diego Calvanese (Free University of Bozen-Bolzano, IT), Jean Christoph Jung (Universität Hildesheim, DE), and Frank Wolter (University of Liverpool, GB)

The participants discussed major trends and challenges in description logic research. For instance,

(i) Ontology-based data access has been a major research topic in description logic over the past 15 years. The development of novel and efficient query answering algorithms has been an important achievement. There are now powerful implemented tools for ontology-based data access that are used in various applications. Meaningful benchmarking for query answering remains an important challenge.

(ii) A comprehensive study of the complexity of query answering is another highlight. This includes novel alternative approaches to complexity such as non-uniform and parametrized complexity.

(iii) Making description logics nonmonotonic has been a major research topic with contributions ranging from the development of defeasible description logics to applying circumscription to description logics. Developing tool support remains a major challenge.

(iv) The development of support for ontology engineering has been another major area of research. Principled approaches to explanation, forgetting, repairing, modularity, and versioning have been suggested and investigated. Tools have been developed for some of these approaches.

(v) Explanation of entailments, abduction, and provenance have also been investigated in depth, with significant development of tool support in particular for explanation.

(vi) The close link between description logics and datalog/existential rules has triggered fruitful interaction with the database community; with results and techniques being transferred in both directions.

(vii) Temporal description logics have been investigated in depth over the past 20 years. Here implemented systems are still missing.

(viii) Inconsistency handling in ontology-based data access has been a major research area, again with significant interaction with the database community.

(xii) Adding features for handling uncertain knowledge to description logics has a long history and many approaches have been proposed and investigated. Tool support is still limited.

(xi) Areas that have recently attracted attention but in which challenging problems remain include description logic and learning, description logic and knowledge graphs, security and privacy of description logic ontologies.

4.8 Current Trends and Challenges in Reasoning about Action

Michael Beetz (Universität Bremen, DE), Anthony Cohn (University of Leeds, GB), Andreas Herzig (Paul Sabatier University – Toulouse, FR), Gerhard Lakemeyer (RWTH Aachen, DE), and Michael Thielscher (UNSW – Sydney, AU)

Participants discussed current trends and challenges regarding reasoning about action. Several trends were identified, including:
(i) epistemic planning
(ii) generalized planning
(iii) controller synthesis (e.g. using LTL, LTLf , timed automata, situation calculus)
(iv) expressive action logics with uncertainty, including regression and progression reasoning, and verification of belief programs
(v) ontologies for robots
(vi) causality
(vii) goal reasoning
(viii) learning action representations, and
(ix) KR techniques for informed reinforcement learning.

As well, various challenges were identified, including
(i) solvers for epistemic planners (the problem is undecidable in general)
(ii) learning representations of actions, affordances, and game rules from data
(iii) rational reconstruction of implemented KR systems for robots, like KnowRob, and
(iv) connections between KR work on causality and machine learning.

4.9 Expanding Knowledge Representation: Attracting People

Anthony Cohn (University of Leeds, GB), Thomas Eiter (TU Wien, AT), Gerhard Lakemeyer (RWTH Aachen, DE), and Michael Thielscher (UNSW – Sydney, AU)

Participants discussed alternatives to attract new talents to the community. Launching tangible practical challenges (e.g. angry birds, robotics), as done by other communities, is an alternative that would catch the attention of students. However, it was also highlighted that the process to attract new students must start early, with undergraduate students. It was identified that students might be interested in the field, but they do not have the necessary background. This way, it is important that universities keep elective courses on symbolic reasoning, for example. In alternative, providing high quality online material would support this lack of background. Finally, the participants discussed the creation of very short videos, possibly with testimonials, to motivate new students.

4.10 Expanding Knowledge Representation: Geographics

Esra Erdem (Sabanci University – Istanbul, TR), Ian Horrocks (University of Oxford, GB), Michael Thielscher (UNSW – Sydney, AU), and Frank Wolter (University of Liverpool, GB)

The group identified that the KR community is underrepresented in 3 important regions: North America, China, and southern Asia. Organising workshops addressing relevant KR topics (e.g. knowledge graphs) in North America could contribute to the promotion of the field in this area. Regarding China, participants identified that there are not many submissions from this area as KR is not ranked as a top conference there. The possibility of contacting acquaintances in China to discuss this issue and try an application for a re-ranking was considered. Finally, the participants agreed that to promote KR in southern Asia members of the KR community could offer to give tutorials and guest lectures in the area. Such tutorials and lectures would also be an opportunity to establish new collaborations in KR with researchers in the region.

4.11 Expanding Knowledge Representation: Other Event Types

Andreas Herzig (Paul Sabatier University – Toulouse, FR), James P. Delgrande (Simon Fraser University – Burnaby, CA), Birte Glimm (Universität Ulm, DE), Sebastien Konieczny (University of Artois/CNRS – Lens, FR), Torsten Schaub (Universität Potsdam, DE), and Steven Schockaert (Cardiff University, GB)

The group discussed several kinds of events differing from the standard format of the KR conference. Four recommendations are proposed.

First, we propose to reconsider holding KR every second year as an online event (either fully online or hybrid). Instead of simply copying the on-site format and having a one-week online event, we propose a series of sessions. This is inspired from successful recent experiences with thematic groups (e.g. the Online Social Choice and Welfare Seminar Series, `https://www.sites.google.com/view/2021onlinescwseminars`), projects (e.g. the EU TAILOR project), local groups advertising their online seminars (e.g. the LUCI Lunch Seminar Series, `https://luci.unimi.it/events/`), and national seminars (e.g. the French KR seminar, `https://www.gdria.fr/seminaire/`). Among the pros of this suggestion are: (i) no travel costs; (ii) good for the planet; (iii) it favors participation from developing countries, (iv) it is easier to attract people from adjacent fields such as roboticists (similar to colocation with another conference), and (v) talks can be recorded easily and cheaply. Among the cons, we list: (i) people from the KR field favor of on-site conferences; (ii) choosing the slot(s) will be problematic: there are several solutions but none is optimal, the merits of each of them should be weighed carefully; (iii) it may become apparent that there is too much heterogeneity in our field, e.g. only argumentation people attend argumentation sessions.

Second, we should try to establish benchmark and system competitions in areas where this makes sense. We might take inspiration from the NSF-funded StarExec platform (`https://www.starexec.org/starexec/public/about.jsp`) for first-order SAT solvers or the ICLP Prolog Programming Contests (`https://people.cs.kuleuven.be/~bart.demoen/PrologProgrammingContests/Contest99.html`).

Third, we propose to get more involved in summer schools such as: (i) ESSLLI (https://2022.esslli.eu/), (ii) NASSLLI (https://ml-la.github.io/nasslli2022/), (iii) the EurAI ACAI Summer School (eurai.org/acai), and (iv) the Reasoning Web Summer School (reasoningweb.org). Beyond KR people individually submitting courses, KR Inc. may propose to sponsor 2 or 3 courses.

Finally, KR Inc. might support tutorials and other small events such as schools in underrepresented countries in order to attract locals; local attendance can be expected to go beyond the KR field.

4.12 Expanding Knowledge Representation: Underrepresented Groups

Renata Wassermann (University of Sao Paulo, BR), Meghyn Bienvenu (University of Bordeaux, FR), Diego Calvanese (Free University of Bozen-Bolzano, IT), Jean Christoph Jung (Universität Hildesheim, DE), Thomas Meyer (University of Cape Town, ZA), Magdalena Ortiz (TU Wien, AT), and Ana Ozaki (University of Bergen, NO)

The group started the discussion with the questioning on whether a session to promote the role of women within KR should be held every year. Such a session results in lots of work for very specific participants and might not reach the expected goals. Alternatives for this session are presenting statistics (or a documentary) during the introduction session or displaying short videos during "food sessions". Next, participants also agreed that funding is necessary for inclusion, but it gets expensive for a single student. Therefore, when looking for sponsors for KR, a specific request for D&I funding can be launched. Finally, participants discussed the creation of a mentoring pool, where people from our community that are interested in supporting underrepresented students can act. Related activities include meeting online with students, discussing career prospects, what to work on, and where to publish.

5 Panel discussions

5.1 Current and Future Challenges in Knowledge Representation and Reasoning: Questionnaire Results

Birte Glimm (Universität Ulm, DE)

The panel session started with a presentation of results from a survey with 24 participants to identify key challenges, areas of increasing and decreasing importance in knowledge representation and reasoning and topics to which participants plan to make contributions. As key successes the participants identified existing industry-strength applications, in particular in the areas of answer set programming, description logics and ontologies, and knowledge graphs. In the following discussion it became clear that applications use KR techniques, but this use is often not very visible. In order to make KR applications more visible, a journal special issue on applications was proposed as well as special tracks with instructions

to reviewers as to how applications papers are to be reviewed were identified. A topic that is seen as very important is the area of hybrid AI, i.e., combining knowledge and learning-based methods. While progress in this direction is seen, this is an area that needs significant attention in the future to which at least some of the seminar participants want to contribute. To support this kind of interdisciplinary work, more tutorials and invited talks from the machine learning community can be incorporated into conferences and (summer) schools. Progress in this area was identified as crucial in order for KR to stay relevant as an area, while it was also clear that KR techniques have clear potential to advance intelligent systems, e.g., in terms of explainability or interpretability and when it comes to integrate existing knowledge. Finally, reaching out to neighboring areas is seen as very important as well as to actively showcase the KR successes and work on the still open challenges that so far hinder a wider adoption of KR techniques such as handling uncertainty or inconsistencies.

5.2 Interaction between Subareas

James P. Delgrande (Simon Fraser University – Burnaby, CA), Birte Glimm (Universität Ulm, DE), Thomas Meyer (University of Cape Town, ZA), and Frank Wolter (University of Liverpool, GB)

The final session of the seminar started with a discussion about the possibility of collocating KR with other conferences. For this, it would be interesting to look for papers that bridge gaps between fields (for whomever we collocate with). Participants also discussed the idea of having a session where anyone can "informally" present something in 3 minutes or, instead, researchers could submit an abstract and present it in 5 minutes. These works would be lightly refereed and would not be available in the KR proceedings, since their aim is exposure, e.g. to promote integrations between fields or propose new ideas. At the end of the session, the idea of creating a forum to instigate discussions and works in KR was launched. As a result, a Discord channel to announce talks and activities in the field was created by Ana Osaki and Renata Wassermann. The session concluded with individual feedback from each participant.

Participants

- Michael Beetz
Universität Bremen, DE
- Meghyn Bienvenu
University of Bordeaux, FR
- Piero Andrea Bonatti
University of Naples, IT
- Diego Calvanese
Free University of Bozen-Bolzano, IT
- Anthony Cohn
University of Leeds, GB
- James P. Delgrande
Simon Fraser University – Burnaby, CA
- Marc Denecker
KU Leuven, BE
- Thomas Eiter
TU Wien, AT
- Esra Erdem
Sabanci University – Istanbul, TR
- Birte Glimm
Universität Ulm, DE
- Andreas Herzig
Paul Sabatier University – Toulouse, FR
- Ian Horrocks
University of Oxford, GB
- Jean Christoph Jung
Universität Hildesheim, DE
- Sebastien Konieczny
University of Artois/CNRS – Lens, FR
- Gerhard Lakemeyer
RWTH Aachen, DE
- Thomas Meyer
University of Cape Town, ZA
- Magdalena Ortiz
TU Wien, AT
- Ana Ozaki
University of Bergen, NO
- Milene Santos Teixeira
Universität Ulm, DE
- Torsten Schaub
Universität Potsdam, DE
- Steven Schockaert
Cardiff University, GB
- Michael Thielscher
UNSW – Sydney, AU
- Francesca Toni
Imperial College London, GB
- Renata Wassermann
University of Sao Paulo, BR
- Frank Wolter
University of Liverpool, GB

Report from Dagstuhl Seminar 22291

Machine Learning and Logical Reasoning: The New Frontier

Sébastien Bardin*[1], Somesh Jha*[2], and Vijay Ganesh*[3]

1 **CEA LIST, FR.** sebastien.bardin@cea.fr
2 **University of Wisconsin-Madison, US.** jha@cs.wisc.edu
3 **University of Waterloo, CA.** vijay.ganesh@uwaterloo.ca

Abstract

Machine learning (ML) and logical reasoning have been the two key pillars of AI since its inception, and yet, there has been little interaction between these two sub-fields over the years. At the same time, each of them has been very influential in their own way. ML has revolutionized many sub-fields of AI including image recognition, language translation, and game playing, to name just a few. Independently, the field of logical reasoning (e.g., SAT/SMT/CP/first-order solvers and knowledge representation) has been equally impactful in many contexts in software engineering, verification, security, AI, and mathematics. Despite this progress, there are new problems, as well as opportunities, on the horizon that seem solvable only via a combination of ML and logic.

One such problem that requires one to consider combinations of logic and ML is the question of reliability, robustness, and security of ML models. For example, in recent years, many adversarial attacks against ML models have been developed, demonstrating their extraordinary brittleness. How can we leverage logic-based methods to analyze such ML systems with the aim of ensuring their reliability and security? What kind of logical language do we use to specify properties of ML models? How can we ensure that ML models are explainable and interpretable?

In the reverse direction, ML methods have already been successfully applied to making solvers more efficient. In particular, solvers can be modeled as complex combinations of proof systems and ML optimization methods, wherein ML-based heuristics are used to optimally select and sequence proof rules. How can we further deepen this connection between solvers and ML? Can we develop tools that automatically construct proofs for higher mathematics?

This Dagstuhl seminar seeks to answer these and related questions, with the aim of bringing together the many world-leading scientists who are conducting pioneering research at the intersection of logical reasoning and ML, enabling development of novel solutions to problems deemed impossible otherwise.

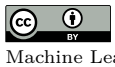
Seminar July 17–22, 2022 – http://www.dagstuhl.de/22291

2012 ACM Subject Classification Theory of computation → Automated reasoning; Computing methodologies → Knowledge representation and reasoning; Theory of computation → Logic; Computing methodologies → Machine learning

Keywords and phrases Logic for ML, ML-based heuristics for solvers, SAT/SMT/CP solvers and theorem provers, Security, reliability and privacy of ML-based systems

Digital Object Identifier 10.4230/DagRep.12.7.80

* Editor / Organizer



Machine Learning and Logical Reasoning: The New Frontier, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 80–111
Editors: Sébastien Bardin, Somesh Jha, and Vijay Ganesh

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Sébastien Bardin
Somesh Jha
Vijay Ganesh

This Dagstuhl seminar is meant to be the first in a series, bringing together researchers from the two main pillars of AI, namely, logical reasoning and machine learning (ML), with a sharp focus on solver-based testing, analysis, and verification (TAV) methods aimed at improving the reliability and security of ML-based systems, and conversely, the use of ML heuristics in improving the power of solvers/provers. A third, albeit smaller focus is neuro-symbolic reasoning (NSR), that aims to combine the power of ML to learn deep correlations with the ability of solvers to perform logical inference as applied to many domains (including but not limited to math and logic).

While many previous Dagstuhl seminars focus on sub-fields of this particular seminar (SAT, SMT, CP or machine learning), we focus here on the synergies and interplay between all them. Our goal is to deepen the understanding of the connections between learning and reasoning, and draw mutually beneficial research directions.

General context: Bringing ML and Logic Reasoning Closer

Since its very inception, Artificial Intelligence (AI) has largely been divided into two broad fields, namely, machine learning (ML) and logical reasoning, that have developed relatively independent of each other. Each of these sub-fields has had a deep and sustained impact on many topics in computer science and beyond, despite the limited interaction between them over the years. However, in recent years new problems and opportunities have come to fore that point towards combinations of ML and logical reasoning as the way forward [1] [1]. In this seminar, we aim to explore combinations of ML and logical reasoning, under the following three specific themes:

Logic Reasoning for ML. Neural Networks (NN) today are ubiquitous and are being deployed as part of critical civilian and defense infrastructure, business processes, automotive software, and governmental decision-making systems. Unfortunately, despite their efficacy in solving many problems, NNs are brittle, unreliable, and pose significant security/privacy challenges [2]. The question of safety and security of NNs has therefore become a great concern to scientists, companies, and governments. In response to this problem, a nascent field of TAV methods for NNs is developing [3]. Key research directions in this context include logics aimed at symbolically representing NNs and their properties [4], novel solving methods [5], as well as solver-based TAV techniques specifically tailored for NNs [6]. A related set of questions focus on explainability and interpretability of NNs [7]. Finally, researchers are also exploring methods that combine logical reasoning within NN learning processes, with the aim of making them adversarially robust [8, 9]. The seminar aims to bring together leading researchers in these topics, enabling cross-fertilization of ideas at a critical juncture in the development of the field.

[1] It goes without saying that it is infeasible to consider all possible combinations of ML and logical reasoning in this seminar. Hence, we focus primarily on problems inspired by testing, analysis, and verification (TAV) of ML and ML-based heuristics for logic solvers, with some forays into neuro-symbolic (a.k.a., neural-symbolic) AI.

ML for Logic Reasoning. In recent years, there has been considerable effort aimed at developing ML-based heuristics for logic reasoning engines such as SAT, SMT, and CP solvers. The premise of this line of research is that logic solvers are a combination of methods that implement proof rules and ML-based heuristics aimed at optimally selecting, sequencing, and initializing such proof rules [10]. This has led to new efficient solving algorithms that can solve real-world formulas with millions of variables and clauses in them. One of the many questions that will be explored in the seminar is how can we further deepen and strengthen this relation between ML and reasoning methods. Yet another line of research being explored is that of replacing rule-based solvers with NN-based logic reasoning (e.g., NeuroSAT [11]). Finally, methods are being developed to combine rule-based methods with reinforcement learning to automatically prove mathematical conjectures [12]. The seminar aims to foster deeper interaction and collaboration among researchers who are pioneers in this intersection of ML-based methods and logical reasoning.

Neuro-symbolic Reasoning. The field of neuro-symbolic reasoning (NSR) aims to combine NNs with symbolic reasoning for the purposes of improving reasoning for many domains (including but not limited to pure math or logic). While at a high-level the field of NSR and logic solvers (with ML heuristics) may seem similar, they employ very different kinds of techniques and have differing goals [1]. For example, NSR researchers have developed methods for translating logical representations of knowledge into neural networks. Others have developed neuro-symbolic methods for concept-learning, and yet others have recently applied NSR to program synthesis. Can these concept-learning methods be adapted to the setting of logic solvers? Could it be that graph neural network (GNN) based representations of mathematical knowledge are easier to analyze? The seminar aims to bring these two disparate communities closer together, that otherwise rarely interact with each other. In a nutshell, the aim of the seminar is to foster cross-fertilization of ideas between the logic reasoning, TAV, and ML communities.

In-depth Description of Focus Areas

Logic Reasoning for ML. As stated above, the reliability, safety, and security of NNs is a critical challenge for society at large. An example of a specific problem in this context is that of adversarial input generation methods against NNs. Many methods have been proposed to address this question, from randomized defense mechanisms to adversarial training to symbolic analysis of NNs via solvers, such as Reluplex [5] that are specifically designed to reason about NNs with ReLU units. Another line of work proposes verification of Binarized Neural Networks (BNNs) via SAT solvers [6]. These initial forays into reasoning for ML bring to fore new challenges, especially having to do with scalability of solvers for NN analysis. Traditional solver methods that scale well for typical software systems, do not seem to scale well for NNs. For example, it is known that solvers, such as Reluplex, are capable of analyzing NNs with only a few thousand nodes. The pressing question of this area of research then is *"How can we develop methods that enable solvers to scale to NNs with millions of nodes in them?"*

A related question has to do with appropriate logics to represent NNs and their properties. Recent work by Soutedeh and Thakur suggests that NNs can be represented symbolically as piecewise linear functions, even though they may use non-linear activation functions such as ReLU [4]. This suggests that there may be efficient solving methods capable of analyzing very large NNs. Yet another question in this setting is how do logic-based methods aimed at testing and verifying NNs compare against hybrid methods that do not require translation of NNs into logic. What are the tradeoffs in this setting?

Another interesting direction where logic reasoning can play a role is in explainability and interpretability of ML models. While both these questions have been long studied in AI and are closely related, they take particular importance in the context of NNs. We say a ML model is explainable, if there is discernable causal relationship between its input and output. Explanations for the behavior of NN, when presented in symbolic form, can be analyzed and debugged using solvers. Researchers have also developed solver-based xAI methods that aim to provide explanations for behavior of NNs [7]. By contrast, interpretable models are ones that have mathematical guarantees regarding their approximation or generalization errors. Solvers can play a role in this context as well via methods for generating counterfactuals (or adversarial examples) [1].

Strong points:

- *Adversarial attacks and defense mechanisms*
- *Neural network testing, analysis, and verification methods*
- *Piecewise linear symbolic representation of NNs*
- *Solvers for NNs*
- *Logic-guided machine learning*
- *Adversarial training*
- *Logic-based explainability and interpretability of NNs*

ML-based Heuristics for Logic Solvers. In recent years, ML-based methods have had a considerable impact on logic solvers. The key premise of this line of research is that logic solvers are a combination of proof systems and ML-based heuristics aimed at optimally selecting, sequencing, and initializing proof rules with the goal of constructing short proofs (if one exists) [10]. A dominant paradigm in this setting is modeling branching heuristics as RL methods to solve the multi-arm bandit (MAB) problem [10]. While this connection seems quite natural today and MAB-style methods have been shown to be empirically powerful, an important question remains as to why these heuristics are effective for industrial instances. A theoretical answer to this question can open up new connections between ML and logic solvers. Another direction of research that has been explored is solving SAT using NNs, *a la* NeuroSAT [11]. Finally, higher-order theorem provers have been developed recently at Google and elsewhere that combine RL with logic reasoning in order to automatically prove theorems from a variety of mathematical fields [13, 12]. The seminar will focus on these recent developments and the next steps in the research on combinations of ML-based methods with logic reasoning with the goal of achieving greater solver efficiency as well as expressive power.

Strong points:

- *ML-techniques for branching and restarts in SAT, SMT, and CP solvers*
- *Supervised learning methods for splitting and initialization in solvers*
- *NN-based methods for logical reasoning*
- *RL for higher-order theorem proving*

Neuro-symbolic Reasoning. Researchers in neuro-symbolic reasoning (NSR) have been independently developing algorithms that combine ML with symbolic reasoning methods with a slightly different focus than solver and theorem prover developers. NSR research has been focused on concept learning in a broader setting than math or logic, and the cross-fertilization of these ideas with logic-based methods can have deep impact both on NSR as well as solver research [1]. One of the key ideas we plan to explore in this context is that of concept learning, i.e., learning of relations or concepts represented in a logical language directly from data. One interesting direction to explore would be how we can incorporate

these methods in logic solvers? Another direction is to explore the synergy between NSR and synthesis of programs from examples. The seminar will focus on bringing NSR and solver researchers closer together, given that they rarely interact in other settings.

Strong points:

- *Concept-learning, with possible applications in higher-order theorem provers*
- *Neuro-symbolic methods for program synthesis*
- *Concept learning for predicate abstraction*

Goals of the Seminar

The aim of this seminar is to bring together the logic reasoning and ML communities, thus shaping and setting the research agenda for ML-based solvers, TAV methods aimed at NNs, and NSR for many years to come.

The seminar will highlight the current challenges with symbolic analysis of NNs, scalability issues with solvers tailored for NNs, state-of-the-art ML-based heuristics for solvers, adapting NSR ideas to the setting of solvers and vice-versa, as well as bring to fore competing TAV methods that don't necessarily rely on symbolic representation of NNs.

Research questions. We highlight some of the main challenges at the intersection of ML and logic reasoning that will be addressed during the seminar from different research perspectives, and discuss how we seek to combine or adapt current techniques to attack them.

- **Symbolic representation of NNs:** Recent work suggests that, while NNs are non-linear functions, they can be effectively modelled symbolically as piecewise linear functions. This is a significant advance since it dramatically simplifies the design of solvers for analyzing NNs. Some of the challenges that remain are algorithmic, i.e., how can NNs be efficiently converted into a symbolic representation.
- **Solvers for NNs:** As of this writing, Reluplex and its successors seem to be among the best solvers for analyzing the symbolic representations of NNs. Unfortunately, these tools scale to NNs with at most a few thousand nodes. There is an urgent need for novel ideas for solving algorithms that enable us to scale to real-world NNs with millions of nodes. Can hybrid methods that combine ML techniques with solvers scale more effectively than pure logic methods?
- **Combining Constraints and NN Learning:** Another set of questions we plan to address is how can we improve the process via which NNs learn using logical constraints. In other words, can the back propagation algorithm be modified to take constraint or domain-specific knowledge into account? Can NNs be combined with logic solvers in a CEGAR-style feedback loop for the purposes of adversarial training?
- **Next steps in ML-based Heuristics for Solvers:** As stated earlier, modern solvers rely in significant ways on ML-based heuristics for their performance. We plan to focus on how we could strengthen this interaction further. For example, are there supervised learning methods for improving the performance of divide-and-conquer parallel SAT solvers. Can we develop ML-based methods for clause sharing in portfolio solvers? How about ML-based restarts and clause deletion policies?
- **Reinforcement learning (RL) and Theorem Provers:** There has been some recent success in combining basic RL methods with reasoning methods in the context of higher-order theorem provers. How can this combination be strengthened further to prove math theorems in a completely automated fashion.

- **Comparison of NN Verification with Testing and Fuzzing Methods:** Researchers have developed a variety of fuzzing methods aimed at NNs. These methods often scale better than verification techniques. On the other hand, unlike verification, testing techniques do not give any guarantees. What are the tradeoffs in this context of complete verification vs. scalability? Can we develop hybrid methods and light-weight verification techniques?
- **Concept Learning and Solvers:** Can we lift the ideas of concept learning from NSR to the setting of solvers, especially in the context of higher-order and combinatorial mathematics?

Synergies. We have also identified the following potential synergies between the ML, Solver, TAV, and NSR communities and expect strong interactions around these points:

- ML researchers in general (and RL in particular) can help refine the ML-based methods used by solver developers;
- Solver developers can propose constraint-based learning strategies for NNs (e.g., combining constraints with gradient-descent in the back propagation algorithm);
- Researchers who work in the space of TAV for NN can benefit greatly by better understanding the realistic security and safety concerns of the ML community;
- Solver developer can substantially benefit by better understanding concept learning from NSR researchers.

Expected results and impact on the research community. One of the core goals of the seminar is to bring together the many different research communities that work in the logic reasoning and ML fields, who unfortunately rarely talk to each other. We believe that the exchange of ideas between them – each with their own methods and perspectives – will help accelerate the future development of combinations of ML and logic reasoning. In terms of concrete outcomes, we believe the workshop is likely to lead to several collaboration projects, especially between members of different communities working on similar or related problems. Common benchmarks and regular meeting forums will also be discussed and we expect for some progress there as well.

References

1. Amel, Kay R. From shallow to deep interactions between knowledge representation, reasoning and machine learning. 13th International Conference Scala Uncertainity Mgmt (SUM 2019), Compiègne, LNCS. 2019.
2. Goodfellow, Ian J and Shlens, Jonathon and Szegedy, Christian. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. 2014
3. Leofante, Francesco and Narodytska, Nina and Pulina, Luca and Tacchella, Armando. Automated verification of neural networks: Advances, challenges and perspectives. arXiv preprint arXiv:1805.09938. 2018
4. Sotoudeh, Matthew and Thakur, Aditya V. A symbolic neural network representation and its application to understanding, verifying, and patching networks. arXiv preprint arXiv:1908.06223. 2019
5. Katz, Guy and Barrett, Clark and Dill, David L and Julian, Kyle and Kochenderfer, Mykel J. Reluplex: An efficient SMT solver for verifying deep neural networks. In CAV 2017
6. Narodytska, Nina and Kasiviswanathan, Shiva and Ryzhyk, Leonid and Sagiv, Mooly and Walsh, Toby. Verifying properties of binarized deep neural networks. AAAI 2018
7. Samek, Wojciech and Wiegand, Thomas and Müller, Klaus-Robert. . Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. 2017

8 Fischer, Marc and Balunovic, Mislav and Drachsler-Cohen, Dana and Gehr, Timon and Zhang, Ce and Vechev, Martin. Dl2: Training and querying neural networks with logic. ICML 2019

9 LGML: Logic Guided Machine Learning. Scott, Joseph and Panju, Maysum and Ganesh, Vijay. AAAI 2020.

10 Jia Hui Liang, Hari Govind V. K., Pascal Poupart, Krzysztof Czarnecki and Vijay Ganesh. An Empirical Study of Branching Heuristics Through the Lens of Global Learning Rate. In SAT 2017.

11 Selsam, Daniel and Lamm, Matthew and Bünz, Benedikt and Liang, Percy and de Moura, Leonardo and Dill, David L. Learning a SAT solver from single-bit supervision. arXiv preprint arXiv:1802.03685. 2018

12 Kaliszyk, Cezary and Urban, Josef and Michalewski, Henryk and Olšák, Miroslav. Reinforcement learning of theorem proving. In Advances in Neural Information Processing Systems. 2018

13 Bansal, Kshitij and Loos, Sarah and Rabe, Markus and Szegedy, Christian and Wilcox, Stewart. HOList: An environment for machine learning of higher order logic theorem proving. ICML 2019.

2 Table of Contents

3 Planning

Monday July 18: (NeuroSymbolic AI Day)

9:00 am – 9:05 am	Welcome and seminar overview (Vijay Ganesh)
9:05 am – 10:05 am	NeuroSymbolic AI: The 3rd Wave (Artur d'Avila Garcez)
10:30 am – 11:00 am	Two birds with one stone? Successes and lessons in building Neuro-Symbolic systems (Xujie Si)
11:00 am – 12:00 pm	Empower Deep Networks with Combinatorial Algorithms (Georg Martius)
2:00 pm – 2:30 pm	CombOptNet: Fit the Right NP-Hard Problem by Learning Integer Programming Constraints (Anselm Paulus)
2:30 pm – 3:00 pm	Learning Modulo Theories (Matt Fredrikson)
3:30 pm – 4:30 pm	Panel on NeuroSymbolic AI (Artur d'Avila Garcez, Georg Martius, Matt Fredrikson) (Moderator: Xujie Si)

Tuesday July 19: (ML for Logic Day)

9:00 am – 10:00 am	Tutorial on ML for Solvers (Vijay Ganesh)
10:00 am – 10:30 am	Machine Learning for Instance Selection in SMT Solving (Pascal Fontaine)
11:00 am – 11:30 am	CrystalBall: Gazing into the Future of SAT Solving (Kuldeep Meel)
11:30 am – 12:00 pm	The importance of memory for mathematical reasoning (Markus Rabe)
2:00 pm – 2:30 pm	AlphaZero from Games to Mathematics (Alhussein Fawzi)
2:30 pm – 3:00 pm	PL vs. AI: The Case for Automated Code Deobfuscation (Sébastien Bardin)
3:30 pm – 4:00 pm	Machine Learning Algorithm Selection for Logic Solvers (Joseph Scott)
5:00 pm – 6:00 pm	Panel on ML for solvers and theorem provers (Markus Rabe, Kuldeep Meel, Alhussein Fawzi, Pascal Fontaine) (Moderator: Sébastien Bardin)

Wednesday July 20: (Testing, Analysis, Verification of ML Systems)

9:00 am – 10:00 am	Tutorial on Temporal Logic for RL (Rajeev Alur)
10:00 am – 10:30 am	Safety Assurance, ML, and Logic – Some Lessons Learned (Chih-Hong Cheng)
11:00 am – 12:00 pm	Perspectives on NeuroSymbolic AI (Luis C. Lamb)
2:00 pm – 2:30 pm	The Foundations of Deep Learning Verification (Matthew Mirman)
2:30 pm – 3:00 pm	Data Usage across the Machine Learning Pipeline (Caterina Urban)

Thursday July 21: (Testing, Analysis, Verification of ML Systems 2)

9:00 am – 10:00 am	Perspectives on Verified AI (Sanjit Seshia)
10:00 am – 10:30 am	There is Plenty of Room at the Bottom : Verification (and Repair) of Small-scale Learning Models (Armando Tacchella)
11:00 am – 11:30 am	Automated Program Analysis: Revisiting Precondition Inference through Constraint Acquisition (Grégoire Menguy)
11:30 am – 12:00 pm	Verification of DNN Controllers (Rajeev Alur)
2:00 pm – 2:30 pm	The Necessity of Run-time Techniques for Safe ML (Ravi Mangal)
2:30 pm – 3:00 pm	Inductive Proofs for Probabilistic Verification and Opportunities in ML (Sebastian Junges)
3:30 pm – 4:00 pm	CGDTest: Constraint-based Gradient Descent Fuzzer for DNNs (Vineel Nagisetty)
4:00 pm – 4:30 pm	Logic for Adversarial ML (Marc Fisher)

Friday July 22: (Synthesis and ML Day)

9:00 am – 10:00 am	Functional Synthesis via Combination of Learning and Reasoning (Kuldeep Meel)
10:00 am – 10:30 am	Synthesizing Pareto-Optimal Interpretations for Black-box Models (Hazem Torfah)
11:00 am – 12:00 pm	Panel on Testing, Analysis, Verification, Security, and Synthesis of AI (Kuldeep Meel, Armando Tacchella, Rajeev Alur, Matt Frederikson) (Moderator: Hazem Torfah)

4 Overview of Talks

4.1 Formal verification for safe autonomy

Rajeev Alur (University of Pennsylvania – Philadelphia, US)

Autonomous systems interacting with the physical world, collecting data, processing it using machine learning algorithms, and making decisions, have the potential to transform a wide range of applications including medicine and transportation. Realizing this potential requires that the system designers can provide high assurance regarding safe and predictable behavior. This motivates research on formally verifying safety (such as avoidance of collisions) of closed-loop systems with controllers based on learning algorithms. In this talk, I will use the experimental platform of the autonomous F1/10 racing car to highlight research challenges for verifying safety for systems with neural-network-based controllers. Our solution to safety verification, incorporated in the tool Verisig at Penn, builds upon techniques for symbolic computation of the set of reachable states of hybrid (mixed discrete-continuous) systems. The case study consists of training the controller using reinforcement learning in a simulation environment, verifying the trained controller using Verisig, and validating the controller by deploying it on the F1/10 racing car.

4.2 Automated Program Analysis: Revisiting Precondition Inference through Constraint Acquisition

Grégoire Menguy (CEA LIST, FR), Sébastien Bardin (CEA LIST, FR)

Joint work of Grégoire Menguy, Sébastien Bardin, Nadjib Lazaar, Arnaud Gotlieb
Main reference Grégoire Menguy, Sébastien Bardin, Nadjib Lazaar, Arnaud Gotlieb: "Automated Program Analysis: Revisiting Precondition Inference through Constraint Acquisition", in Proc. of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022, pp. 1873–1879, ijcai.org, 2022.
URL https://doi.org/10.24963/ijcai.2022/260

Automated program analysis enables to prove code properties like correctness or incorrectness and more generally to help understanding software. Such methods are usually white-box, i.e., they rely on the code syntax to deduce code properties through logical reasoning. While white-box methods have proven to be very powerful, being used for example at Microsoft, Facebook and Airbus, they also suffer from some limitations. First, they need the source code, which is not always available (e.g., proprietary software, malware). Second, the code size and the complexity of data structures manipulated degrade their efficiency drastically. Third, they are highly impacted by syntactic code complexity, which can be amplified by optimizations (improving code speed and memory consumption) and obfuscation (impeding end-users from extracting intellectual property contained in the code).

In this talk, we propose a new method, completely black-box, which infers code annotations from observed code executions only. Indeed, annotations, under the form of function pre/postconditions, are crucial for many software engineering and program verification applications. Unfortunately, they are rarely available and must be retrofit by hand. Thus, we explore how Constraint Acquisition (CA), a learning framework from Constraint Programming, can be leveraged to automatically infer program preconditions in a black-box manner, from input-output observations. We propose PreCA, the first ever framework based on active constraint acquisition dedicated to infer memory-related preconditions. PreCA overpasses prior techniques based on program analysis and formal methods, offering well-identified guarantees and returning more precise results in practice.

References

1 Bessiere, C., Koriche, F., Lazaar, N., O'Sullivan, B. (2017). Constraint acquisition. Artificial Intelligence, 244, 315-342.
2 Rossi, F., Van Beek, P., Walsh, T. (Eds.). (2006). Handbook of constraint programming. Elsevier.
3 Hoare, C. A. R. (1969). An axiomatic basis for computer programming. Communications of the ACM, 12(10), 576-580.
4 Dijkstra, E. W. (1968). A constructive approach to the problem of program correctness. BIT Numerical Mathematics, 8(3), 174-186.
5 Floyd, R. W. (1993). Assigning meanings to programs. In Program Verification (pp. 65-81). Springer, Dordrecht.
6 Menguy, G., Bardin, S., Lazaar, N., Gotlieb, A. Automated Program Analysis: Revisiting Precondition Inference through Constraint Acquisition. IJCAI 2022

4.3 PL vs. AI: The Case of Automated Code Deobfuscation

Sébastien Bardin (CEA LIST, FR)

Joint work of Grégoire Menguy, Sébastien Bardin, Richard Bonichon, Cauim de Souza Lima
Main reference Grégoire Menguy, Sébastien Bardin, Richard Bonichon, Cauim de Souza Lima: "Search-Based Local Black-Box Deobfuscation: Understand, Improve and Mitigate", in Proc. of the CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, Republic of Korea, November 15 – 19, 2021, pp. 2513–2525, ACM, 2021.
URL https://doi.org/10.1145/3460120.3485250

In this talk, we slightly deviate from the main *"Logic # Machine Learning"* topic of the seminar, to consider another closely related inference vs. deduction scheme, namely the link between Program Analysis (PL) and Artificial Intelligence (AI), with a focus on the case of reverse engineering attacks and code deobfuscation. Especially, we discuss how PL and AI both hep in the field, highlight their complementarity strengths and weaknesses and draw lines for future research directions.

Reverse attacks consist in trying to retrieve sensitive information (e.g., secret data, secret algorithms, sensitive business details, etc.) form a program under analysis. We usually consider a very strong attacker with unlimited access to the executable code. Hence, the attacker can for example perform static analysis other the executable, trace the execution, rewrite part of the binary, etc. The goal for the defender is to delay as much as possible the information retrieval, through the use of so-called obfuscation techniques aiming at making the code "hard to understand". We call deobfuscation the effort of removing obfuscations from a program, or helping a reverse engineer to understand an obfuscated program.

Since the middle of the 2010's, several authors managed to adapt PL techniques to perform deobfuscation, with very strong and unexpected results other standard protections. Still, as these *white-box* attacks are based on deduction from the code syntax, they can at some point be fooled by dedicated protections aiming to increase the syntactic complexity of the code in such ways that program analyzer become ineffective.

More recently, *black-box* attacks, based on the observations of input-output relationships together with AI-based synthesis methods in order to rebuild a simple view of the beahviour of some obfuscated parts of the code, show very effective against certain kinds of local obfuscations – even anti-PL obfuscations. Still, these methods are sensitive to semantic complexity, and we show how dedicated protections can take advantage of that.

Finally, as future direction, it seems natural to try combining these dual trends (AI & PL, deduction & inference, blackbox & whitebox) into some combined form of hybrid attacks. From a more theoretical point of view, we could also see this problem as an instance of "AI vs. PL", as PL techniques are also used for the protection side, with questions such as how to train an AI to bypass code protections, or how to create code resistant to AI-augmented attackers.

References

1 Grégoire Menguy, Sébastien Bardin, Richard Bonichon, Cauim de Souza Lima: Search-Based Local Black-Box Deobfuscation: Understand, Improve and Mitigate. CCS 2021
2 Sébastien Bardin, Robin David, Jean-Yves Marion: Backward-Bounded DSE: Targeting Infeasibility Questions on Obfuscated Codes. Symposium on Security and Privacy 2017
3 Mathilde Ollivier, Sébastien Bardin, Richard Bonichon, Jean-Yves Marion: How to kill symbolic deobfuscation for free (or: unleashing the potential of path-oriented protections). ACSAC 2019

4 Mathilde Ollivier, Sébastien Bardin, Richard Bonichon, Jean-Yves Marion. Obfuscation: Where Are We in Anti-DSE Protections? In Proceedings of the 9th Software Security, Protection, and Reverse Engineering Workshop (SSPREW 2019).

5 Sebastian Schrittwieser, Stefan Katzenbeisser, Johannes Kinder, Georg Merz- dovnik, and Edgar Weippl. 2016. Protecting Software Through Obfuscation: Can It Keep Pace with Progress in Code Analysis? ACM Comput. Surv. 49, 1, Article 4 (2016)

6 Babak Yadegari, Brian Johannesmeyer, Ben Whitely, and Saumya Debray. A Generic Approach to Automatic Deobfuscation of Executable Code. In Symposium on Security and Privacy, 2015.

7 Tim Blazytko, Moritz Contag, Cornelius Aschermann, and Thorsten Holz. Syntia: Synthesizing the Semantics of Obfuscated Code. In Usenix Security. 2017

4.4 Safety Assurance, ML, and Logic – Some Lessons Learned

Chih-Hong Cheng (Fraunhofer IKS – München, DE)

Main reference Chih-Hong Cheng, Chung-Hao Huang, Thomas Brunner, Vahid Hashemi: "Towards Safety Verification of Direct Perception Neural Networks", in Proc. of the 2020 Design, Automation & Test in Europe Conference & Exhibition, DATE 2020, Grenoble, France, March 9-13, 2020, pp. 1640–1643, IEEE, 2020.
URL https://doi.org/10.23919/DATE48585.2020.9116205

In this talk, I summarize some of my experiences in engineering ML in safety-critical applications. Within the industry, the emerging consensus is that one requires a systematic & holistic approach to address all potential problems that might occur in the complete life cycle. One can use logic and theorem-proving to tighten the "leap of faith" in the safety argumentation. For formal verification of DNN in autonomous driving, the real challenge lies in creating an implicit specification that characterizes the operational design domain. We use an assume-guarantee reasoning approach, where we learn the operational design domain via abstracting the feature vectors collected by the training data. The formal proof is conditional to the assumption that any input in the operational domain falls inside the abstraction. The abstraction is then deployed in the field as an OoD detector.

4.5 Neurosymbolic AI

Artur d'Avila Garcez (City – University of London, GB), Luis C. Lamb (Federal University of Rio Grande do Sul, BR)

Main reference Artur S. d'Avila Garcez, Luís C. Lamb: "Neurosymbolic AI: The 3rd Wave", CoRR, Vol. abs/2012.05876, 2020.
URL https://arxiv.org/abs/2012.05876

Current advances in Artificial Intelligence (AI) and Machine Learning (DL) have achieved unprecedented impact across research communities and industry. Nevertheless, concerns around trust, safety, interpretability and accountability of AI were raised by influential thinkers. Many identified the need for well-founded knowledge representation and reasoning to be integrated with Deep Learning (DL). Neural-symbolic computing has been an active area of research for many years seeking to bring together robust learning in neural networks with reasoning and explainability by offering symbolic representations for neural models. In [5], recent and early research in neurosymbolic AI is analysed with the objective of

identifying the most important ingredients of neurosymbolic AI systems. Our focus is on research that integrate in a principled way neural network learning with symbolic knowledge representation and logical reasoning. Insights from the past 20 years of research in neural-symbolic computing were discussed and shown to shed new light onto the increasingly prominent role of safety, trust, interpretability and accountability of AI. We also identify promising directions and challenges for the next decade of AI research from the perspective of neural-symbolic computing, commonsense reasoning and causal explanation.

Over the past decade, AI and in particular DL has attracted media attention, has become the focus of increasingly large research endeavours and changed businesses. This led to influential debates on the impact of AI in academia and industry [3]. It has been claimed that deep learning caused a paradigm shift not only in AI, but in several Computer Science fields, including speech recognition, computer vision, image understanding, natural language processing (NLP), and machine translation [2]. Others have argued eloquently that the building of a rich AI system, semantically sound, explainable and ultimately trustworthy, will require a sound reasoning layer in combination with deep learning. Parallels have been drawn between Daniel Kahneman's research on human reasoning and decision making, reflected in his book "Thinking, Fast and Slow [1], and so-called "AI systems 1 and 2, which would in principle be modelled by deep learning and symbolic reasoning, respectively.

We seek to place 20 years of research in the area of neurosymbolic AI, known as neural-symbolic integration, in the context of the recent explosion of interest and excitement around the combination of deep learning and symbolic reasoning. We revisit early theoretical results of fundamental relevance to shaping the latest research, such as the proof that recurrent neural networks can compute the semantics of logic programs, and identify bottlenecks and the most promising technical directions for the sound representation of learning and reasoning in neural networks. As well as pointing to the various related and promising techniques, such as [4], we aim to help organise some of the terminology commonly used around AI, ML and DL. This is important at this exciting time when AI becomes popularized among researchers and practitioners from other areas of Computer Science and from other fields altogether: psychology, cognitive science, economics, medicine, engineering and neuroscience, to name a few.

The first wave of AI in the 1980's was symbolic – based on symbolic logic and logic programming, and later Bayesian networks; the second wave of AI in the 2010's was neural (or connectionist), based on deep learning. Having lived through both waves and having seen the contributions and drawbacks of each technology, we argue that the time is right for the third wave of AI: neurosymbolic AI. Specifically, we summarise the current debate around neurons vs. symbols from the perspective of the long-standing challenges of variable grounding and commonsense reasoning. We survey some of the prominent forms of neural-symbolic integration. We address neural-symbolic integration from the perspective of distributed and localist forms of representation, and argue for a focus on logical representation based on the assumption that representation precedes learning and reasoning. We delve into the fundamentals of current neurosymbolic AI methods and systems and identify promising aspects of neurosymbolic AI to address exciting challenges for learning, reasoning and explainability. Finally, based on all of the above, we propose the list of ingredients for neurosymbolic AI and discuss promising directions for future research to address the challenges of AI.

References

1 D. Kahneman, Thinking, Fast and Slow, 2011, Farrar, Straus and Giroux, New York.
2 Y. LeCun and Y. Bengio and G. Hinton, Deep Learning, Nature 521(7553):436-444, 2015.

3 G. Marcus, The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence, CoRR abs/1801.00631, 2020.
4 Son N. Tran and Artur d'Avila Garcez, Logical Boltzmann Machines, CoRR abs/2112.05841, 2021. https://arxiv.org/abs/2112.05841.
5 Artur d'Avila Garcez and Luis C. Lamb, Neurosymbolic AI: The 3rd Wave, CoRR abs/2012.05876, 2020. https://arxiv.org/abs/2012.05876.

4.6 AlphaZero from games to maths

Alhussein Fawzi (Google DeepMind – London, GB)

Main reference Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J. R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, David Silver, Demis Hassabis, Pushmeet Kohli: "Discovering faster matrix multiplication algorithms with reinforcement learning", Nat., Vol. 610(7930), pp. 47–53, 2022.
URL https://doi.org/10.1038/s41586-022-05172-4

In this talk, I will describe how we can extend AlphaZero, a reinforcement learning agent developed for playing games such as Go and Chess, to mathematical problems. I will go in detail through the different components of AlphaZero, and focus on some of the challenges of applying it to mathematical problems. To illustrate my talk, I will focus precisely on "decomposition problems", where the task is to decompose a hard mathematical object (e.g., a tensor) into a sum of atoms (e.g., rank one tensors).

4.7 Logic & Adversarial Machine Learning

Marc Fischer (ETH Zürich, CH)

Joint work of Christian Sprecher, Anian Ruoss, Dimitar I. Dimitrov, Mislav Balunović, Timon Gehr, Gagandeep Singh, Dana Drachsler-Cohen, Ce Zhang, Martin Vechev
Main reference Dana Drachsler-Cohen Marc Fischer Mislav Balunović: "DL2: Training and Querying Neural Networks with Logic", in Proc. of the International Conference on Machine Learning, 2019.
URL https://www.sri.inf.ethz.ch/publications/fischer2019dl2

The discovery of adversarial examples, small semantic-preserving perturbations that mislead neural networks, highlighted the need to study the robustness of machine learning systems. In this talk, I discuss three perspectives connecting this study of robustness with logic:

- First, how can we use techniques for finding and defending against adversarial examples to query and train neural networks with logic specifications?
- Second, how can we leverage relations between different inputs and input specifications in a robustness analysis based on abstract interpretation – the symbolic propagation of input sets through programs?
- Third, how can we combine these techniques to enforce and verify notions of individual fairness?

References

1 Anian Ruoss, Mislav Balunovic, Marc Fischer, Martin T. Vechev: Learning Certified Individually Fair Representations. NeurIPS 2020

2 Marc Fischer, Christian Sprecher, Dimitar I. Dimitrov, Gagandeep Singh, Martin T. Vechev: Shared Certificates for Neural Network Verification. CAV 2022
3 Marc Fischer, Mislav Balunovic, Dana Drachsler-Cohen, Timon Gehr, Ce Zhang, Martin T. Vechev: DL2: Training and Querying Neural Networks with Logic. ICML 2019

4.8 Learning Modulo Theories: Leveraging Theory Solvers for Machine Learning

Matt Fredrikson (Carnegie Mellon University – Pittsburgh, US)

A recently proposed class of techniques, which aim to integrate solver layers within Deep Neural Networks (DNNs), has shown considerable promise in bridging a long-standing gap between inductive learning and symbolic reasoning techniques. This approach brings the capabilities of a decision procedure to a learned model, both during training and inference. Such an approach is particularly useful in solving problems that have both a perceptual as well as logical or combinatorial sub-tasks. Statistical learning excels at the perceptual, while progress in solver technology continues to open new horizons for the logical.

We will present a new framework, ERM(ϕ), and an associated set of methods for integrating solver layers that encompass a broad range of symbolic knowledge into an ML system. Using this framework, we demonstrate several fundamental challenges and opportunities for this direction. Further, we provide a set of algorithms for computing the forward (inference) and backward (training) passes of a DNN layer that makes calls to an SMT solver, with few restrictions on the user-provided constraints that the solver can query. For example, the theory solver does not need to be differentiable. The talk will conclude by giving an overview of an implementation of our approach within Pytorch, using Z3 to solve constraints, and show how to construct vision and natural language models that incorporate symbolic knowledge during training and inference, and can outperform conventional models – especially in settings where training data is limited or when the cost of fine-grained labeling is prohibitive.

4.9 Opportunities for Neurosymbolic Approaches in the Context of Probabilistic Verification

Sebastian Junges (Radboud University Nijmegen, NL)

Joint work of Steven Carr, Nils Jansen, Ufuk Tocu, Kevin Batz, Mingshuai Chen, Benjamin Lucien Kaminski, Joost-Pieter Katoen, Christoph Matheja

In this overview, we outline some challenges for neurosymbolic approaches in the context of probabilistic verification. In short, probabilistic verification aims to show that a specification holds on a system with probabilistic uncertainties. Such systems can be modelled as probabilistic programs, as stochastic Petri nets, as Bayesian dynamic networks, or any other description language to describe a Markov models. One step beyond verification is synthesis,

in this context often policy synthesis, in which the goal is to find policies for agents making decisions under uncertainty, such that the joint behavior of agent and environment satisfies the given specification. We discuss two directions: *"Solver Inside"* and *"Learning Inside"*.

For solver inside, we discuss shielding in reinforcement learning [1]. In particular, we report on shielding for partially observable Markov decision processes and the integration with state-of-the-art deep reinforcement learning [2]. We briefly discuss how shields are computed by an iterative application of SAT-solvers [3]. We discuss the framework and the relative strengths and weaknesses of addings shields in sparse-reward settings. Among others, we show how bootstrapping with a shield can help guide the learning process.

For learning inside, we consider various inductive synthesis frameworks. We may aim to learn inductive invariants for probabilistic models and programs, alike to learning inductive invariants for deterministic programs. A major challenge over learning deterministic invariants is the continuous search space. While results for property-directed-reachability (PDR) on Markov decision processes are mixed [4], the use of a CEGIS-style loop are more promissing [5]. It remains an important challenge how to guess the right form of templates. Data-driven approaches may be helpful. We furthermore briefly discuss inductive synthesis for policies described by small finite-state controllers [6]. Data-driven approaches to come up with good, diverse policies will be able to boost the state-of-the-art.

References

1 Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, Ufuk Topcu: *Safe Reinforcement Learning via Shielding.* AAAI 2018: 2669-2678.

2 Steven Carr, Nils Jansen, Sebastian Junges, Ufuk Topcu: *Safe Reinforcement Learning via Shielding for POMDPs.* CoRR abs/2204.00755 (2022)

3 Sebastian Junges, Nils Jansen, Sanjit A. Seshia: Enforcing Almost-Sure Reachability in POMDPs. CAV 2021: 602-625

4 Kevin Batz, Sebastian Junges, Benjamin Lucien Kaminski, Joost-Pieter Katoen, Christoph Matheja, Philipp Schröer: *PrIC3: Property Directed Reachability for MDPs.* CAV 2020: 512-538

5 Kevin Batz, Mingshuai Chen, Sebastian Junges, Benjamin Lucien Kaminski, Joost-Pieter Katoen, Christoph Matheja: *Probabilistic Program Verification via Inductive Synthesis of Inductive Invariants.* CoRR abs/2205.06152 (2022)

6 Roman Andriushchenko, Milan Ceska, Sebastian Junges, Joost-Pieter Katoen: *Inductive synthesis of finite-state controllers for POMDPs.* UAI 2022: 85-95

4.10 From Learning to Reasoning in Neurosymbolic AI

Luis C. Lamb (Federal University of Rio Grande do Sul, BR)

Joint work of Luis C. Lamb, Artur d'Avila Garcez

Neurosymbolic AI aims to bring together the statistical nature of machine learning and the logical essence of reasoning in AI systems. Such integration demands a shift as regards research methodology, since the connectionist and symbolic schools of AI have been developed under distinct technical foundations over the last 50 years [1, 4]. Nonetheless, leading technology companies and research groups have put forward agendas for the development of the field, as modern AI systems require sound reasoning, interpretability, and improved explainability [7, 5]. Moreover, AI and deep learning researchers have also pointed out that Neurosymbolic AI is

one of *"the most promising approach to a broad AI [..], that is, a bilateral AI that combines methods from symbolic and sub-symbolic AI"*[2]. In this talk, we highlight how the evolution of Neurosymbolic AI research results can lead to applications and novel developments towards building robust, explainable AI systems. We summarize how Neurosymbolic AI evolved over the years and how it might contribute to improved explainability and the effective integration of learning and reasoning in the construction of robust AI systems. This talk is motivated by the evolution of our work on the integration of modal, temporal, and intuitionistic logics and neural learning [4]. Over the years, we showed that the proper integration of logical methods and neural learning can lead to applications in classical benchmarks in multiagent systems [3], modelling the evolution of software requirements specifications, and possibly to a better understanding of the learnability of rich graph-based and optimization problems [1, 6].

References

1 Artur d'Avila Garcez and Luís C. Lamb. Neurosymbolic AI: The 3rd Wave. CoRR, abs/2012.05876. 2020.
2 Sepp Hochreiter. Toward a broad AI. CACM 2022.
3 Ronald Fagin, Joseph Y. Halpern, Yoram Moses and Moshe Y. Vardi. Reasoning About Knowledge. MIT Press, 1995.
4 Artur S. d'Avila Garcez, Luís C. Lamb and Dov M. Gabbay. Neural-Symbolic Cognitive Reasoning. Cognitive Technologies. 2009
5 Leslie Valiant. Probably approximately correct: nature's algorithms for learning and prospering in a complex world. Basic Books (AZ), 2013
6 Luís C. Lamb, Artur S. d'Avila Garcez, Marco Gori, Marcelo O. R. Prates, Pedro H. C. Avelar and Moshe Y. Vardi. Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective. IJCAI 2020.
7 Gary Marcus. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. CoRR, abs/2002.06177, 2020.

4.11 The Necessity of Run-time Techniques for Safe ML (and how to deal with the pitfalls)

Ravi Mangal (Carnegie Mellon University – Pittsburgh, US)

Joint work of Ravi Mangal, Matt Fredrikson, Aymeric Fromherz, Klas Leino, Bryan Parno, Corina Păsăreanu, Chi Zhang
Main reference Klas Leino, Aymeric Fromherz, Ravi Mangal, Matt Fredrikson, Bryan Parno, Corina Păsăreanu: "Self-Correcting Neural Networks For Safe Classification", arXiv, 2021.
URL https://doi.org/10.48550/ARXIV.2107.11445

Neural networks are increasingly being used as components in systems where safety is a critical concern. Although pre-deployment verification of these networks with respect to required specifications is highly desirable, the specifications are not always amenable to verification. In particular, adversarial robustness, a popular specification, requires that a neural network f exhibit local robustness at every input x in the support of its input data distribution D. Local robustness at an input x is the property that $\forall x'.||x - x'|| \leq \epsilon \Rightarrow f(x) = f(x')$. Unfortunately, neither the distribution D nor its support are known in advance. We advocate for the use of run-time or inference-time (i.e., post-deployment) checks to deal with such distribution-dependent specifications. For instance, to ensure adversarial robustness, a network should be used for prediction only if it passes a local robustness check at run-time, otherwise it should abstain from prediction.

While run-time checks can ensure that neural networks do not misbehave, each abstention incurs a cost since one has to default to an expensive fall-back mechanism (typically, human decision-makers). For run-time checks that encode a class of constraints called *safe-ordering constraints*, we propose a mechanism for repairing the outputs of a neural network whenever the run-time check fails. These constraints relate requirements on the order of the classes output by a classifier to conditions on its input. Though local robustness cannot be encoded as a safe-ordering constraint, this fragment is expressive enough to encode various interesting examples of neural network safety specifications from the literature.

Our repair mechanism is based on a self-repairing layer which performs constraint solving and provably yields safe outputs regardless of the characteristics of the network input. We compose this layer with an existing neural network to construct a self-repairing network (SR-Net), and show that in addition to providing safe outputs, the SR-Net is guaranteed to preserve the classification accuracy of the original network whenever possible. Our approach is independent of the size and architecture of the neural network used for classification, depending only on the specified property and the dimension of the network's output; thus it is scalable to large state-of-the-art networks. We show that our approach can be optimized for a GPU, introducing run-time overhead of less than 1ms on current hardware – even on large, widely-used networks containing hundreds of thousands of neurons and millions of parameters. Designing a run-time repair mechanism to handle failures of local robustness checks is an interesting direction for future research.

References

1 Klas Leino, Aymeric Fromherz, Ravi Mangal, Matt Fredrikson, Bryan Parno, and Corina Păsăreanu. *Self-Correcting Neural Networks for Safe Classification*. 5th International Workshop on Software Verification and Formal Methods for ML-Enables Autonomous Systems (FoMLAS), 2022.

2 Klas Leino, Chi Zhang, Ravi Mangal, Matt Fredrikson, Bryan Parno, and Corina Păsăreanu. *Degradation Attacks on Certifiably Robust Neural Networks*. Transactions on Machine Learning Reasearch (TMLR), 2022.

4.12 The importance of memory for mathematical reasoning

Markus Rabe (Google – Mountain View, US)

Joint work of Markus Rabe, Christian Szegedy, Yuhuai Tony Wu, DeLesley Hutchins, Charles Staats, and others
Main reference Yuhuai Wu, Markus N. Rabe, DeLesley Hutchins, Christian Szegedy: "Memorizing Transformers", CoRR, Vol. abs/2203.08913, 2022.
URL https://doi.org/10.48550/arXiv.2203.08913

In this talk I was discussing astonishing progress in using large language models for mathematical reasoning. One of the main bottlenecks at the moment is the ability to process long documents (books, papers, ...) at once instead of looking at small (page-length) snippets of the data. Our paper on Memorizing Transformers opens a path to equipping existing large language models with the ability to process book-length data, which we improves the performance of large language models on code and mathematical data significantly.

4.13 Blackbox Differentiation: Empower Deep Networks with Combinatorial Algorithms

Georg Martius (MPI für Intelligente Systeme – Tübingen, DE), Anselm Paulus (MPI für Intelligente Systeme – Tübingen, DE)

Joint work of Marin Vlastelica Pogancic, Anselm Paulus, Vít Musil, Georg Martius, Michal Rolínek
Main reference Marin Vlastelica Pogancic, Anselm Paulus, Vít Musil, Georg Martius, Michal Rolínek: "Differentiation of Blackbox Combinatorial Solvers", in Proc. of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, OpenReview.net, 2020.
URL https://openreview.net/forum?id=BkevoJSYPB

Machine Learning has achieved great successes on solving problems that seemed unsolvable just a decade ago. Examples are the mastering of the game of Go, automatic machine translation, and learning in-hand manipulation with a robotic hand. Besides many technical innovations, these advances have been enabled by two main ingredients: highly flexible differentiable function approximators (deep networks) and huge amounts of data. While deep networks can extract very complicated patterns from data, there is a certain sense of dissatisfaction when it comes to their performance on tasks with combinatorial or algorithmic complexity. For example, think of learning to find the shortest path in an environment when provided only with raw birds-eye maps (images). Current, deep networks can learn this task on maps they were trained on, but perform poorly on new maps. The reason is that part of the problem has an algorithmic nature: the same shortest path algorithm works on all maps, if suitably represented as a graph. However a normal deep network cannot perform the same computations and thus can only learn to imitate the process.

The next big step for researchers in machine learning and artificial intelligence is to enhance the ability of the methods to reason. This sentiment was for example expressed by Battaglia et. al. [1] who advocate that "combinatorial generalization must be a top priority for AI".

Importantly, there are decades worth of research contributions in graph algorithms and discrete optimization. We have optimal runtimes for sorting algorithms, clever tricks for various algorithmic problems over graphs/networks such as for shortest path or various cuts or matching-based problems. In other words, when faced with combinatorial or algorithmic problems in isolation and with a clean specification, we already have very strong methods for solving them. This should not be ignored.

While there is some level of success in designing deep learning architectures with "algorithmic behavior", the classical methods are still miles ahead when it comes to performance in purely combinatorial setups. We believe the right approach is to build bridges between the two disciplines so that progress can freely flow from one to another. In that spirit, we would rephrase the earlier sentiment as "merging techniques from combinatorial optimization and deep learning must be a top priority for AI".

We have recently developed a method [2] that allows to embed a large class of combinatorial algorithms in deep neural networks while maintaining the usual training procedure unchanged. In the talk, I will explain the fundamental problem that we had to overcome and show examples of what can be done with the new architecture. This includes the shortest path problem on raw images [2], finding correspondences in pairs of images [3], and directly optimizing for rank-based loss functions [4].

We have two blog-posts on this topic:
https://towardsdatascience.com/the-fusion-of-deep-learning-and-combinatorics-4d0112a74fa7
https://towardsdatascience.com/rambo-ranking-metric-blackbox-optimization-36811a5f52dd

References

1 P. Battaglia et al. *Relational inductive biases, deep learning, and graph networks.* https://arxiv.org/abs/1806.01261, 2018
2 M. Vlastelica, A. Paulus, V. Musil, G. Martius, and M. Rolínek. *Differentiation of blackbox combinatorial solvers.* In International Conference on Learning Representations, ICLR, 2020.
3 M. Rolínek, P. Swoboda, D. Zietlow, A. Paulus, V. Musil, and G. Martius. *Deep graph matching via blackbox differentiation of combinatorial solvers.* In European Conference on Computer Vision ECCV, 2020
4 M. Rolínek, V. Musil, A. Paulus, M. Vlastelica, C. Michaelis, and G. Martius. *Optimizing ranking-based metrics with blackbox differentiation.* In Conference on Computer Vision and Pattern Recognition, CVPR, 2020

4.14 Democratizing SAT Solving

Kuldeep S. Meel (National University of Singapore, SG)

Joint work of Mate Soos, Raghav Kulkarni, Kuldeep S. Meel
Main reference Mate Soos, Raghav Kulkarni, Kuldeep S. Meel: "CrystalBall: Gazing in the Black Box of SAT Solving", in Proc. of the Theory and Applications of Satisfiability Testing – SAT 2019 – 22nd International Conference, SAT 2019, Lisbon, Portugal, July 9-12, 2019, Proceedings, Lecture Notes in Computer Science, Vol. 11628, pp. 371–387, Springer, 2019.
URL https://doi.org/10.1007/978-3-030-24258-9_26

Boolean satisfiability is a fundamental problem in computer science with a wide range of applications including planning, configuration management, design and verification of software/hardware systems. The annual SAT competition continues to witness impressive improvements in the performance of the winning SAT solvers largely thanks to the development of new heuristics arising out of intensive collaborative research in the SAT community. Modern SAT solvers achieve scalability and robustness with complex heuristics that are challenging to understand and explain. Consequently, the development of new algorithmic insights has been largely restricted to experts in the SAT community.

I will describe our project that aims to democratize the design of SAT solvers. In particular, our project, called CrystalBall, seeks to develop a framework to provide white-box access to the execution of SAT solver that can aid both SAT solver developers and users to synthesize algorithmic heuristics for modern SAT solvers? We view modern conflict-driven clause learning (CDCL) solvers as a composition of classifiers and regressors for different tasks such as branching, clause memory management, and restarting, and we aim to provide a data-driven automated heuristic design mechanism that can allow experts in domains outside SAT community to contribute to the development of SAT solvers.

4.15 Functional Synthesis – An Ideal Meeting Ground for Formal Methods and Machine Learning

Kuldeep S. Meel (National University of Singapore, SG)

Joint work of Priyanka Golia, Subhajit Roy, Kuldeep S. Meel, Friedrich Slivovsky
Main reference Priyanka Golia, Subhajit Roy, Kuldeep S. Meel: "Manthan: A Data-Driven Approach for Boolean Function Synthesis", in Proc. of the Computer Aided Verification – 32nd International Conference, CAV 2020, Los Angeles, CA, USA, July 21-24, 2020, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 12225, pp. 611–633, Springer, 2020.
URL https://doi.org/10.1007/978-3-030-53291-8_31

Don't we all dream of the perfect assistant whom we can just tell what to do and the assistant can figure out how to accomplish the tasks? Formally, given a specification $F(X, Y)$ over the set of input variables X and output variables Y, we want the assistant, aka functional synthesis engine, to design a function G such that $F(X, G(X))$ is true. Functional synthesis has been studied for over 150 years, dating back Boole in 1850's and yet scalability remains a core challenge. Motivated by progress in machine learning, we design a new algorithmic framework Manthan, which views functional synthesis as a classification problem, relying on advances in constrained sampling for data generation, and advances in automated reasoning for a novel proof-guided refinement and provable verification. The significant performance improvements call for interesting future work at the intersection of machine learning, constrained sampling, and automated reasoning.

References

1 Priyanka Golia, Subhajit Roy, and Kuldeep S. Meel. Manthan: A data-driven approach for boolean function synthesis. CAV 2020
2 Priyanka Golia, Subhajit Roy, and Kuldeep S. Meel. Program synthesis as dependency quantified formula modulo theory. In IJCAI 2021
3 Priyanka Golia, Friedrich Slivovsky, Subhajit Roy, and Kuldeep S. Meel. Engineering an efficient boolean functional synthesis engine. In ICCAD 2021

4.16 Verifiable Neural Networks: Theoretical Capabilities and Limitations

Matthew Mirman (ETH Zürich, CH)

Famously, deep learning has become an integral part of many high stakes applications, from autonomous driving to health care. As the discovery of vulnerabilities and flaws in these models has become frequent, so has the interest in ensuring their robustness and reliability. In recent years, many methods have been developed to build deep learning models amenable to analysis with efficient formal methods. However, these techniques, known together as provability training, have failed to produce models with nearly the empirical quality as traditional training. This stagnation has opened up questions as to the theoretical foundations of provability training. In this talk I will explain our theoretical results on both the possibility and impossibility of constructing verifiable neural networks. To motivate continued search for provable training methods, I will present our possibility result: a stronger form of the universal approximation theorem for the case of interval-certifiable

neural networks. To begin to explain the barriers to provable training, I will present our impossibility results: (i) that for any neural network classifying just three points, there is a valid specification over these points that interval analysis can not prove, and (ii) given any radius, there is a set of points that no one-hidden-layer network can be proven to interval-robustly classify.

4.17 CGDTest: Constraint-based Gradient Descent Fuzzer for DNNs

Vineel Nagisetty (Borealis AI – Toronto, CA)

Joint work of Vineel Nagisetty, Guanting Pan, Piyush Jha, Dhananjay Ashok, Laura Graves, Christopher Srinivasa, Vijay Ganesh

In this work we propose a new Deep Neural Network (DNN) testing algorithm, called the Constrained Gradient Descent (CGD) method, and an implementation we call CGDTest aimed at exposing security issues such as testing for adversarial robustness and fairness in DNNs. Our CGD algorithm is a gradient-descent (GD) method, with the twist that the user can also specify logical properties that characterize a specific type of input that they want. This functionality allows us to specify constraints so as to test DNNs for standard Lp ball-based adversarial robustness as well as other properties such as non-standard adversarial robustness and individual fairness. We perform extensive experiments where we use CGDTest to test for both standard and non-standard adversarial robustness in the vision domain, adversarial robustness in the NLP domain, and individual fairness in the tabular domain, comparing against 18 state-of-the-art methods over the 3 domains. Our results indicate that CGDTest is comparable to state-of-the-art tools in testing for standard definitions of robustness and fairness, and is significantly superior in testing for non-standard robustness, with improvements in PAR2 score of over 1500% in some cases over the next best tool. Our evaluation shows that CGD method outperforms all other methods in terms of scalability (i.e., can be applied to very large real-world models with millions of parameters), expressibility (i.e., test for a variety of properties from disparate domains), and generality (i.e., handle a variety of architectures).

References

1 Fischer M, Balunovic M, Drachsler-Cohen D, Gehr T, Zhang C, Vechev M. Dl2: Training and querying neural networks with logic. InInternational Conference on Machine Learning 2019 May 24 (pp. 1931-1941). PMLR.
2 Kimmig A, Bach S, Broecheler M, Huang B, Getoor L. A short introduction to probabilistic soft logic. InProceedings of the NIPS workshop on probabilistic programming: foundations and applications 2012 (pp. 1-4).
3 Liu C, Arnon T, Lazarus C, Strong C, Barrett C, Kochenderfer MJ. Algorithms for verifying deep neural networks. Foundations and Trends® in Optimization. 2021 Feb 10;4(3-4):244-404.
4 Katz G, Huang DA, Ibeling D, Julian K, Lazarus C, Lim R, Shah P, Thakoor S, Wu H, Zeljić A, Dill DL. The marabou framework for verification and analysis of deep neural networks. InInternational Conference on Computer Aided Verification 2019 Jul 15 (pp. 443-452). Springer, Cham.
5 Singh G, Gehr T, Püschel M, Vechev M. An abstract domain for certifying neural networks. Proceedings of the ACM on Programming Languages. 2019 Jan 2;3(POPL):1-30.

6 Zhang H, Chen H, Xiao C, Gowal S, Stanforth R, Li B, Boning D, Hsieh CJ. Towards stable and efficient training of verifiably robust neural networks. arXiv preprint arXiv:1906.06316. 2019 Jun 14.

4.18 Machine Learning for Instance Selection in SMT Solving

Pascal Fontaine (University of Liège, BE)

Joint work of Daniel El Ouraoui, Cezary Kaliszyk, Jasmin Blanchette, Pascal Fontaine
Main reference Jasmin Christian Blanchette, Daniel El Ouraoui, Pascal Fontaine, Cezary Kaliszyk: "Machine Learning for Instance Selection in SMT Solving", in Proc. of the AITP 2019 – 4th Conference on Artificial Intelligence and Theorem Proving, 2019.
URL https://hal.archives-ouvertes.fr/hal-02381430

Satisfiability Modulo Theories (SMT) solvers are powerful tools used to check specifications of critical systems and to discharge proof obligations in proof assistants. For many such applications, quantifiers are necessary to express the problems. It often happens that SMT solvers fail finding proofs when too many quantifiers occur in the input problem. To deal with quantifiers, SMT solvers rely on instantiation, and use heuristic techniques to generate instances. Often, thousands of instances are generated and since most of them are useless, they impede the solver.

We use machine learning to predict the usefulness of an instance in order to decrease the number of instances generated and handled by the SMT solver. For this, we propose a meaningful way to characterize the state of an SMT solver, we collect instantiation learning data, and we integrate a predictor in the core of a state-of-the-art SMT solver. This ultimately leads to more efficient SMT solving for quantified problems.

4.19 CombOptNet: Fit the Right NP-Hard Problem by Learning Integer Programming Constraints

Anselm Paulus (MPI für Intelligente Systeme – Tübingen, DE) and Georg Martius (MPI für Intelligente Systeme – Tübingen, DE)

Joint work of Anselm Paulus, Michal Rolínek, Vit Musil, Brandon Amos, Georg Martius
Main reference Anselm Paulus, Michal Rolínek, Vit Musil, Brandon Amos, Georg Martius: "CombOptNet: Fit the Right NP-Hard Problem by Learning Integer Programming Constraints", in Proc. of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research, Vol. 139, pp. 8443–8453, PMLR, 2021.
URL https://proceedings.mlr.press/v139/paulus21a.html

Over recent years, deep learning has revolutionized multiple fields, such as computer vision, robotics, and natural language processing. This progress has predominantly built on the astonishing ability of neural networks to extract valuable information from raw data, an essential skill for approaching real-world problems. However, despite these successes, neural networks still notoriously struggle at algorithmic and logical reasoning tasks.

Such tasks can often be solved efficiently by combinatorial solvers, such as SAT or ILP solvers, which can build on a long development history. However, these solvers usually require a clean abstract formulation of the problem, such as boolean clauses or cost and constraint coefficients, instead of operating on raw data.

How can we bridge the gap between these two worlds? Ideally, we would like to use combinatorial solvers as building blocks of neural networks to build hybrid architectures that leverage both the feature extraction and the algorithmic reasoning capabilities of neural networks and combinatorial solvers, respectively. However, as combinatorial solvers typically operate on discrete structures, there is no continuously differentiable relationship between the inputs and outputs. In contrast, deep learning at its core relies on differentiability for end-to-end learning. Overcoming this fundamental conflict poses a significant challenge.

Recently proposed methods have addressed this challenge by considering continuously differentiable relaxations or by relying on informative gradient replacements that exploit the structure of the solver [1]. These methods have enabled the integration of dedicated combinatorial solvers into end-to-end trainable architectures, which extract the cost coefficients of the solver from raw data. While achieving promising results in computer vision and natural language processing applications, the strong prior information required to guide the choice of the problem-tailored combinatorial solver remains a limiting factor.

As an answer to this limitation, this talk introduces the recently developed method CombOptNet [2]. The goal of this method is to remove the restriction of a priori specifying a dedicated solver and to instead rely on a more general combinatorial building block. It is well known that many combinatorial problems can be formulated as ILPs, in which the constraint set determines the nature of the problem. Based on this observation, we aim to integrate a general ILP Solver into deep learning architectures. CombOptNet provides informative gradient replacements for both the cost and constraint coefficients. By learning the constraints from raw data, the architecture infers the nature of the combinatorial problem at hand. Thereby the architecture strives to achieve universal combinatorial expressivity.

References

1 M. Vlastelica, A. Paulus, V. Musil, G. Martius, and M. Rolínek. *Differentiation of BlackBox combinatorial solvers*. In International Conference on Learning Representations, ICLR, 2020.

2 A. Paulus, M. Rolinek, V. Musil, B. Amos, and G. Martius. *CombOptNet: Fit the Right NP-Hard Problem by Learning Integer Programming Constraints*. In International Conference on Machine Learning, ICML, 2021.

4.20 Machine Learning Algorithm Selection for Logic Solvers

Joseph Scott (University of Waterloo, CA)

Joint work of Joseph Scott, Vijay Ganesh, Aina Niemetz, Mathias Preiner, Saeed Nejati, Maysum Panju, Tony Pan

In this two-part talk, I present recent work in machine learning applied to logic and logic applied to machine learning.

First, I present some recent results on algorithm selection for logic solvers, specifically in the context of SMT solvers and neural network verification. As is typical for hard search problems, no single solver is expected to be the fastest on all inputs. This insight suggests using algorithm selection techniques that automatically select the fastest solver for a given input. We present MachSMT, an algorithm selection tool for SatisfiabilityModulo

Theories (SMT) solvers. MachSMT supports the entirety of the SMT-LIB. We provide an extensive empirical evaluation of MachSMT to demonstrate the efficiency and efficacy of MachSMT over three broad usage scenarios on theories and theory combinations of practical relevance (e.g., bit-vectors,(non-)linear integer and real arithmetic, arrays, and floating-point arithmetic). Additionally, we present Goose, an adaptive algorithm selection tool, which we dub a *meta-solver*, for deep neural network verification. We evaluate Goose by simulating VNN-COMP '21 and observe a 37% improvement over the competition winner.

Second, we introduce Logic Guided Machine Learning (LGML), a novel approach that symbiotically combines machine learning (ML) and logic solvers with the goal of learning mathematical functions from data. LGML consists of two phases, namely a learning-phase and a logic-phase with a corrective feedback loop, such that, the learning-phase learns symbolic expressions from input data, and the logic-phase cross verifies the consistency of the learned expression with known auxiliary truths. If inconsistent, the logic-phase feeds back "counterexamples" to the learning-phase. This process is repeated until the learned expression is consistent with auxiliary truth. Using LGML, we were able to learn expressions that correspond to the Pythagorean theorem and the sine function, with several orders of magnitude improvements in data efficiency compared to an approach based on an out-of-the-box multilayered perceptron (MLP).

References

1 Joseph Scott, Maysum Panju and Vijay Ganesh. LGML: Logic Guided Machine Learning (Student Abstract), AAAI 2020
2 Joseph Scott, Aina Niemetz, Mathias Preiner, Saeed Nejati and Vijay Ganesh. MachSMT: A Machine Learning-based Algorithm Selector for SMT Solvers. TACAS 2021.
3 Joseph Scott, Guanting Pan, Elias B. Khalil and Vijay Ganesh. Goose: A Meta-Solver for Deep Neural Network Verification. SMT workshop 2020.
4 Dhananjay Ashok, Joseph Scott, Sebastian Johann Wetzel, Maysum Panju and Vijay Ganesh. Logic Guided Genetic Algorithms (Student Abstract). AAAI 2021.
5 Lin Xu, Frank Hutter, Holger H. Hoos and Kevin Leyton-Brown. SATzilla: Portfolio-based Algorithm Selection for SAT. J. Artif. Intell. Res., 2008.

4.21 Two birds with one stone? Successes and lessons in building Neuro-Symbolic systems

Xujie Si (McGill University – Montréal, CA)

Joint work of Sever Topan, David Rolnick, Xujie Si
Main reference Sever Topan, David Rolnick, Xujie Si: "Techniques for Symbol Grounding with SATNet", in Proc. of the Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual, pp. 20733–20744, 2021.
URL https://proceedings.neurips.cc/paper/2021/hash/ad7ed5d47b9baceb12045a929e7e2f66-Abstract.html

Deep learning models have achieved remarkable successes in many challenging fields but suffer from a lack of interpretability, poor generalization ability, and difficulty in integrating human knowledge. Symbolic systems on the other hand address these limitations by design but heavily rely on hardcoded knowledge and have a very limited capability of learning. A promising design is perhaps building neuro-symbolic systems, combining the benefits of both worlds. However, such a design inevitably combines the challenges from both worlds and also

faces some unique challenges in itself. In this talk, I will share some successes and lessons in building two neuro-symbolic systems – a data-driven optimization for symbolic software model checking and an end-to-end visual sudoku solver.

References

1 Sever Topan, David Rolnick, and Xujie Si. *Techniques for Symbol Grounding with SATNet*. Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual
2 Nham Le, Xujie Si, Arie Gurfinkel. *Data-driven Optimization of Inductive Generalization*. Formal Methods in Computer Aided Design, FMCAD 2021, New Haven, CT, USA, October 19-22, 2021

4.22 There is plenty of room at the bottom: verification and repair of small scale learning models

Armando Tacchella (University of Genova, IT)

Joint work of Armando Tacchella, Luca Pulina, Giorgio Metta, Lorenzo Natale, Shashank Pathank, Erika Abraham, Nils Jansen, Dario Guidotti, Francesco Leofante, Joost-Pieter Katoen, Simone Vuotto, Nina Narodytska, Andrea Pitto
Main reference Dario Guidotti, Francesco Leofante, Luca Pulina, Armando Tacchella: "Verification of Neural Networks: Enhancing Scalability Through Pruning", in Proc. of the ECAI 2020 – 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 – September 8, 2020 – Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020), Frontiers in Artificial Intelligence and Applications, Vol. 325, pp. 2505–2512, IOS Press, 2020.
URL https://doi.org/10.3233/FAIA200384

With the growing popularity of machine learning, the quest for verifying data-driven models is attracting more and more attention, and researchers in automated verification are struggling to meet the scalability and expressivity demands imposed by the size and the complexity of state-of-the-art machine learning architectures. However , there are applications where relatively small-scale learning models are enough to achieve industry-standard performances, yet the issue of checking whether those models are reliable remains challenging. Furthermore, in these domains, verification is just half of the game: providing automated ways to repair models that are found to be faulty is also an important task in practice. In this talk, I will touch upon some research directions that I have pursued in the past decade, commenting the results and providing some connections with related efforts. In particular we consider the following case studies:

- The problem of ensuring that a multi-agent robot control system is both safe and effective in the presence of learning components. In particular, we focus on a robot playing the air hockey game against a human opponent, where the robot has to learn how to minimize opponent's goals (defense play). This setup is paradigmatic since the robot must see, decide and move fastly, but, at the same time, it must learn and guarantee that the control system is safe throughout the process. We attack this problem using automata-theoretic formalisms and associated verification tools, showing experimentally that our approach can yield safety without heavily compromising effectiveness.
- Verification of Neural Networks known as Multi-Layer Perceptrons (MLPs), where we propose a solution to verify their safety using abstractions to Boolean combinations of linear arithmetic constraints. We show that our abstractions are consistent, i.e., whenever the abstract MLP is declared to be safe, the same holds for the concrete one. Spurious counterexamples, on the other hand, trigger refinements and can be leveraged to automate the correction of misbehaviors.

- Verification of Reinforcement Learning, a well-known AI paradigm whereby control policies of autonomous agents can be synthesized in an incremental fashion with little or no knowledge about the properties of the environment. We are concerned with safety of agents whose policies are learned by reinforcement, i.e., we wish to bound the risk that, once learning is over, an agent damages either the environment or itself. We propose a general-purpose automated methodology to verify, i.e., establish risk bounds, and repair policies, i.e., fix policies to comply with stated risk bounds. Our approach is based on probabilistic model checking algorithms and tools, which provide theoretical and practical means to verify risk bounds and repair policies. Considering a taxonomy of potential repair approaches tested on an artificially-generated parametric domain, we show that our methodology is also more effective than comparable ones.
- Verification of deep neural networks, particularly when it comes to enabel state-of-the-art verification tools to deal with neural networks of some practical interest. We propose a new training pipeline based on network pruning with the goal of striking a balance between maintaining accuracy and robustness, while also making the resulting networks amenable to formal analysis.

References

1. Stefano Demarchi, Dario Guidotti, Andrea Pitto, and Armando Tacchella. Formal verification of neural networks: A case study about adaptive cruise control. In Proceedings of the 36th ECMS International Conference on Modelling and Simulation, ECMS 2022, European Council for Modeling and Simulation, 2022.
2. Dario Guidotti, Francesco Leofante, Luca Pulina, and Armando Tacchella. Verification of neural networks: Enhancing scalability through pruning. In ECAI 2020 – 24th European Conference on Artificial Intelligence, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 2505–2512. IOS Press, 2020.
3. Dario Guidotti, Luca Pulina, and Armando Tacchella. pynever: A framework for learning and verification of neural networks. In *Automated Technology for Verification and Analysis* – 19th International Symposium, ATVA 2021
4. Francesco Leofante, Nina Narodytska, Luca Pulina, and Armando Tacchella. Automated verification of neural networks: Advances, challenges and perspectives. *CoRR*, abs/1805.09938, 2018.
5. Francesco Leofante, Simone Vuotto, Erika Ábrahám, Armando Tacchella, and Nils Jansen. Combining static and runtime methods to achieve safe standing-up for humanoid robots. In Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques – 7th International Symposium, ISoLA 2016
6. Shashank Pathak, Erika Ábrahám, Nils Jansen, Armando Tacchella, and Joost-Pieter Katoen. A greedy approach for the efficient repair of stochastic models. In NASA Formal Methods – 7th International Symposium, 2015
7. Shashank Pathak, Luca Pulina, Giorgio Metta, and Armando Tacchella. Ensuring safety of policies learned by reinforcement: Reaching objects in the presence of obstacles with the icub. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems
8. Shashank Pathak, Luca Pulina, and Armando Tacchella. Verification and repair of control policies for safe reinforcement learning. *Appl. Intell.*, 48(4):886–908, 2018.
9. Luca Pulina and Armando Tacchella. An abstraction-refinement approach to verification of artificial neural networks. In Computer Aided Verification, 2010.
10. Luca Pulina and Armando Tacchella. Never: a tool for artificial neural networks verification. *Ann. Math. Artif. Intell.*, 62(3-4):403–425, 2011.
11. Luca Pulina and Armando Tacchella. Challenging SMT solvers to verify neural networks. *AI Commun.*, 25(2):117–135, 2012.

4.23 Synthesizing Pareto-Optimal Interpretations for Black-Box Models

Hazem Torfah (University of California – Berkeley, US)

Joint work of Hazem Torfah, Shetal Shah, Supratik Chakraborty, S Akshay, Sanjit A. Seshia
Main reference Hazem Torfah, Shetal Shah, Supratik Chakraborty, S. Akshay, Sanjit A. Seshia: "Synthesizing Pareto-Optimal Interpretations for Black-Box Models", in Proc. of the Formal Methods in Computer Aided Design, FMCAD 2021, New Haven, CT, USA, October 19-22, 2021, pp. 153–162, IEEE, 2021.
URL https://doi.org/10.34727/2021/isbn.978-3-85448-046-4_24

We present a new multi-objective optimization approach for synthesizing interpretations that "explain" the behavior of black-box machine learning models. Constructing human-understandable interpretations for black-box models often requires balancing conflicting objectives. A simple interpretation may be easier to understand for humans while being less precise in its predictions vis-a-vis a complex interpretation. Existing methods for synthesizing interpretations use a single objective function and are often optimized for a single class of interpretations. In contrast, we provide a more general and multi-objective synthesis framework that allows users to choose (1) the class of syntactic templates from which an interpretation should be synthesized, and (2) quantitative measures on both the correctness and explainability of an interpretation. For a given black-box, our approach yields a set of Pareto-optimal interpretations with respect to the correctness and explainability measures. We show that the underlying multi-objective optimization problem can be solved via a reduction to quantitative constraint solving, such as weighted maximum satisfiability. To demonstrate the benefits of our approach, we have applied it to synthesize interpretations for black-box neural-network classifiers. Our experiments show that there often exists a rich and varied set of choices for interpretations that are missed by existing approaches.

4.24 Data Usage across the Machine Learning Pipeline

Caterina Urban (INRIA – Paris, FR)

In this talk, I give an overview of past and ongoing work in developing abstract interpretation-based static analyses for reasoning about data and input usage across the machine learning development pipeline. I present work targeting data processing software (Python and Jupyter Notebooks) or trained machine learning models (neural networks but also decision tree ensembles and support vector machines), as well as model training itself.

References

1 Caterina Urban, Peter Müller. *An Abstract Interpretation Framework for Input Data Usage*. In Proceedings of the 27th European Symposium on Programming (ESOP 2018).
2 Caterina Urban, Maria Christakis, Valentin Wüstholz, Fuyuan Zhang. *Perfectly Parallel Fairness Certification of Neural Networks*. In Proceedings of the ACM on Programming Languages, Volume 4, Number OOPSLA, 2020.
3 Denis Mazzucato, Caterina Urban. *Reduced Products of Abstract Domains for Fairness Certification of Neural Networks*. In Proceedings of the 28th Static Analysis Symposium (SAS 2021).

4.25 On The Unreasonable Effectiveness of SAT Solvers: Logic + Machine Learning

Vijay Ganesh (University of Waterloo, CA)

Joint work of Vijay Ganesh, Jimmy Liang, Ed Zulkoski, Krzysztof Czarnecki, Pascal Poupart
Main reference Jia Hui Liang, Vijay Ganesh, Pascal Poupart, Krzysztof Czarnecki: "Learning Rate Based Branching Heuristic for SAT Solvers", in Proc. of the Theory and Applications of Satisfiability Testing – SAT 2016 – 19th International Conference, Bordeaux, France, July 5-8, 2016, Proceedings, Lecture Notes in Computer Science, Vol. 9710, pp. 123–140, Springer, 2016.
URL https://doi.org/10.1007/978-3-319-40970-2_9

In this talk, we discuss a framework for viewing solver branching heuristics as optimization algorithms where the objective is to maximize the learning rate, defined as the propensity for variables to generate learnt clauses. By viewing online variable selection in SAT solvers as an optimization problem, we can leverage a wide variety of optimization algorithms, especially from machine learning, to design effective branching heuristics. In particular, we model the variable selection optimization problem as an online multi-armed bandit, a special-case of reinforcement learning, to learn branching variables such that the learning rate of the solver is maximized. We develop a branching heuristic that we call learning rate branching or LRB, based on a well-known multi-armed bandit algorithm called exponential recency weighted average.

5 Conclusion

In conclusion, this 5-day Dagstuhl seminar on topics at the intersection of machine learning and logic was very productive, enabling new collaborations and connections between researchers in the two camps of AI. We had over 25 talks that can be broadly categorized into the following four categories: 1) Neurosymbolic AI, 2) machine learning for solvers, 3) the testing, analysis, verification of machine learning systems, and 4) the use of machine learning in program synthesis. Key takeaways from the seminar included a greater need for intensification of collaboration between researchers in both camps, especially given the increasing importance of robust, secure, trustworthy, privacy-preserving, and interpretable AI. Most participants were very happy with the quality and diversity of the talks, the outcomes, new collaborations, and with our plan to continue organizing seminars in this series into the foreseeable future.

Participants

- Rajeev Alur
University of Pennsylvania – Philadelphia, US
- Sébastien Bardin
CEA LIST, FR
- Chih-Hong Cheng
Fraunhofer IKS – München, DE
- Jonathan Chung
University of Waterloo, CA
- Judith Clymo
University of Liverpool, GB
- Artur d'Avila Garcez
City – University of London, GB
- Alhussein Fawzi
Google DeepMind – London, GB
- Marc Fischer
ETH Zürich, CH
- Pascal Fontaine
University of Liège, BE
- Matt Fredrikson
Carnegie Mellon University – Pittsburgh, US
- Vijay Ganesh
University of Waterloo, CA
- Sebastian Junges
Radboud University Nijmegen, NL
- Chunxiao (Ian) Li
University of Waterloo, CA
- Ravi Mangal
Carnegie Mellon University – Pittsburgh, US
- Georg Martius
MPI für Intelligente Systeme – Tübingen, DE
- Kuldeep S. Meel
National University of Singapore, SG
- Grégoire Menguy
CEA LIST – Nano-INNOV, FR
- Matthew Mirman
ETH Zürich, CH
- Anselm Paulus
MPI für Intelligente Systeme – Tübingen, DE
- Markus N. Rabe
Google – Mountain View, US
- Joseph Scott
University of Waterloo, CA
- Xujie Si
McGill University – Montréal, CA
- Armando Tacchella
University of Genova, IT
- Hazem Torfah
University of California – Berkeley, US
- Caterina Urban
INRIA – Paris, FR
- Saranya Vijayakumar
Carnegie Mellon University – Pittsburgh, US

Remote Participants

- Somesh Jha
University of Wisconsin-Madison, US
- Luis C. Lamb
Federal University of Rio Grande do Sul, BR
- Vineel Nagisetty
Borealis AI – Toronto, CA
- Sanjit A. Seshia
University of California – Berkeley, US

Report from Dagstuhl Seminar 22292

Computational Approaches to Digitised Historical Newspapers

Maud Ehrmann*[1], Marten Düring*[2], Clemens Neudecker*[3], and Antoine Doucet*[4]

1 **EPFL – Lausanne, CH.** maud.ehrmann@epfl.ch
2 **University of Luxembourg, LU.** marten.during@uni.lu
3 **Staatsbibliothek zu Berlin, DE.** clemens.neudecker@sbb.spk-berlin.de
4 **University of La Rochelle, FR.** antoine.doucet@univ-lr.fr

Abstract

Historical newspapers are mirrors of past societies, keeping track of the small and great history and reflecting the political, moral, and economic environments in which they were produced. Highly valued as primary sources by historians and humanities scholars, newspaper archives have been massively digitised in libraries, resulting in large collections of machine-readable documents and, over the past half-decade, in numerous academic research initiatives on their automatic processing. The Dagstuhl Seminar 22292 "Computational Approaches to Digitised Historical Newspaper" gathered researchers and practitioners with backgrounds in natural language processing, computer vision, digital history and digital library involved in computational approaches to historical newspapers with the objectives to share experiences, analyse successes and shortcomings, deepen our understanding of the interplay between computational aspects and digital scholarship, and discuss future challenges. This report documents the program and the outcomes of the seminar.

Seminar July 17–22, 2022 – http://www.dagstuhl.de/22292

2012 ACM Subject Classification Computing methodologies → Information extraction; Computing methodologies → Machine learning; Information systems → Digital libraries and archives; Applied computing → Arts and humanities; Applied computing → Document management and text processing; Information systems → Information retrieval; Information systems → Data mining; Information systems → Document representation; Information systems → Document structure; Information systems → Structure and multilingual text search; Information systems → Users and interactive retrieval

Keywords and phrases historical document processing, document structure and layout analysis, natural language processing, information extraction, natural language processing, digital history, digital scholarship

Digital Object Identifier 10.4230/DagRep.12.7.112

* Editor / Organizer

Computational Approaches to Digitised Historical Newspapers, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 112–179
Editors: Maud Ehrmann, Marten Düring, Clemens Neudecker, and Antoine Doucet

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Maud Ehrmann (EPFL – Lausanne, CH)
Marten Düring (Luxembourg Centre for Contemporary and Digital History, LU)
Clemens Neudecker (Staatsbibliothek zu Berlin, DE)
Antoine Doucet (University of La Rochelle, FR)

Context

For long held on library and archive shelving, historical newspapers are undergoing mass digitisation and millions of facsimiles, along with their machine-readable content captured via optical character recognition (OCR), are becoming accessible via a variety of online portals.[1] While this represents a major step forward in terms of preservation and access, it also opens up new opportunities and poses timely challenges for both computer scientists and humanities scholars [2, 1, 3, 14].

As a direct consequence, the last ten years have seen a significant increase of academic research on historical newspaper processing. In addition to decisive grassroots efforts led by libraries to improve OCR technology,[2] individual works dedicated to the development and application of tools to digitised newspaper collections have multiplied [12, 13, 10, 11], as well as events such as evaluation campaigns or hackathons [8, 7, 6, 5].[3] Besides, several large consortia projects proposing to apply computational methods to historical newspapers at scale have recently emerged, including ViralTexts[4], Oceanic Exchanges[5], *impresso* – Media Monitoring of the Past[6], NewsEye[7], Living with Machines[8], and Data-kbr-be[9] [9].

This momentum can be attributed not only to the long-standing interest of scholars in newspapers coupled with their recent digitisation, but also to the fact that these digital sources concentrate many challenges for computer science, which are all the more difficult – and interesting – since addressing them requires taking digital (humanities) scholarship needs and knowledge into account. Within interdisciplinary frameworks, various and complementary approaches spanning the areas of natural language processing, computer vision, large-scale computing and visualisation, are currently being developed, evaluated and deployed. Overall, these efforts are contributing a pioneering set of tools, system architectures, technical infrastructures and interfaces covering several aspects of historical newspaper processing and exploitation.

[1] Such as those discussed in [16] and by this seminar's working group on transparency and fairness, see Section 4.3 hereafter.
[2] See e.g., the OCR-D project, an ecosystem for improving OCR on historical documents: `https://ocr-d.de/en/about` and [4].
[3] See the 2017 edition of the Coding Da Vinci cultural hackathon or the 2019 edition of the Helsinki Digital Humanities Hackathon.
[4] A project aiming at mapping networks of reprinting in 19th-century newspapers and magazines (US, 2012-2016): `https://viraltexts.org`
[5] A project tracing global information networks in historical newspaper repositories from 1840 to 1914 (US/EU, 2017-2019): `https://oceanicexchanges.org`
[6] A project which tackles the challenge of enabling critical text mining of newspaper archives (CH, 2017-2020): `https://impresso-project.ch`
[7] A digital investigator for historical newspapers (EU, 2018-2022): `https://www.newseye.eu`
[8] A project which aims at harnessing digitised newspaper archives (UK, 2018-2023): `https://www.turing.ac.uk/research/research-projects/living-machines`
[9] A project which aims at facilitating data-level access to KBR's collections (mainly newspapers) for open science (2020-2022):`https://www.kbr.be/en/projects/data-kbr-be/`.

Objectives

The aim of the seminar was to bring together researchers and practitioners involved in computational approaches to historical newspapers to share experiences, analyse successes and shortcomings, deepen our understanding of the interplay between computational aspects and digital scholarship, and begin to design a road map for future challenges. Our seminar was guided by the vision of methodologically reflected, competitive and sustainable technical frameworks capable of providing an extensive, sophisticated and possibly dynamic access to the content of digitised historic newspapers in a way that best serves the needs of digital scholarship. We are convinced that in order to meet the many challenges of newspaper processing and to accommodate the demands of humanities scholars, only a global and interdisciplinary approach that looks beyond technical solutionism and embraces the complexity of the source and its study can really move things forward.

Participants and Organisation

The seminar gathered 22 researchers[10] with backgrounds in natural language processing, computer vision, digital history and digital library, the vast majority of whom had previously worked on historical newspapers and were familiar with interdisciplinary environments. To structure and coordinate the work of the seminar, the organisers proposed a mixture of plenary sessions, working groups, and talks, as follows (see also Fig.1):

- **Spotlight talks** on day 1, where each participant briefly introduced him/herself and gave an opinion or statement on his/her current view of the main topic of the seminar (3-minute/1-slide).
- **Demo session**, where some participants shortly introduced a relevant asset (e.g., a dataset, tool, interface, on-going experiment).
- **Working group sessions**, during which groups composed of computer scientists and humanities scholars focused on a specific question. Work within a group featured different moments, with: expert group discussion, where people with similar backgrounds exchanged in order to align their understanding of the question at hand and to prioritise problems; observation of concrete research and workflow practices on existing approaches and/or tools; cross-interviews, where people from one domain interviewed one person from another domain about a specific point; mixed group discussion, where everybody jointly reflected; and writing time, where the group wrote a report summarising its findings.
- **Reporting sessions**, where working groups reported their discussion and presented their main conclusions and recommendations in a plenary session.
- **Morning presentations**, where researchers shared their experience from a project and/or their view on a specific topic, followed by a discussion with the participants.
- **Evening talks**, where researchers shared their experience and views on a topic at large. We proposed three evening lectures that addressed the field of digitised newspapers from the perspective of computer science, digital history, and digital libraries.

[10] Some participants had to cancel at the last moment due to the pandemic; we thank them for their initial commitment and hope that there will be future opportunities.

<table>
<tr><th></th><th>Monday</th><th>Tuesday</th><th>Wednesday</th><th>Thursday</th><th>Friday</th></tr>
<tr><td>9:00 - 10:00</td><td rowspan="3">- General introduction
- Spotlight talks (ca. 22*3min) with coffee break
- Topics presentation and expressions of interest
- Identification of groups
(please refer to the Wiki for detailed schedule)</td><td>Morning Talks (2* 20+10min)</td><td rowspan="2">Demo Session</td><td>Group Reporting</td><td rowspan="2">Group Reporting & Working Group Session</td></tr>
<tr><td>10:00 - 11:00</td><td>Group Reporting (30')</td><td rowspan="2">Working Group Session</td></tr>
<tr><td>11:00 - 12:00</td><td>Working Group Session</td><td>Working Group Session</td><td>Concluding session</td></tr>
<tr><td>12:15 - 13:30</td><td colspan="4">lunch</td><td>end of seminar</td></tr>
<tr><td>13:30 - 15:30</td><td>Room assignment and working group session</td><td>Working group session</td><td rowspan="5">excursion</td><td>Working Group Session</td><td rowspan="5"></td></tr>
<tr><td>15:30 - 16:00</td><td>coffee break</td><td>coffee break</td><td>coffee break</td></tr>
<tr><td>16:00 - 17:30</td><td rowspan="2">Working Group Session</td><td rowspan="2">Working Group Session</td><td rowspan="2">Working Group Session</td></tr>
<tr><td>17:30 - 18:00</td></tr>
<tr><td>18:00 - 19:00</td><td>dinner</td><td>dinner</td><td>dinner</td></tr>
<tr><td>20:15 - 21:15</td><td>Evening talk (computer science perspective)</td><td>Evening talk (digital history perspective)</td><td>outside dinner</td><td>Evening talk (digital library perspective)</td></tr>
</table>

Figure 1 Schedule of the seminar.

Topics

The topics and modus operandi of the seminar were not set in stone but discussed and validated with all participants during the first day. First, the organisers proposed three main topics (and several corresponding sub-questions) for the participants to discuss and reflect on during the seminar. On this basis, participants were then invited to express the specific themes, questions and issues they wished to work on, in a traditional post-it session. Finally, these propositions were examined and structured by the organisers, who defined four working groups.

Proposed topics

As a starting point, organisers proposed to consider three closely intertwined topics, which are detailed below to further illustrate the background knowledge of this seminar.

1. **Document Structure and Text Processing.** While recent work on the semantic enrichment of historical newspapers has opened new doors for their exploration and data-driven analysis from a methodological perspective (e.g., n-grams, culturomics), results up to now often confirmed common knowledge and were not always considered relevant by historians. The next natural and eagerly-awaited step consists in enriching newspaper contents and structure with semantic annotations which allow for the exploration of far more nuanced research questions. In this regard, several issues arise, among others:
   - **Q1.1** – ***Complex structures and heterogeneity of contents.*** Newspapers are typically composed of a diverse mix of content including text, image/graphical elements, as well as tabular data and various other visual features. The proper segmentation of the page content into individual information pieces is key for enabling advanced research and analysis. This includes the modelling and detection of logical units on the document (or specifically, issue) level as e.g. articles can span across multiple pages. Also of high relevance to researchers is the more advanced classification and semantic labelling of content units, separating categories such as information, opinion,

stock market indices, obituaries, humour, etc. Despite a growing interest, a good understanding of these complex structures as well as methods and technologies for identifying, classifying and accessing diverse content types through appropriate data models and search interfaces are still lacking.

- **Q1.2** – ***Diachronic processing.*** Historical newspaper material poses severe challenges for computational analysis due to their heterogeneity and evolution over time. At language level, besides historical spelling variation which leads to major problems in text recognition and retrieval, sequential labelling tasks such as named entity recognition and disambiguation are problematic and often require time-specific resources and solutions. At document level, text classification or topic modelling need to pay attention to the necessary historical contextualisation of their category schemes or corpus time-spans in order to avoid anachronisms. Finally, at structure level, layout processing faces similar challenges and its application needs to adapt to changing sources.

2. **Visualisation, Exploitation and Digital Scholarship.** Historians and other user groups require tools for content discovery and management to reflect their iterative, exploratory research practices. The opportunities and challenges posed by mass digitised newspapers and other digitised sources require them to adjust their current workflows and to acquire new skills.
   - **Q2.1** – ***Transparency and digital literacy.*** In the context of research, humanists' trust in computer systems is dependent on sufficient comprehension of the quality of the underlying data and the performance of the tools used to process it. One way to generate such trust is to create transparency, here understood as: information on the provenance and quality of digital sources; information which allows users to make informed decisions about the tools and data they use; and information which allows their peers to retrace their steps. Such transparency empowers users to use the system in a reflective way. But there is to-date no shared understanding of which information exactly is required to achieve transparency: technical confidence scores are themselves hard to interpret and do not translate easily into actionable information. Likewise, historians can not be expected to be aware of the consequences of all the algorithmic treatments to which their digitised sources have been exposed. Instead, the identification of the most relevant biases and their concrete consequences for users appears to be a more realistic approach. Once these are understood, counter-action can be taken.
   - **Q2.2** – ***Iterative content discovery and analysis.*** In contrast to many other applications in computer science, the discovery of relevant content is of greater interest to historians than the detection of patterns in datasets following a priori hypotheses. Historical research is typically iterative: The study of documents yields new insights which determine future exploration strategies and allow scholars to reassess the value of the sources they have consulted. Semantically enriched content offers multiple ways to support this iterative exploration process. New tools for content discovery also require "generous" interfaces, i.e. interfaces which allow users to discover content rather than relying on narrow keyword search [15].
3. **System Architecture and Knowledge Representation.** The application of various natural language processing and computer vision components which transform noisy and unstructured material into structured data also poses challenges in terms of system architecture and knowledge representation. If those two well-studied fields already offer a strong base to build upon, many questions arise from newspaper source specificities and

the digital humanities context.

- **Q3.1** – ***Managing provenance, complexity, ambivalence and vagueness.*** Lots of factual and non factual information is extracted from newspaper material and need to be stored and interlinked. In this regard, two points require great attention. First, newspapers – like any other historical source – represent past realities which do not necessarily align with present-day realities: institutions and countries change names or merge, country borders move or become disputed and professions change or disappear. These temporal shifts, ambivalences and contradicting information cause historical data to be highly complex and sometimes disputed, and the representation of this complexity poses interesting challenges for computer science. Second, if processing steps, and possibly intermediary representations, of algorithms are recorded for the purpose of transparency, this meta-knowledge need to be stored alongside the data.
- **Q3.2** – ***Dynamic processing.*** Historical newspaper processing outputs are useless if not used by scholars who wish to investigate research questions. If all methods and practices can not be transposed as they stand from analogue to digital, careful consideration must be given to how best to accommodate scholarship requirements in digital environments where primary sources are turned into data.

Selected Topics and Working Groups

The discussion around topics led to the definition of four working groups which the participants joined on the first day (on a voluntary basis) and in which they worked throughout the week. No guidelines were given and the groups were free to adapt the direction of their work. Each group wrote a report summarising their activities and findings in Section 4.

1. **Working Group on Information Extraction.** Initiated around the topic of information extraction, this group eventually settled on the specific topic of person entity mentions found in historical newspapers but not present in knowledge bases, a.k.a "hidden people". The group defined a number of challenges and worked – in a productive hackathon style – on several experiments (see Section 4.1).
2. **Working Group on Segmentation and Classification.** Members of this group quickly discarded the segmentation question to focus on classification only, considering classification scope and practices in relation to digitised newspapers (see Section 4.2).
3. **Working Group on Transparency, Critics and Newspapers as Data.** This group (the largest) worked on a set of recommendations regarding the different aspects of transparency and fairness needed for the analysis of digitised and enriched historical newspaper collections (see Section 4.3).
4. **Working Group on Infrastructure and Interoperability.** This group discussed the issue of consolidation, growth and sustainability of current and future achievements in digitisation, access, processing and exchange of historical newspapers (see Section 4.4).

Spotlight Talks on the Main Challenges Ahead for Digitised Historical Newspapers

On the first morning of the seminar, the organisers asked participants to briefly present their views on some questions they had to consider in advance. These questions were:

- What are the main challenges we need to address in relation with historical newspapers?
- What is the most exciting opportunity you would like to explore during this seminar?

- If you were given €1 million to spend in the next 6 months on historical newspapers, what would you do?

As well as being a good ice-breaker and kick-off to the seminar, the series of responses to these questions documents what a community of researchers in July 2022 believe to be the next challenges for computational approaches to historical newspapers. In total, 21 researchers formulated no less than 67 statements as responses to the first question. In what follows, we provide a summary of the main ideas and suggestions which we have grouped in 8 themes that cover more or less the whole spectrum of activities around digitised newspapers. Apart from this grouping, no further reflection or refactoring has been done on these statements. While most of the answers are not a surprise to those familiar with the subject, they confirm existing needs, reflect on-going trends, and reveal new lines of research.

► **Document processing.** A first group of statements relates to optical character recognition and optical layout recognition (OLR), two critical processes when working with newspapers. These two document image refinement techniques are extremely difficult when applied to such sources (especially for collections digitised long ago), which explains why they are still high on the agenda despite all the efforts invested in recent years. The views expressed highlight and confirm several dimensions, namely: OCR and OLR quality needs to be improved, finer-grained segmentation and classification of news items is necessary, and processes should be more robust across time and collections. Intensive work is being carried out in these areas.

Verbatim statements:

- *How to make available digitised newspaper collections in high quality OCR+layout* (Clemens Neudecker);
- *High quality article segmentation and classification* (Maud Ehrmann, Mickaël Coustaty);
- *The (massive) segmentation bottleneck* (Antoine Doucet);
- *Robust layout recognition* (Matteo Romanello);
- *A level playing field: OCR and fine-grained content segmentation* (Marten Düring);
- *Improving article segmentation (e.g., wrt advertisements and classifieds)* (Mariona Coll-Ardanuy);
- *Layout recognition (e.g., article separation, recognition of headings and authors' names)* (Dario Kampkaspar);
- *OCR+, layout recognition, article segmentation and classification* (Eva Pfanzelter);
- *Better article segmentation, and a way to deal with heterogeneous qualities of segmentation in DL* (Axel Jean-Caurant);
- *Quality of OCR across periods, languages and original document qualities* (Yves Maurer);
- *Standardised approaches to segment historical newspaper pages* (Stefan Jänicke).

► **Text and image processing.** This group encompasses all types of content processing applied to OCR and OLR outputs in view of enriching newspaper contents with further information, usually in the form of semantic annotations and item classification. The main challenges that emerge are: robustness (i.e. approaches that perform well on challenging, noisy input), finer-grained information extraction, few-short learning (to compensate lack of training data), transferability (approaches that perform well work across settings), multilinguality, multimodality, entity linking, interlinking of collections, and transmedia approaches.

Verbatim statements:

- *Robust multilingual information extraction* (Maud Ehrmann);
- *Developing methods that are robust to OCR errors* (Mariona Coll-Ardanuy);

- *Words with meaning change over time* (Martin Volk);
- *Text summarisation and text classification (monolingual and across languages)* (Martin Volk);
- *How to automatically detect genres (in particular film reviews and film listings)* (Julia Noordegraaf);
- *Multilinguality* (Eva Pfanzelter);
- *Ease multilingual scholarship (qualitative and quantitative)* (Antoine Doucet);
- *Investigating the relation between (or intertwining of) image and text* (Kaspar Beelen);
- *Embedding newspaper content within the media landscape* (Kaspar Beelen);
- *Data mining in newspapers (e.g. biographies, TV/radio programmes)* (Marten Düring);
- *Robust entity linking (multilingual historical documents)* (Matteo Romanello);
- *Entity Linking and visualisations over time, space and networks* (Martin Volk);
- *Linking with other data sources (parliamentary protocols, wiki-data, (historic) names-db, other newspaper portals, (historic) place names db, etc.)* (Eva Pfanzelter);
- *Create links between newspaper contents (topics, entities) and knowledge bases* (Simon Clematide);
- *Automate content analysis (discourses, argumentation, events, meaning, topics) to enable historical research* (Eva Pfanzelter);
- *Learn with few samples and human interactions* (Mickaël Coustaty).

▶ **Digitisation and Content Mining Evaluation.** Here we have grouped together views on the evaluation of technical approaches and tools at large, and the means to implement it. Important points that emerge are: better metrics, more and diverse gold standards, and better contextualisation and understanding of (sources of) errors.

Verbatim statements:

- *How to arrive at common methods and metrics for quality of digitised newspapers* (Clemens Neudecker);
- *Sustainably sharing ground truth datasets and training models* (Sally Chambers);
- *Developing a variety of NLP benchmarks for different tasks across different languages and types of publications* (Mariona Coll-Ardanuy);
- *Build a general taxonomy of content items (including ads, service tables, etc.) and prepare well-sampled data sets from a variety of publication places and time periods* (Simon Clematide);
- *Disentangling correlation of errors and missingness with time, place, language, network position, etc.* (David Smith).

▶ **Exploration of (enriched) newspaper collections and beyond.** One of the opportunities that researchers have been working on in recent years is new ways of exploring newspaper content. This group of statements is part of this context and highlights some of the long-awaited next steps: unified access to newspaper collections, support for data-driven research, and connection to other archives.

Verbatim statements:

- *Access to newspaper content across collections/projects/platforms* (Matteo Romanello);
- *Unified access to all collections with advanced exploration capacities* (Maud Ehrmann);
- *A unified framework to REALLY make collections accessible, usable and interoperable* (Antoine Doucet);
- *Access across collections and copyright hurdles* (Marten Düring);
- *Silos* (Jana Keck);

- *Data-driven linking and analysis of multiple types of sources (e.g. radio, TV, parliamentary records) and datasets (e.g. land ownership, migration)* (Marten Düring);
- *User-driven (from novice to expert) image, information and metadata etc. extraction* (Eva Pfanzelter);
- *Offer our users more than search, but what ? Topics, recommenders, . . . ?* (Yves Maurer);
- *Contextual information extracted from the corpus: hints on rubrics, themes, top keyword per month etc.* (Estelle Bunout);
- *Comparative analyses of contents (political targets of publishers), ordering of articles and time-based development of topics* (Stefan Jänicke).

▶ **Working with data.** In addition to working with enriched sources that can be semantically indexed and thus more easily retrieved and analysed, researchers (especially historians) also express the wish to work directly with raw data – digitised documents, annotations, or both – and be able to build their own corpora.

Verbatim statements:

- *How to create useful datasets and corpora from digitised newspapers* (Clemens Neudecker);
- *Availability of digitisation output (images, text) for further use (beyond interfaces)* (Estelle Bunout);
- *Newspapers as Data: how to facilitate dataset / corpus building* (Sally Chambers);
- *Ease the building and sharing of corpora, taking into account the context of creation (queries, quality, etc)* (Axel Jean-Caurant).

▶ **Collections, source and tool criticism; Documentation; Inclusivity.** The validity of any conclusions drawn in empirical research depends on a solid understanding of the data used for the analysis. Digitised and enriched newspapers contain multiple levels of processing which often vary significantly across titles in terms of processing quality and extent of enrichment. The statements below point to key challenges and opportunities to advance our reflected analysis of digitised newspapers.

Verbatim statements:

- *How to support and perform source criticism on digitised newspaper collections* (Julia Noordegraaf);
- *Methodological guidelines for the computational analysis of newspaper content* (Julia Noordegraaf);
- *Describing how biases arise in digitised newspapers collections ("full-stack bias")* (Kaspar Beelen);
- *Understanding how structured missingness and data quality affect (historical) research* (Kaspar Beelen);
- *Selection criteria guidelines for what is being selected, digitised, accessible and how it is represented, searchable, and available* (Jana Keck);
- *Trustable and/or understandable approaches to meet users' needs* (Mickaël Coustaty);
- *How do we make collections as well as access mechanisms inclusive?* (Laura Hollink);
- *How do we monitor fairness of computational approaches to historic newspapers?* (Laura Hollink);
- *How well does the collection support different user groups?* (Laura Hollink);
- *How do we make perspectives in the data explicit? (e.g in NL context: words signalling a colonial perspective)* (Laura Hollink);
- *Information on the scope, contents and quality of a collection, e.g., included titles, covered time periods, granularity of items (page vs. article), OCR quality, corpus statistics* (Estelle Bunout);
- *Investigate the role of attributes like font face and style, margins, layout, paper, etc.* (Stefan Jänicke);

- *Book-historical studies of editorship and publishing (costs, layout, format, advertising, syndicates, networks) crossing national and cataloguing (newspaper/magazine) boundaries* (David Smith) ;
- *Investigate the role of attributes like font face and style, margins, layout, paper, etc.* (Stefan Jänicke).

▶ **Workflows.** The combination of multiple processes, moreover between different actors, requires the design of more standardised and efficient workflows encompassing the many processing steps that have emerged in recent years.

Verbatim statements:

- *Advanced digitisation workflows: from digitisation to OCR to article segmentation* (Sally Chambers);
- *Workflows that conflate search, annotation, classification, corpus construction* (David Smith).

▶ **Legal matters.** Finally, a last set of (unavoidable) challenges concerns legal issues, with questions of copyright clearance and management, and of personal data, whether it be user data handled by platforms, or the right to be forgotten.

Verbatim statements:

- *Find sustainable ways to work with copyright-restricted data sets* (Yves Maurer);
- *Access across collections and copyright hurdles* (Marten Düring);
- *Copyrights and proprietary rights, image rights etc.* (Eva Pfanzelter);
- *Legal questions (copyright, personal rights, etc.)* (Dario Kampkaspar).

Acknowledgements

This seminar was originally planned for September 2020 but was cancelled due to the COVID-19 pandemic and rescheduled 2 years later. We would like to thank the administrative and scientific teams at Dagstuhl for their support and professionalism throughout the (re)organisation of this seminar as well as the staff on site for their valuable every day help and care. We also thank all the participants for accepting our invitation to spend a week exchanging views, examining, questioning, debating (and writing) about computational approaches to historical newspapers.

References

1 Bingham, A. The Digitization of Newspaper Archives: Opportunities and Challenges for Historians. *Twentieth Century British History*.21, 225-231 (2010,6)
2 Deacon, D. Yesterday's Papers and Today's Technology: Digital Newspaper Archives and "Push Button" Content Analysis. *European Journal Of Communication*. 22, 5-25 (2007)
3 Nicholson, B. The Digital Turn. *Media History*. 19, 59-73 (2013,2)
4 Neudecker, C., Baierer, K., Federbusch, M., Boenig, M., Würzner, K., Hartmann, V. & Herrmann, E. OCR-D: An End-to-End Open Source OCR Framework for Historical Printed Documents. *Proceedings Of The 3rd International Conference On Digital Access To Textual Cultural Heritage*. pp. 53-58 (2019,5)
5 Ehrmann, M., Romanello, M., Najem-Meyer, S., Doucet, A. & Clematide, S. Extended Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. *Proceedings Of The Working Notes Of CLEF 2022 – Conference And Labs Of The Evaluation Forum*. 3180 (2022) https://infoscience.epfl.ch/record/295816

6 Ehrmann, M., Romanello, M., Flückiger, A. & Clematide, S. Extended Overview of CLEF HIPE 2020: Named Entity Processing on Historical Newspapers. *Working Notes Of CLEF 2020 – Conference And Labs Of The Evaluation Forum*. 2696 pp. 38 (2020)

7 Clausner, C., Antonacopoulos, A., Pletschacher, S., Wilms, L. & Claeyssens, S. PRImA, DMAS2019, Competition on Digitised Magazine Article Segmentation (ICDAR 2019). (2019), `https://www.primaresearch.org/DMAS2019/`

8 Rigaud, C., Doucet, A., Coustaty, M. & Moreux, J. ICDAR 2019 Competition on Post-OCR Text Correction. *2019 International Conference On Document Analysis And Recognition (ICDAR)*. pp. 1588-1593 (2019)

9 Ridge, M., Colavizza, G., Brake, L., Ehrmann, M., Moreux, J. & Prescott, A. The Past, Present and Future of Digital Scholarship with Newspaper Collections. *DH 2019 Book Of Abstracts*. pp. 1-9 (2019), `http://infoscience.epfl.ch/record/271329`

10 Kestemont, M., Karsdorp, F. & Düring, M. Mining the Twentieth Century's History from the Time Magazine Corpus. *Proceedings Of The 8th Workshop On Language Technology For Cultural Heritage, Social Sciences, And Humanities (LaTeCH)*. pp. 62-70 (2014)

11 Lansdall-Welfare, T., Sudhahar, S., Thompson, J., Lewis, J., Team, F. & Cristianini, N. Content Analysis of 150 Years of British Periodicals. *Proceedings Of The National Academy Of Sciences*. **114**, E457-E465 (2017)

12 Moreux, J. Innovative Approaches of Historical Newspapers: Data Mining, Data Visualization, Semantic Enrichment. *IFLA News Media Section, Lexington, August 2016, At Lexington, USA*. (2016,8), `https://hal-bnf.archives-ouvertes.fr/hal-01389455`

13 Wevers, M. Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990. *Proc. Of The 1st International Workshop On Computational Approaches To Historical Language Change*. (2019), `https://www.aclweb.org/anthology/W19-4712`

14 Bunout, E., Ehrmann, M. & Clavert, F. (editors) Digitised Newspapers – A New Eldorado for Historians? Tools, Methodology, Epistemology, and the Changing Practices of Writing History in the Context of Historical Newspaper Mass Digitisation. De Gruyter (2022, in press), doi:10.1515/9783110729214, `https://www.degruyter.com/document/isbn/9783110729214/html`

15 Whitelaw, M. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*. **9** (2015), `http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html`

16 Ehrmann, M., Bunout, E. & Düring, M. Historical Newspaper User Interfaces: A Review. *Proceedings Of The 85th International Federation Of Library Associations And Institutions (IFLA) General Conference And Assembly*. pp. 24 (2019), `https://infoscience.epfl.ch/record/270246?ln=en`

2 Table of Contents

3 Overview of Talks

3.1 Memoirs of Extraordinary Popular Delusions

David A. Smith (Northeastern University – Boston, US)

Joint work of David A. Smith, Ryan Cordell

Newspapers present an extraordinary window into modern language, history, and culture and a revealing mode of information production. By tracing how texts are exchanged, edited, composed, laid out, and generally reprinted, we can learn about historical communications, political, social, and transportation networks. The sample of newspapers digitised by most projects, however, is not representative of all aspects of the historical population. We can correct for this mismatch using regression on observed features of undigitised papers from catalogues and historical directories. Finally, we can use the structure of text reprinting to correct and retrain transcription and layout analysis models.

3.2 Living with Machines: Exploring Newspapers at Scale

Kaspar Beelen (The Alan Turing Institute – London, GB)
Mariona Coll-Ardanuy (The Alan Turing Institute – London, GB)

Joint work of The Living with Machines Project
URL https://livingwithmachines.ac.uk/team-2/

Living with Machines (LwM) is an interdisciplinary research project focused on the lived experience of Britain's industrialisation during the long nineteenth century (roughly 1780 to 1918). The project develops approaches made possible by rapid digitisation and computational methods, to analyse and link historical records. In this presentation, we focused on our work on historical newspapers. The digitised press provides an immense amount of varied, fine-grained, and often neglected information. However large and rich, newspapers remain difficult to navigate as existing tools struggle with the particularities of digitised historical data.

News was often about place, its discourse anchored in space. In our talk, we show that historically sensitive methods improve performance for both toponym recognition and resolution. Moreover, news content was embedded in social and historical contexts. We presented the "Environmental Scan' which explores questions of representativeness and bias based on insights derived from contemporaneous reference sources about the press such as newspaper press directories.

References

1 Ardanuy, M., Beavan, D., Beelen, K., Hosseini, K., Lawrence, J., McDonough, K., Nanni, F., Strien, D. & Wilson, D. A Dataset for Toponym Resolution in Nineteenth-Century English Newspapers. *Journal Of Open Humanities Data*. **8** pp. 3 (2022,1)
2 Beelen, K., Lawrence, J., Wilson, D. & Beavan, D. Bias and Representativeness in Digitized Newspaper Collections: Introducing the Environmental Scan. *Digital Scholarship In The Humanities*. https://doi.org/10.1093/llc/fqac037 (2022,7)

3 Hosseini, K., Nanni, F. & Coll Ardanuy, M. DeezyMatch: A Flexible Deep Learning Approach to Fuzzy String Matching. *Proceedings Of The 2020 Conference On Empirical Methods In Natural Language Processing: System Demonstrations*. pp. 62-69 (2020,10)

3.3 NLP on Historical Documents: Experience (from impresso), Challenges, Opportunities

Simon Clematide (Universität Zürich, CH)

URL https://impresso-project.ch
Joint work of Impresso project team: Maud Ehrmann, Marten Düring, Simon Clematide, Matteo Romanello, Estelle Bunout, Daniele Guido, Philipp Ströbel, Roman Kalyakin, Lars Wieneke, Andreas Fickers, Martin Volk, Frédéric Kaplan

In this talk we discussed the application of NLP techniques on historical newspapers in light of the *impresso* project experience. In a nutshell, the project '*impresso* – Media Monitoring of the Past' tried to answer the question of how best to accommodate text analysis research tools and their usage by humanities scholars. Using a co-design approach involving text miners, UX/UI designers and historian users, we worked on bringing the content of digitised newspaper silos – often consisting of big messy data – into an interface that allows search, exploration and visualisation of the texts and their semantic enrichment[11]. NLP and text mining techniques involve basic IR keyword indexing, OCR improvements (lately visual transformers such as TrOCR) and language identification. Although not rocket science, it has to be done carefully to keep all downstream processing language-aware. Data-driven word embeddings built on the historical corpora help organise the semantic space and support query expansion. Keyphrase extraction uses NLP and word embeddings to summarise content items in the most concise terms. Topic modelling clusters the documents into topic distributions and serves as a search filter and recommender backbone. Lastly, named entities are indexed and linked to Wikidata. Their distribution in the corpus is highly informative for historians. Being able to represent linguistic elements from words to terms, sentences, documents as comparable vectors in multilingually aligned vector spaces will enable semantic search in the future.

References

1 Ehrmann, M., Romanello, M., Clematide, S., Ströbel, P. & Barman, R. Language Resources for Historical Newspapers: The Impresso Collection. *Proceedings of The 12th Language Resources And Evaluation Conference*. pp. 958-968 (2020,5), `https://www.aclweb.org/anthology/2020.lrec-1.121`

2 Romanello, M., Ehrmann, M., Clematide, S. & Guido, D. The Impresso System Architecture in a Nutshell. *EuropeanaTech Insights*. (2020), `https://pro.europeana.eu/page/issue-16-newspapers#the-impresso-system-architecture-in-a-shell`, `https://infoscience.epfl.ch/record/283595`

3 Ehrmann, M., Düring, M., Clematide, S., Romanello, M., Bunout, E., Guido, D., Ströbel, P., Kalyakin, R., Wieneke, L., Fickers, A., Volk, M. & Kaplan, F. *Impresso*: Historical Newspapers Beyond Keyword Search. *(in preparation)*.

[11] `https://impresso-project.ch/app/`

3.4 Integrating Computational Processing into Historical Research and the Steps Leading to it.

Estelle Bunout (Luxembourg Centre for Contemporary and Digital History, LU)

The efforts in digitising newspapers have been massive in the past decades and many experiences have been gathered by humanists and libraries using and publishing these collections. While the diversity of formats in which their content is accessible remains high, engaged users can collect different and complementary information from different sources. Humanists and in particular historians, need access to the source material but also to contextual information, often provided in form of metadata. The issue relies most commonly not so much in the lack of experience in producing relevant metadata – being derived from catalogue information or content-based computed metadata, but rather in lack of wide habit of using them by researchers and the tendency for some untypical metadata to be incomplete or unstable. Nevertheless, once access and basic contextual information has been made available, many options of operationalising humanities research questions with the support of computational tools, most commonly text mining, open up. In the talk, a case was presented of using a naive Bayes classifier to look for similar texts to anti-modern articles published in Swiss interwar press. This study highlights the need for contextualisation of findings to interpret them properly.

References

1 Bunout, E. & Düring, M. Collections of Digitised Newspapers as Historical Sources – Parthenos Training. (2019), `https://training.parthenos-project.eu/sample-page/digital-humanities-research-questions-and-methods/collections-of-digital-newspapers-as-historical-sources/`

3.5 Newspapers as Data: Challenges and Solutions

Sally Chambers (Ghent University, BE & KBR, Royal Library of Belgium, Brussels, BE)

Digital cultural heritage collections in libraries, archives, and museums are increasingly being used for digital humanities research. However, traditional ways of providing access to such collections, for example through digital library interfaces, are less than ideal for researchers who are looking to build datasets around specific research questions. Originating in the United States, the "Collections as Data" movement was established to encourage cultural heritage professionals to start thinking differently about how they provide access to their collections to facilitate analysis using digital tools and methods. "Collections as Data" encourages the provision of "data-level access" to the underlying files of digitised and born-digital cultural heritage resources to facilitate data analysis by means of tools and methods developed in the field of digital humanities. This presentation explores whether the application of "Collections as Data" to digitised historical newspapers could help facilitate corpus building for digital humanities research.

References

1 Padilla, T., Allen, L., Frost, H., Potvin, S., Roke, E. & Varner, S. Final Report – Always Already Computational: Collections as Data. *Zenodo. Texas Digital Library.* **10** (2019), https://zenodo.org/record/3152935

2 Chambers, S., Lemmers, F., Pham, T., Birkholz, J., Ducatteeuw, V., Jacquet, A., Dillen, W., Ali, D., Milleville, K. & Verstockt, S. Collections as Data : Interdisciplinary Experiments with KBR's Digitised Historical Newspapers : A Belgian Case Study. *DH Benelux 2021, Abstracts.* (2021), http://hdl.handle.net/1854/LU-8712404

3 Tasovac, T., Chambers, S. & Tóth-Czifra, E. Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper. (2020,10), https://hal.archives-ouvertes.fr/hal-02961317

4 Working groups

4.1 Tracking Discourses on Public and Hidden People in Historical Newspapers

Simon Clematide (Department of Computational Linguistics, Switzerland, siclemat@cl.uzh.ch)
Mariona Coll-Ardanuy (The Alan Turing Institute, UK, mcollardanuy@turing.ac.uk)
Yves Maurer (National Library of Luxembourg, Luxembourg, yves.maurer@bnl.etat.lu)

This working group focused on information extraction on historical newspapers. Following an initial brainstorming session with the whole seminar group, we decided to narrow down the topic of information extraction to the more specific question of the coverage of public (notable) and hidden (non-notable) people in newspapers.[12] From the historian's perspective, it is difficult to identify and study groups of people who are not already notable for something, and work has been undertaken to develop approaches to represent the under-represented [2]. During the seminar, we built on the idea (and hope) that newspapers could serve as a source of stories about and insights into the lives of people who are not in the public eye. Given the group's interest in working with real datasets and tools, we agreed to conduct actual experiments in the form of a case study to understand if our questions could be addressed in a practical way.

4.1.1 Discussed Problems

Our overarching research question is whether computational methods can help shed some light on how newspapers report on *public* versus *hidden* people. Our aim was to explore ways to address this question computationally. To do so, we broke it into the following methodological questions:

[12] We decided to use the terms 'public' and 'hidden' in this report. The term 'notable' is also often used for public figures, especially in the context of Wikipedia. We chose 'hidden' not for the reason that they are not mentioned at all in newspapers, but that they cannot be found in (cultural) knowledge bases for famous or public people.

- Can we use the Wikidata linkability of historical newspaper ground-truth datasets on entity linking[13] as a proxy for the distinction of public vs hidden people?
- What type of public figures are actually linked in historical newspaper datasets?
- Are there linguistic cues in the context of person entity mentions in newspapers that are reliable enough to be used to detect whether a name is that of a public or hidden person?
- Can we approximate the Wikidata linkability of person mentions by these contextual linguistic features? In other words, can we avoid linking mentions that are unlikely to be in Wikidata?
- How much training material is necessary to achieve a reasonable performance to classify name mentions into public vs hidden? Can these methods be used to effectively collect mentions of hidden people?
- Where do newspapers typically mention hidden people by name (i.e. proper names, as opposed to definite descriptions)? Where do they mention public figures?
- Is there a way to visualise the occurrences of these mentions with respect to page numbers, layout regions, or just within the content of a page?
- Can we identify and cluster typical content items in large newspaper datasets that mention either mostly public, mostly hidden, or a mixture of public and hidden people?

The main goal of the week was to create a proof-of-concept for a subset of the raised problems. We focused our efforts on prototyping a system that would be able to determine, from the context, whether a person mentioned in a newspaper article is a public figure (i.e. someone present in Wikidata), or a hidden person (someone not represented in Wikidata). Additionally, we decided to experiment with innovative visualisation ideas to better understand and explore where public and hidden figures are mentioned in newspapers.[14]

4.1.2 Related work

We approached the task of distinguishing between *hidden* and *public* person entities as a sequence tagging problem, basically an extension of the more general named entity recognition and classification task (NERC), which *detects* potential mentions in running text and *classifies* them into different named entity types (such as person, location, and organisation). Ehrmann *et al.* [3] recently surveyed approaches to NERC on historical data.

The task of assigning a unique identifier from a knowledge base (or NIL if the mentioned entity is not represented in the knowledge base) to a mention of a name in running text is known as (named) entity linking (EL), named entity disambiguation/normalisation or, if the link is to Wikipedia or Wikidata, as wikification. A recent survey introduces the approaches to solve this task on modern textual material [4], and the HIPE-2022 overview reports results of a shared task on entity recognition and linking in historical newspaper data in English, German, French, Finnish and Swedish [1].

The topic of notable people in knowledge bases such as Wikidata is presented in [5]. The authors aggregate, validate and analyse the content of different editions of Wikipedia and Wikidata, ending up with a validated subset of 2.29 million persons, from which they conclude

[13] In other words, whether a person mentioned in a news article is present in Wikidata (represented in entity linking datasets with a QID, i.e. a unique Wikidata identifier) or not (represented with 'NIL').

[14] We are grateful for the suggestions, ideas and discussions on visualisations that Stefan Jänicke brought into our working group.

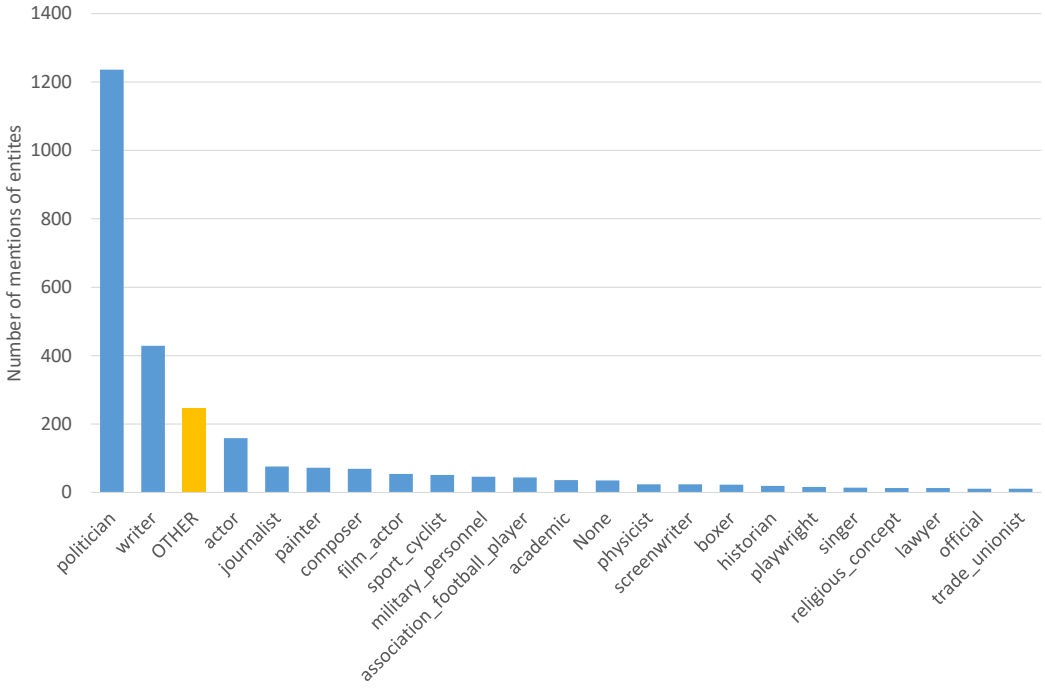


Figure 2 Distribution of Wikidata linked entities' labels (most often their professions in the case of persons) in our French "hipe2020" and "newseye" newspaper dataset. For each person, the chosen profession (among the ones that can be found for the person) is the most common label in the collection (if we take all possible labels for all linked persons into consideration). The OTHER category accumulates entity professions that are mentioned less than 10 times in the dataset.

that it roughly covers an elite of 1/43,000 of humans having ever lived. The distribution of the main occupations of the individuals in this subset is: Culture (31%) including journalists; Sports (28%); Leadership (27%) in politics, religion, nobility etc.; Science/Discovery (12%).

4.1.3 Data

For our experiments, we used two of the French datasets provided as part of the HIPE-2022 shared task on named entity processing in multilingual historical documents[15] [1]: namely the `hipe2020` and `newseye` datasets. They consist of historical newspaper pages or articles in French, ranging from mid-19th century to mid-20th century, annotated for the tasks of named entity recognition and entity linking to Wikidata. In these articles, each entity mention has been manually annotated and classified into one of several categories (including 'person'), and linked to its corresponding Wikidata QID (i.e. the Wikidata identifier) if existing, or tagged as NIL otherwise.

To better understand the professions of people mentioned in historical newspapers that are linked to Wikidata, and compare it to the distribution of occupations in the Wikidata people dataset mentioned above [5], we analysed the linked person entities mentioned in the `hipe2020` and `newseye` datasets used for this case study.[16] Figure 2 shows the distribution

[15] For more information on the shared task and on the datasets, see `https://hipe-eval.github.io/HIPE-2022/` and `https://github.com/hipe-eval/HIPE-2022-data` respectively.
[16] We used the CLOCQ API [6] for efficient retrieval of the persons' Wikidata labels.

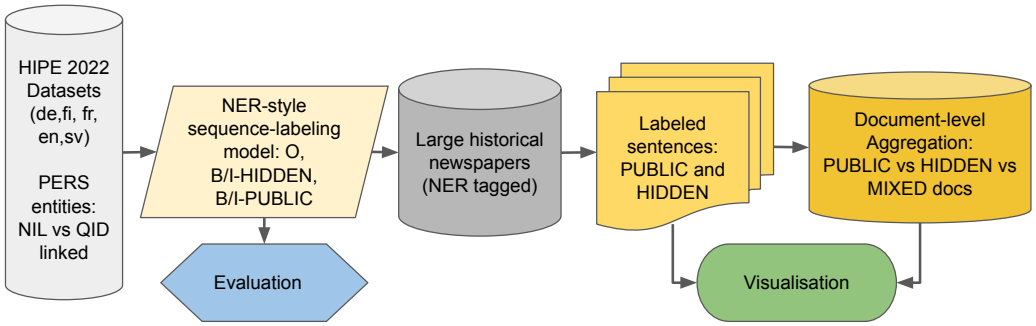


Figure 3 Proposed experimental pipeline of our case study. Using HIPE-2022 datasets for training a sequence tagger for the recognition of hidden/public persons from NER-tagged input. The results of the application of this model can be visually explored, aggregated on the sentence or document level, and further analysed on the respective levels.

of profession labels of linked persons, with *politician* as the dominant one. Many of the long tail of rare professions are accumulated in the category OTHER. Here it is important to know that a person can have several labels (professions) in Wikidata. For our statistics, we chose for each person the profession that was the most frequent one in our data set. This might explain the imbalanced distribution to some degree.

4.1.4 Approach

Figure 3 shows an overview of the proposed approach and the experiments that we tried to implement. Starting with an entity-linked dataset, we create training and evaluation material by labelling all person entities that are linked to Wikidata as PUBLIC and all entities that are not linked as HIDDEN. In this step, we only consider the proper name parts of the annotations and ignore titles, professions or further specifications attached to the names. The proper name tokens are then masked by generic `[PERS]` tokens.[17] Next, a traditional NER-style sequence labelling model is trained on the IOB-formatted representation of the training material. The performance of this model on the test set serves as an indicator of whether there is sufficient signal in the linguistic contexts for a reliable distinction between public and hidden. This model can then be applied to further newspaper data. This data needs to be NE-tagged according to the training material, and proper name parts need to be masked by `[PERS]`. In our case, we just trained an NER tagger with the unmasked training material, but any other tagger could serve the purpose. Finally, given a single newspaper page or any aggregated page subset (e.g. all front pages), we can then visualise where the occurrences of PUBLIC or HIDDEN names are.[18] Another use of the output of the PUBLIC/HIDDEN tagger could be the collection on the sentence level (without layout information playing a role) or on the content item level (e.g., international news vs local news, advertisements, obituaries, radio programs).

[17] We decided to mask person names for two reasons. On the one hand, this was done to force the system to learn from linguistic context, and to avoid the system learning cues that may be implicit or explicit in the person's name. On the other hand, we were aware that public figures appear recurrently in entity linking datasets, and we wanted to avoid our system from basing its predictions on seen dataset-specific training examples.

[18] Alignment between person mention offsets and OCR token image coordinates is required for this.

Table 1 Descriptive statistics of our French datasets compiled from HIPE-2022 datasets `hipe2020` and `newseye`.

Dataset	PUBLIC	%	HIDDEN	%	TOTAL
`hipe2020`	2143	55.1%	1743	44.9%	3886
`newseye`	2695	43.7%	3475	56.3%	6170
ALL	4838	48.1%	5218	51.9%	10056

4.1.5 Experiment 1: Hidden vs Public

For the task of HIDDEN vs PUBLIC classification of person names, we fine-tuned a historical BERT model for French (`bert-base-french-europeana-cased`[19]) and worked with HuggingFace's token classification approach.[20] We fine-tuned the model to the HIPE-2022 data, focusing on the `hipe2020` and the `newseye` datasets. In both datasets, mentions of persons in the data are ideally annotated with a link to Wikidata if the person exists there, and alternatively tagged NIL if the person is absent from Wikidata.[21] Table 1 shows the amount of HIDDEN and PUBLIC entities in our material. Note that overall we have slightly more HIDDEN entities. Interestingly, the two sources have complementary distributions of the two classes: `hipe2020` has more PUBLIC, whereas `newseye` has more HIDDEN. An explanation for this could be that `newseye` pages (full pages were annotated) were randomly sampled, whereas in `hipe2020` the randomly sampled content items (OLR-segmented journal articles) were manually checked before annotation to exclude unsuitable material, such as advertisements. It is also important to note that some of the instances of NIL annotations in the training material are due to the name being ambiguous, or context or time for humans to investigate was insufficient. To balance the proportion between non-person tokens and person tokens, we excluded all sentences in the training material that did not contain any person names. As already mentioned, to facilitate that the model learns from the context, we masked the person names using the `[PERS]` special token for each proper name token.[22] We used 4,283 sentences for training, 535 for validation, and 536 for evaluation.

Findings

We report the results of a single trained model in Table 2. We observe that the performance of this binary classification task in terms of F1 is equal for both classes. However, there is a clear tendency for the model to over-predict PUBLIC compared to HIDDEN. The performance is not perfect overall, but the model clearly finds signal in the context to make the correct prediction in almost 70% of the cases.

[19] `https://huggingface.co/dbmdz/bert-base-french-europeana-cased`

[20] We used the HuggingFace implementation with a learning rate of $2e-5$, a batch size of 16, and weight decay of 0.01, for 10 epochs. We adapted the implementation from the example notebook provided in their Github repository: `https://github.com/huggingface/notebooks/blob/main/examples/token_classification.ipynb`

[21] Note that we take the presence or absence of a person on Wikidata as a proxy of the notability of that person. However, we are aware that this is just a proxy, and we do not consider it a perfect ground truth. A proper evaluation would require annotating the data for the specific objective of distinguishing public and hidden figures.

[22] For example, *'M. Pierre Dupond'* would be masked as *'M. [PERS] [PERS]'*.

Table 2 Performance of the PUBLIC-vs-HIDDEN sequential tagger on our test set split for French.

	Precision	Recall	F1
HIDDEN	0.757	0.625	0.685
PUBLIC	0.634	0.744	0.685
Overall	0.688	0.682	0.685

Examples

Below, we provide some examples of sentences from the test set, in which all proper name parts of person mentions have previously been masked with the special token `[PERS]`. This is the input to our HIDDEN-vs-PUBLIC sequence labelling model and represents real examples extracted from historical newspaper articles.

▶ **Example 1.** *«[PERS], de la Côte-aux-Fées, fusilier dans la compagnie Germann, et [PERS], fusilier, natif de Neuchâtel ; sont prévenus que s'ils estiment pouvoir prétendre une part aux susdits dons, ils doivent adresser le plutôt possible à la Chancellerie soussignée, les certificats qui constatent les blessures qu'ils ont reçues et les droits qu'ils ont de participer à ces dons, afin que ces pièces puissent être envoyées avant le 15 juillet au comité prémentionné siégant à Berne.»*
→ both mentions are identified as HIDDEN.

▶ **Example 2.** *«Le roi, la reine, le prince héritier, la princesse Louise, le prince Waldemar et sa femme se sont rendus d'hui à bord du vapeur Tantallon-Castle, qui est ancré dans le port de Copenhague, pour rendre visite à [PERS].»*
→ identified as PUBLIC.

▶ **Example 3.** *«M. [PERS] [PERS], mineur, habitant Garnich, a été happé par un convoi de wagonnets Rodange.»*
→ identified as HIDDEN.

▶ **Example 4.** *«[PERS], ministre d'Etat belge et chef de la vieille-droite, dont nous avons annoncé le décès à l'âge de 86 ans, était une personnalité de premier plan.»*
→ identified as PUBLIC.

▶ **Example 5.** *«[PERS] [PERS], 37 ans, a été poignardé alors qu'il se rendait à l'opéra.»*
→ identified as HIDDEN.

▶ **Example 6.** *«Monsieur et Madame [PERS] [PERS] et leur fils [PERS]»*
→ all identified as HIDDEN.

▶ **Example 7.** *«Le télégraphe nons apportait lundi malin czar, Alexandre III, a été clamé immédiatement ; c'est le second fils d [PERS] [PERS]*
→ all identified as PUBLIC.

▶ **Example 8.** *«Le Président [PERS] [PERS] a assisté aux funérailles de M. [PERS] [PERS], mineur, habitant Garnich, a été happé par un convoi de wagonnets Rodange.*
→ where the president is identified as PUBLIC, and the miner as HIDDEN.

Table 3 Performance of PERS named-entity tagger trained on the HIPE-2022 French datasets `hipe2020` and `newseye`.

	Precision	Recall	F1
PERS	0.801	0.813	0.811

Limitations

Whereas in Example 8, the system adequately predicts the president as a public figure and the miner as a hidden figure, based on observation, we find this is a case in which the system struggles, as it appears to push the prediction of one entity in the direction of the other entities in the sentence. Therefore, further quantitative investigation regarding the performance of the tagger in mixed PUBLIC/HIDDEN sentences is needed.

4.1.6 Experiment 2: Applying the Recognition of Public vs Hidden Person Mentions on Other Newspaper Data

In order to test on other newspaper material than HIPE-2022 data, we create a pipeline that first recognises person (PERS) mentions in texts and then classifies them as PUBLIC or HIDDEN. The output of this pipeline serves as input for visualisation experiments. We choose the French-language *Indépendance Luxembourgeoise* newspaper for this experiment because its ALTO XML format contains the page position for each token of the texts. This allows us to explore visualisation ideas that combine HIDDEN/PUBLIC information with layout information.

An Application-Specific PERS Tagger

Using the same French HIPE-2022 data as for the HIDDEN/PUBLIC classifier, we trained a HuggingFace model with the same setup to classify PERS entities as HIDDEN or PUBLIC in the Luxembourgish title. The performance of this simple model (F1 score 81%) as shown in Table 3 does not match the state of the art, but we deem it good enough to serve as a NER component in our case study.

Dataset Description

The data used for these experiments is the French-language *Indépendance Luxembourgeoise*[23]. It includes the coordinates of individual words in the `ALTO`[24] XML format, so the distribution of hidden vs public person entities over pages can be analysed. To examine diachronic changes, two distinct years were selected: 1872 and 1928. See Table 4 for detailed data statistics.

[23] `https://eluxemburgensia.lu/periodicals/indeplux`, digitised by the National library of Luxembourg
[24] `https://www.loc.gov/standards/alto/`

Table 4 Data statistics and results on *Indépendance Luxembourgeoise*.

Year	Issues	Pages	Sentences	Entities	PUBLIC	HIDDEN
1872	152	608	130,562	17,243	6,166	11,077
1928	141	564	133,512	19,180	7,925	11,166

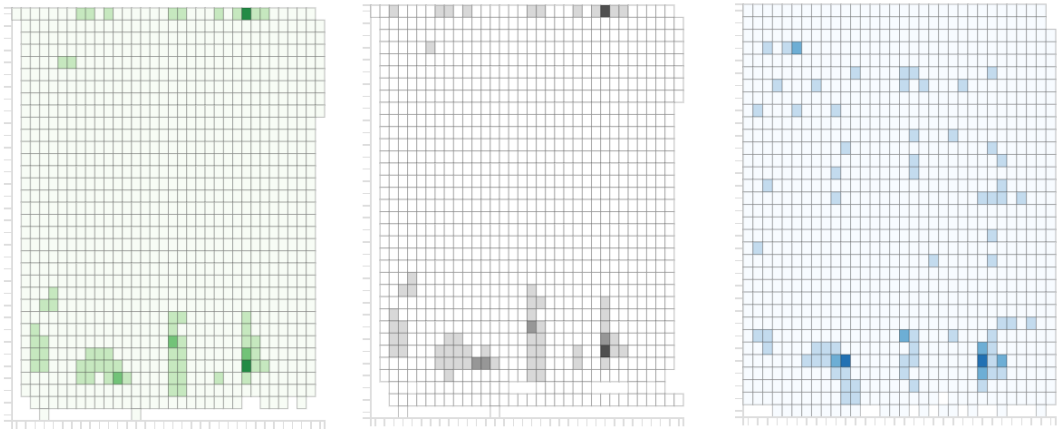


Figure 4 Distribution of entities over all pages from 1872 and 1928. The green heat map shows all person mentions, the grey heat map renders only HIDDEN ones and the blue heat map the PUBLIC ones.

4.1.7 Visualisations

Our Processing and Visualisation Pipeline

In order to make these visualisations, the ALTO files have been run through the code from `dagstuhl-vips`.[25] The ALTO blocks are first transformed into a JSONL file with the script `extract-alto-blocks.py`. Then the blocks are split into sentences using *spaCy*[26] with `splittextblocks.py`. These sentences are run through our sequence tagger for PERS recognition and HIDDEN/PUBLIC classification. Finally, the tagged sentences from our processing pipeline are combined with the ALTO block data to extract positions of all person mention tokens of HIDDEN/PUBLIC entities on the page by the script `tagged-to-wordpos.py`. Multi-token person names can span more than one line in a newspaper column layout. To keep it simple, we represent each person name by the coordinates of its start token that is tagged either as B-HIDDEN or B-PUBLIC and we refer to them as HIDDEN and PUBLIC respectively. These positions are then loaded into an *elasticsearch*[27] index and visualised using *Kibana*.[28] The visualisations below are all produced from Kibana.

[25] `https://github.com/ymaurer/dagstuhl-vips`
[26] `https://spacy.io/`
[27] `https://www.elastic.co/`
[28] `https://www.elastic.co/kibana/`

Figure 5 Projection of word coordinates onto the top-left to bottom-right diagonal. The top-left position of the token is taken and for person names that span multiple ALTO strings, only the first one is considered. The resulting value is between 0 (top-left) and 1 (bottom right). E.g. the words surrounded by the green and yellow boxes are projected to the red diagonal along the coloured (green and yellow) lines.

Page Heat Maps

The page heat maps in Figure 4 show that the persons classified as HIDDEN tend to be at the very top or the bottom of the page. The PUBLIC persons are distributed more evenly.

Projection on the Page Diagonal

A suggestion by Stefan Jänicke[29] was to simplify the visualisation by projecting the 2-dimensional coordinates of a word onto the 1-dimensional page diagonal. This dimensionality reduction facilitates the visualisation and allows for more abstract comparisons between data points.

The natural reading order for French historical newspapers is from the top-left to the bottom right. The most important information is usually on the top-left because that is the first thing the editor wants the reader to see. Therefore, this projection compresses the dimensions but preserves the order of importance of words to some degree. In that way, all pages can be shown on the same graph or an individual graph per page can be computed as shown in Figure 6. This is simplified by the fact that most newspaper issues at the time had exactly 4 pages.

[29] Associate Professor at the Department of Mathematics and Computer Science at the University of Southern Denmark and participant to this seminar `https://imada.sdu.dk/~stjaenicke/`

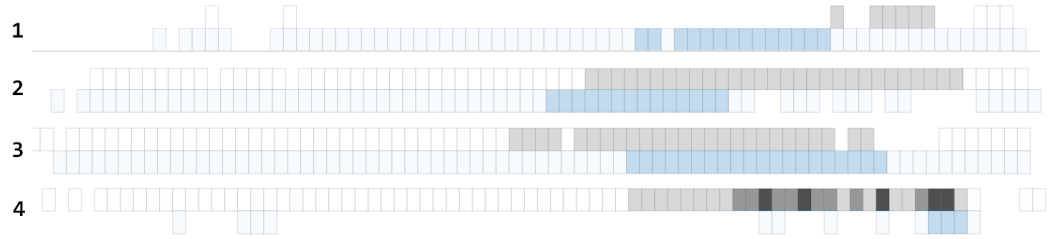


Figure 6 Heat map of HIDDEN (grey) and PUBLIC (blue) entities aggregated per page number.

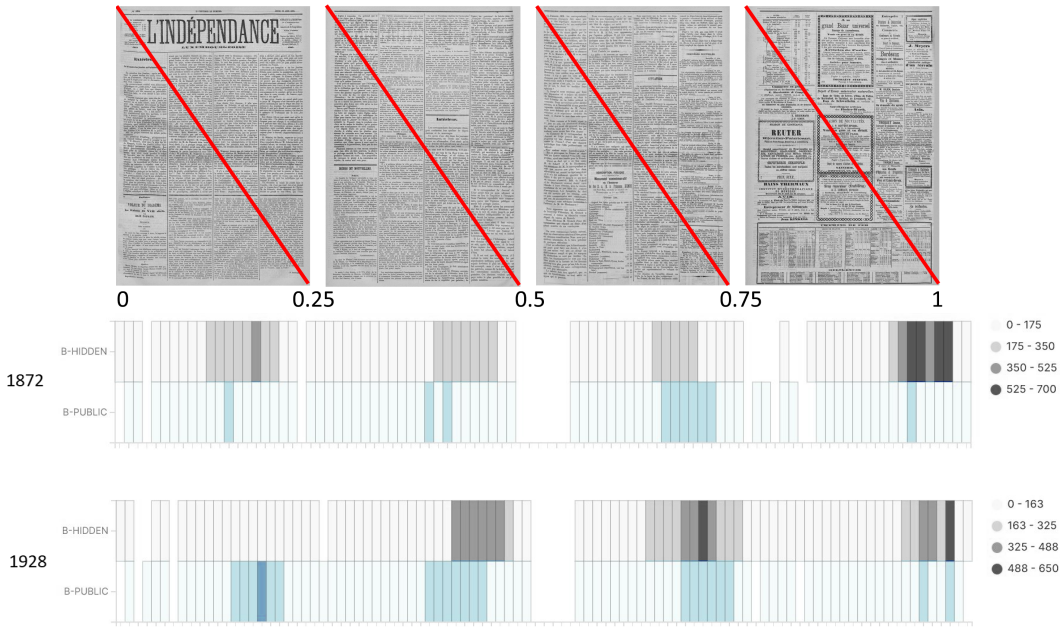


Figure 7 Distribution of entities over pages 1 to 4 from 1872 and 1928. The horizontal axis spans from the top left of page 1 to the bottom right of page 4 as illustrated in the upper part of the figure. The top heat map shows 1872 and the bottom one 1928, the grey part the entities of type HIDDEN and the blue part the entities of type PUBLIC.

Per Page Projection

Figure 6 illustrates that the persons and their PUBLIC/HIDDEN categories are not distributed uniformly on the pages, and page 4 is notably different from the others. There are several possible interpretations to this. First, the last page contains plenty of advertisements for local companies, classifieds and civil registries; these result in a large number of identified HIDDEN entities. Secondly, on the first page, there are clearly more PUBLIC entities, and they tend to be closer to the top left. We assume that the editors lead with stories about public figures because those interest readers more.

Diachronic Comparison

The projection along the diagonal can be generalised to a projection over the diagonals of all pages. As shown in Figure 7, this results in a single heat map where the first quarter represents page 1, the second one page 2, etc. This allows a quick visual comparison between different subsets of the data and is therefore particularly suitable for comparing data distributions

for different years (i.e., one can select any set of entity positions on newspaper pages and compare it to another set). The heat maps of Figure 7 show that the distribution of named entities of type person is different for the two years examined, and the distribution of PUBLIC and HIDDEN are also different.

In particular, HIDDEN entities are clustered on the bottom of page 4 for the year 1872, but in 1928 they are on both page 3 and page 4. There are in general more person entities on page 3 in 1928. This again is probably explained by advertisements, which have become more numerous in 1928 and are also filling up the bottom of page 3.

Another notable difference is that on the first pages of 1928 there are more PUBLIC entities than for 1872. There is no clear explanation why this should be the case.

A further difference of a smaller magnitude is the fact that the frequencies of entity types on page 1 are inverted between 1872 and 1928. While 1872 has more hidden entities and few public ones, it's nearly the opposite for 1928. This could be the result from different editorial choices by the editors, by a different layout or by the use of different language around public and hidden figures 56 years apart.

4.1.8 Conclusion and Open Problems

Our case study on French material indicates that linguistic contexts can help identifying person mentions as either public or hidden. Additionally, we show that innovative visualisation techniques could shed light on the distribution of these two categories of peoples in terms of layout position and page number. As mentioned throughout this report, this is an exploratory case study (conducted in a few days during the seminar), and much more needs to be done to reproduce or improve these findings on further material and languages, such as creating task-specific annotations for the distinction between hidden and public. Another point to investigate are ensemble approaches that can help to deal with the model variance that we observed during training and to improve the performance of the public/hidden classification generally.

References

1 Ehrmann, M., Romanello, M., Najem-Meyer, S., Doucet, A. & Clematide, S. Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. *Experimental IR Meets Multilinguality, Multimodality, And Interaction. Proceedings Of The Thirteenth International Conference Of The CLEF Association (CLEF 2022)*. (2022)

2 Stranisci M.A., Paii, V. & Damiano R. Representing the under-represented: A dataset of post-colonial, and migrant writers. *3rd Conference on Language, Data and Knowledge, LDK 2021, OpenAccess Series in Informatics*, vol 93, pp 1-14 (2021), `https://dx.doi.org/10.4230/OASIcs.LDK.2021.7`

3 Ehrmann, M., Hamdi, A., Pontes, E., Romanello, M. & Doucet, A. Named Entity Recognition and Classification on Historical Documents: A Survey. (arXiv,2021), `https://arxiv.org/abs/2109.11406`

4 Shen, W., Li, Y., Liu, Y., Han, J., Wang, J. & Yuan, X. Entity Linking Meets Deep Learning: Techniques and Solutions. *IEEE Transactions On Knowledge And Data Engineering*. pp. 1-1 (2021)

5 Laouenan, M., Bhargava, P., Eyméoud, J., Gergaud, O., Plique, G. & Wasmer, E. A cross-verified database of notable people, 3500BC-2018AD. *Scientific Data*. **9**, 290 (2022,6,9), `https://doi.org/10.1038/s41597-022-01369-4`

6 Christmann, P., Saha Roy, R. & Weikum, G. Beyond NED: Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases. *Proceedings Of The*

Fifteenth ACM International Conference On Web Search And Data Mining. pp. 172-180 (2022), https://doi.org/10.1145/3488560.3498488

7 Allen, R. Lost and Now Found: The Search for the Hidden and Forgotten. *M/C Journal*. **20** (2017), https://journal.media-culture.org.au/index.php/mcjournal/article/view/1290

4.2 Current Practices of Iterative Classification Approaches for Digitised Historical Newspaper Collections

Mickaël Coustatsy (University of La Rochelle, FR)
Estelle Bunout (University of Luxembourg, LU)
Jana Keck (German Historical Institute Washington, US)
David A. Smith (Northeastern University – Boston, US)

The objective of this working group was to bring together scholars from different disciplinary backgrounds (Digital History, Computational Literary Studies, Natural Language Processing, and Computer Vision) to collect and compare current practices of iterative classification processes for historical newspaper collections. Current work on classification covers several tasks ranging from improving OCR or OLR to enriching semantic annotation of newspaper texts. These methods of classifier refinement, however, do not take advantage of the structure of historical newspaper collections, the punctuated equilibrium of newspaper layout, the evolution of language, the spatially distributed information cascades that spread news and other cultural artefacts. We propose, therefore, a process of structurally-informed exploration as important for building historically useful classification systems.

Digitised newspapers enable a variety of classification tasks. The digital turn in the study of historical newspapers and other sources rests on the creation of digital metadata and digital editions. At the lowest level, this includes digital photography, automatic image classification, and automatic transcription of text via optical character recognition. Recent work applying models from machine learning and artificial intelligence to historical sources, therefore, builds on the output of earlier digital work. Lara Putnam, in her essay on "The transnational and the text-searchable" [7], distinguishes this prior phase of computational research as the digitised turn, which receives less theoretical attention due to its congruence with (current) research habits. As Putnam writes, "*Precisely because web-enabled digital search simply accelerates the kinds of information-gathering that historians were already doing, its integration into our practice has felt smooth rather than revolutionary... How can typing words into a search box – which feels as revolutionary as oatmeal – be a sea change?*". The digital format opens up these collections to the detection of patterns [2] that can be significant for humanist research, such as Long *et al.* [9] have shown with their study of haikus published in the US press. There have been contributions from the (digital) media history that show how text mining can help identify and discuss the emergence and distribution of rubrics in the press, using e.g. classifiers [3]. These few examples illustrate the exciting potentials the combination of digitisation and text mining tools offers; these research outputs, however, remain too often inaccessible for further uses by the community of (digital) humanists, hindering a deeper engagement with the source material beyond keyword search.

Current classification approaches for historical newspapers are numerous. At first glance, they seem very different depending on the disciplinary background: scholars use, for instance, search queries to explore documents and categorise them into relevant or non-relevant items to study events; they also use supervised machine learning to classify documents into newspaper genres to examine how genres have developed over time; or they combine textual and visual embeddings to group different items of a newspaper page (e.g. image and text) that belong together to enhance segmentation. As this shows, digitised newspapers are a very special kind of source given their abundance, heterogeneity, or seriality, and people from different backgrounds approach them from completely different perspectives. These differences have an impact on how scholars construct classification processes. Therefore, classification approaches for historical newspaper collections need to be reflected upon from a transdisciplinary perspective, considering the steps that precede the classification.

What all these multi-level classification approaches have in common is that they aim to improve the usability, searchability, and analytic capabilities for studying news of the past. Documenting these steps are necessary to share them with different disciplines. At the same time, reflecting upon these research projects brings to the surface the flaws of digitised collections and digital archives. However, this information is relevant and can be used as feedback for institutions that are digitising material and making it available online. Especially as in the case of digitised newspapers, two other idiosyncrasies make the findings more difficult to interpret for humanist researchers: the colossal size and the structure of each issue (itself being subject to historical changes). Searching for a word can generate interesting connections between contexts of its uses. For instance, looking for "morphine" in Dutch newspapers in the 19th century leads to hits in rubrics of hard news containing medical topics but also in fiction [8]. And this contextual information is very useful to interpret the results, not only to understand the distribution of the word, but also to reflect on the intertwining of themes across article types at a given time.

Classification is a research output. Current work on classification in historical newspapers covers several tasks. Layout analysis and page segmentation determine article breaks, figures, and reading order. Genre classification helps researchers to cluster articles. Entity linking connects mentions of named entities or definite descriptions (e.g., "the present king of France") to entries in a knowledge base (where such entries exist; see Section 4.1). Although large pretrained language models and image representations have improved many tasks in natural language processing and computer vision, the main paradigm for applying machine learning to classification in historical newspapers is supervised machine learning. Researchers wishing to train a supervised classifier must annotate items – e.g., pages, or lines, or articles, or names, depending on the task – with labels indicating their class. But to perform this labelling, a researcher or research team needs to agree on what classes they are interested in. Determining a useful classification scheme for a research project, we propose, is closer to the information-seeking process researchers employ when formulating search queries and combing through results than it is to the item-wise labelling used in training a simple classifier. Importantly, researchers engaged in search do not usually stop with the results of one query; instead, they often reformulate a query to see if it returns more useful results or to answer some subsidiary question. In addition to these general issues with classification, the variation of historical materials across space and time makes it important to reason not only about our classifier's average error rate but also about other properties of the error distribution. In this report, we explore how existing and speculative search capabilities for digitised historical newspapers can support the development of classifiers and classification schemes.

Keyword search as a classification task. Current research infrastructures for digitised newspapers supports some possible first steps in a classification workflow. Many digital collection search platforms allow users to assign labels to individual items or to save groups of items or search results in collections or work sets. Many platforms offer similarity-based recommendations, linking items (books or articles) to similar ones ("more like this"). Some of these similarity recommendations use only metadata, others use low-dimensional representations (e.g., embeddings or the output of black-box classification algorithms) to compute item similarity. Unfortunately, we are not aware of any bibliographic or newspaper search interfaces that allow user control over these representations for similarity search. To this functionality for supporting classification, we should also add keyword search. When users specify keywords and phrases, they are already engaged in a process of model building. Of course, until they examine the results, they may be, in Nicholas Belkin's formulation [4], in an "anomalous state of knowledge". If you know exactly how some concepts would be described in a historical document, then you would already know most of what you were searching for in that collection; because you need to search to fill gaps in your knowledge, you will inevitably not be able to describe the concept fully. This concept slippage or "vocabulary mismatch" is often easier to see with historical distance, as when, for example, nineteenth-century descriptions of reproductive health or infectious diseases do not match current terminology.

We can therefore think of a user query to a search engine as an incomplete and poorly calibrated model of the concept the user wants to find: incomplete, because it often contains only a few words or phrases representing the concept; and poorly calibrated, because humans are bad at estimating probabilities and assigning quantitative importance to the constituent terms of a query. Once users have obtained the first results of a query, they can however modify the original "model" by adding, deleting or modifying search terms. When performed automatically, this modification of the original query in response to users' judgements of document relevance is termed "relevance feedback".

A call to share the classification models applied to digitised newspapers. When users have labelled some initial examples – either by manual annotation or as the result of a search query – we can see several possibilities for model refinement that will be easily achievable in the short term. First, we mention supervised training, the focus of many current attempts to apply machine-learning methods to historical newspapers. There is scope for feature engineering, prompt engineering, and model selection, depending on the model architecture chosen. Improvements in creating annotated training and test sets have also been the focus of several projects. Secondly, and less widely used, has been query expansion through relevance feedback. Creating the model via manipulation of a human-readable query has some advantages. Finally, similarity search is supported by many newspaper platforms at the level of individual items, but not at the level of sets of items. Especially for those systems that compute similar items by projecting text or images into a low-dimensional embedding space, suggesting items that are slightly farther away than the nearest neighbours might improve the recall of concept expansion.

These methods of classifier refinement, however, do not take advantage of the structure of historical newspaper collections: the punctuated equilibrium of newspaper layout, the evolution of language, the spatially-distributed information cascades that spread news and other cultural artefacts. We propose, therefore, a process of structurally-informed exploration as important for building historically useful classification systems. For instance, when collecting training data for page layout models, we should sample a full range of historical periods for the newspapers of interest. Alternatively, we could have users check the predictions of a trained layout model over a broad temporal range.

Classification as part of digital source criticism. Interpretative work with digitised newspapers is mediated through image processing, text transcription, and search technologies. In addition to the filters of who and what gets recorded and archived, we observe differences in the effectiveness of optical character recognition, image analysis, and document classification. If these archival and digital filters removed or corrupted data uniformly at random, they might not affect our analyses, but they are often correlated with variables of interest, such as document date. To take a simple OCR example, the *Chronicling America* portal to the data from the US Digital Newspaper Program contains all issues of the *Richmond Daily Dispatch* from 1852 until its change of name in 1884. The word 'Virginia' appears at least once on 96% of these pages, but this average conceals an uneven trajectory over time. Starting in 1880, "Virginia" appears on only 84% of pages, compared to 98% before that time. The beginning of 1880 also corresponds to new microfilm rolls and a new batch (`vi_journey_ver01`) in the digitisation workflow. Comparing the *Daily Dispatch* before and after the beginning of 1880 falls prey to this time-dependent distortion. Although a recall of 84% in these noisier issues might still be useful, we are less able to generalise about the prevalence of terms. If, for example, we are interested in the rise of "scientific" racism in the 19th century and search the *Dispatch* for 'Caucasian" (as used in such anti-Reconstruction organs as the *Lexington* [Missouri] *Weekly Caucasian*), we find it on 0.68% of pages before 1880 and 0.33% thereafter. Should we conclude that "Caucasian" was used only half as often after 1880?

We observe similar effects with more complex problems. In the Newspaper Navigator experiments reported by [5], for example, page-element classifiers based on image features achieve an average precision of 74% at detecting headlines in Chronicling America pages from the first quarter of the twentieth century but 52% for 1875-–1900 and 21% for 1850–1875. Similar differences in classification accuracy were observed for advertisements, illustrations, and other categories. These models are still useful for many applications – just as many other retrieval systems can still be useful at 20% average precision – but varying accuracy over time makes it more difficult to answer questions about changing page layout, the advertising basis for newspaper publication, and other topics. These errors analysing images and transcribing text in historical newspapers arise from mismatched training sets and data shifts.

At these error rates, it is difficult to ensure that any individual word or document is correctly classified, just as an inaccurate medical test may lead to improper decisions in particular patients' cases – and just as machine learning can cause harm in other domains when applied to individuals. But in epidemiology, as in many social scientific and historical investigations, we can frame questions about the prevalence of a particular disease (or behaviour, or linguistic feature), even if we remain unsure of any given individual's medical state. In a classic result from epidemiology, Levy *et al.* [6] showed how to derive unbiased estimates of the proportions of a population falling into two classes (e.g., infected and not infected) given noisy tests. We need to know the test sensitivity – i.e., $p(\hat{T} = 1 \mid T = 1)$, the probability that a positive case will be correctly detected – and its specificity – i.e., $p(\hat{T} = 0 \mid T = 0)$, the probability that a negative case will be correctly detected. If the proportion of the population whose tests are measured as positive is $p(\hat{T} = 1))$, then the corrected estimate for the proportion of positive cases is

$$p(T = 1) = \frac{p(\hat{T} = 1) - -[1 - -p(p(\hat{T} = 0 \mid T = 0)]}{p(\hat{T} = 1 \mid T = 1) - -[1 - -p(\hat{T} = 0 \mid T = 0)]}$$

Combining classifications to mitigate their individual limitations and explore digitised newspaper collections. If we train a classifier to estimate the proportions of a document collection falling into various classes, we can use information on the error distribution of

this classifier to correct these estimates. In an example of search in noisy OCR, assume for now that specificity for most queries is nearly 1 – i.e., it is very unlikely that one long word would be corrupted into a different long word. Further assume that sensitivity scales with the length of the query word. If we estimate the character error rate for the *Richmond Dispatch* as 10% before 1880 and 20% thereafter, the corrected percentage of pages with "Caucasian" before 1880 would go from 0.68% to 1.8%, and for pages from 1880 from 0.33% to 5.1%, suggesting this term's frequency continued to increase.

Other possible directions for structurally-aware exploration include:

- Analysing the variation in classifiers and representations trained on multiple datasets as input;
- Speculatively exploring the consequences of alternate annotations on classifier predictions, toggling between document-level, feature-level, and corpus-level predictions; and
- Checking our understanding of what the classifier is learning by generating synthetic data, e.g., using a language model to generate new examples of a genre.

Where to share classifications of digitised newspaper content? Classification activities are dependent on existing infrastructures since they build on previous digital work for building input document representations, features, and annotations. We expect that many collections of digitised newspapers will provide application programming interfaces (APIs) not only for accessing images or OCR transcriptions of individual pages, articles, issues, or other metadata, but also for submitting search queries and retrieving results. To describe how these APIs might support classification workflows, we can distinguish at least five general kinds of queries they accept:

- unweighted keyword search, possibly with boolean and phrase operators;
- weighted keyword search, where individual terms, phrases or predicates may be assigned weights in the relevance function;
- dense text embedding retrieval, which takes a fixed-dimension vector representation of a query and returns documents or passages by their similarity to this vector;
- dense image embedding retrieval, which takes a vector representation of (part of) an image and returns (parts of) images; and
- document similarity retrieval, which accepts a single document as a query and returns other documents.

These query types do not exhaust the space of possible retrieval systems for digitised periodicals, as illustrated by the range of capabilities in *impresso*[30], *NewsEye*[31], and others. They do, however, form a useful set of primitive operations supported by several of these platforms.

Many document classification systems achieve acceptable accuracy using bag-of-words models with linear decision functions (e.g., logistic regression). We could thus retrieve likely members of particular classes using weighted keyword search and perform online updates to our model using relevance feedback. Unweighted keyword search would require more scaffolding on top of methods such as decision lists and random forests, but could be optimised end-to-end.

Several libraries have also experimented with similarity search by encoding text and image data using large pre-trained neural models. After mapping, say, an image into a fixed-dimensional vector space, these systems then perform a nearest-neighbour search in that

[30] `https://impresso-project.ch/app/`
[31] `https://www.newseye.eu/`

space to retrieve similar images. Where these vector-similarity searches could be exposed via an API, they can form useful primitives for classification systems. In some cases, a retrieval system will encode documents using a known published model such as BERT-BASE-UNCASED. We could directly encode queries using this same model. If the document-encoding model is not known, or if we wish to improve on this baseline performance, we could train our classification systems to learn improved query encoders.

In some cases, a retrieval API might only return nearest-neighbours for individual items in the collection rather than allowing arbitrary vectors as queries. In addition to supporting fast approximate relevance feedback [1], this clustering information can form the basis for a classifier.

References

1 Cartright, M.-A., Allan, J., Lavrenko, V., McGregor, A. Fast query expansion using approximations of relevance models. In *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)* (2010)

2 Dzogang, F., Lansdall-Welfare, T., Team, F. N., & Cristianini, N. (2016). Discovering Periodic Patterns in Historical News. PLOS ONE, 11(11), e0165736. https://doi.org/10.1371/journal.pone.0165736

3 Langlais, P.-C. (2022). Classified News. Revisiting the history of newspaper genre with supervised models. In E. Bunout, M. Ehrmann, & F. Clavert (Eds.), Digitised Newspapers – A New Eldorado for Historians?: Tools, Methodology, Epistemology, and the Changing Practices of Writing History in the Context of Historical Newspaper Mass Digitisation. De Gruyter Oldenbourg. https://www.degruyter.com/document/isbn/9783110729214/html?lang=en

4 Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5(1), 133–143.

5 Lee, B.C.G., Mears, J., Jakeway, E., Ferriter, M., Adams, C., Yarasavage, N., Thomas, D., Zwaard, K., Weld, D.S. The Newspaper Navigator Dataset: Extracting And Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America. http://arxiv.org/abs/2005.01583 (2020)

6 Levy, P.S., Kass, E.H. A three-population model for sequential screening for bacteriuria. *American Journal of Epidemiology*, 91(2):148–154 (1970)

7 Putnam, L. The Transnational and the Text-Searchable. *American Historical Review*, 121(2):377–402, (2016)

8 Walma, L. W. B. (2015). Filtering the "News": Uncovering Morphine's Multiple Meanings on Delpher's Dutch Newspapers and the Need to Distinguish More Article Types. TS: Tijdschrift Voor Tijdschriftstudies. http://dspace.library.uu.nl/handle/1874/324205

9 Long, H., & So, R. J. (2015). Literary Pattern Recognition: Modernism between Close Reading and Machine Learning. Critical Inquiry, 42(2), 235–267. https://doi.org/10.1086/684353

4.3 Fairness and Transparency throughout a Digital Humanities Workflow: Challenges and Recommendations

Kaspar Beelen (The Alan Turing Institute – London, GB)
Sally Chambers (Ghent University, BE & KBR, Royal Library of Belgium, Brussels, BE)
Marten Düring (Luxembourg Centre for Contemporary and Digital History, LU)
Laura Hollink (CWI – Amsterdam, NL)
Stefan Jänicke (University of Southern Denmark – Odense, DK)
Axel Jean-Caurant (University of La Rochelle, FR)
Julia Noordegraaf (University of Amsterdam, NL)
Eva Pfanzelter (Universität Innsbruck, AT)

4.3.1 Main challenges and aim

How can we achieve sufficient levels of transparency and fairness for (humanities) research based on historical newspapers? Which concrete measures should be taken by data providers such as libraries, research projects and individual researchers? We approach these questions from the vantage point that digitised newspapers are complex sources with a high degree of heterogeneity caused by a long chain of processing steps, ranging, e.g., from digitisation policies, copyright restrictions to the evolving performance of tools for their enrichment such as OCR or article segmentation. Overall, we emphasise the need for careful documentation of data processing, research practices and the acknowledgement of support from institutions and collaborators.

Increasingly, historical newspaper data undergoes automatic processing using probabilistic methods. For example, topic modelling may inspire the identification of semantic facets within a set of articles, and word embeddings can suggest new keywords and as such different contexts or semantic shifts over time. The acknowledgement of such input matters inasmuch as it holds novel analytical potential and constitutes opportunities to broaden researchers' views on their sources. At the same time, it can mislead researchers due to the underlying principles which govern their creation and make them neither neutral nor objective. We therefore emphasise that researchers benefit from accessible information regarding the processing of data and its fairness. Still, at some point they will nevertheless have to trust systems' output and accept that their findings also depend on factors beyond their understanding, e.g., the impact of different constellations of search engine settings or the outcome provided by topic modelling tools.

Our goal is to compile recommendations for different aspects of transparency and fairness required for the analysis of digitised and enriched historical newspaper collections. We focus on aspects with a potentially high impact on the outcome of research. We distinguish between the need of researchers to obtain information for processes which lie beyond their control, such as institutional digitisation policies and OCR, and their obligation to provide information on aspects they can control, such as the documentation of their modus operandi and sharing research data to allow the traceability of their research. In this report we focus on the former.

The authors of this report have backgrounds in computer science (AI, visualisation, engineering), history (media history, contemporary history, digital history) and library science. This report is the result of one week of exchange and discussion on the topic of data transparency and fairness.

4.3.2 Approach

In a first exploration phase we started with a round-table discussion about fairness and transparency in the context of humanities research based on digitised historical newspapers. For a more formal and systematic review of interface features for historical newspapers see [4, 15].

Second, we performed an initial exploration of seven portals that provide access to historical newspapers. Several issues related to fairness and transparency surfaced in the round-table discussion and in the platform exploration which was centred on the needs of researchers in the historical disciplines.

Third, we used the output of the exploration phase to identify six focus areas which play a key role for historical newspaper research and formulated accompanying recommendations for measures to improve transparency and fairness. The focus areas and measures are organised along the lines of a typical digital humanities workflow.

In a final application phase we used the identified focus areas and recommendations to evaluate the *impresso* interface[32] which was developed with particular attention to transparency. We tested the portal and discussed to what extent each issue plays a role, and to what extent *impresso* implements or enables the recommended strategies. This resulted in insights regarding how far one of the state-of-the-art portals is when it comes to facilitating fair and transparent research on digitised historic newspapers.

4.3.3 Definitions

User Various types of persons work with digitised historical newspaper data, for example humanities scholars, interested lay people, collection owners, and portal developers, as well as scientists from other fields, such as natural language processing (NLP) researchers, who use newspapers as training sets. In this report, our point of reference are foremost the needs of historians, but we nevertheless expect that our recommendations are also relevant for other user groups.

Collection A comprehensive body of materials, in our case digitised historical newspapers, that is curated by a library, museum, or archive.

Corpus or Research Dataset The dataset that a researcher has compiled and on which they will do their analysis. The research dataset may be a subset of one or more collections. The researcher may have used one or more portals to compile the research dataset. The research dataset has often undergone multiple (iterative) processing and enrichment steps.

Portal An access point to one or more collections of digitised historic newspapers, providing functionality such as keyword search, faceted search, browsing or the inspection of raw scans.[33]

Workflow A sequence of actions executed by a researcher to find, collect, transform, enrich, and/or analyse documents.

Fairness We define fairness as the absence of bias. Fairness and equity can relate to a collection, a corpus, or the input and output of a tool. Fairness can be improved by raising awareness of biases as well as unwanted over- and under-representations. Lacking fairness can either be the result of culturally ingrained biases or technical processing on any stage of newspaper digitisation and enrichment.

[32] `https://impresso-project.ch/app/`
[33] In this report the terms *platform*, *interface*, *web application* describe the same thing.

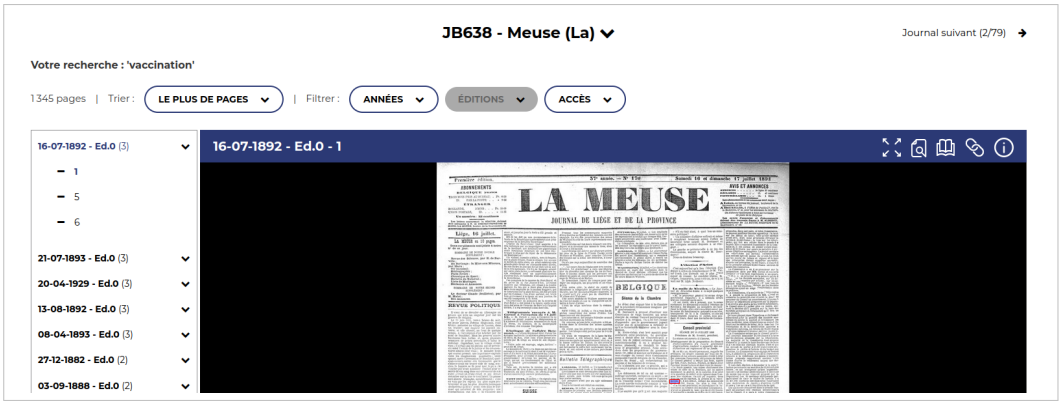


Figure 8 Screenshot of the KBR Belgica Press search interface (© KBR, Royal Library of Belgium).

Transparency Explicit, accessible information regarding the content of a collection or corpus regarding the workflow that was followed to create, process, and enrich it and/or regarding what is known about its fairness.

4.3.4 Exploration: Initial use case-based exploration of platforms

Here we present the findings of our initial exploration of platforms that provide access to historical newspapers. The findings were used as input for the workflow requirements regarding fairness and transparency described in the next section. Seven platforms were investigated. This list is not complete: not covered are, for example, Delpher[34], the CLARIN Newspapers Resource Family[35], and *impresso*.

The initial exploration was guided by a use case on the topic of "vaccination". We have documented this exploration in the form of short reviews which are structured as follows:

- Overview of the portal and its collections including a characterisation of the titles and main features for search, exploration and opportunities to interact with the data.
- Vaccination case study with a focus on the following questions: When was the first article which mentions vaccination published? Which bursts/peaks can be observed in the coverage?
- Summary and assessment of the level of transparency and fairness.

In the following sections we provide reports on the results of these experiments for different portals.

BelgicaPress

The landing page of the Belgica Press portal[36](Figure 2) gives information about the content of the available collection: 121 titles published between 1814 and 1970. Some details are given about the selection of this collection, as well as the information that only one title has been digitised until 1970. However, there is no further information concerning the availability of other titles.

[34] https://www.delpher.nl/
[35] https://www.clarin.eu/news/clarin-resource-families-newspaper-corpora
[36] https://www.belgicapress.be

The interface itself is simple. A search bar can be used to query for keywords and an advanced search allows for date filtering as well as Boolean conditions on the presence or absence of keywords in the results. A first query for "vaccin" yields 12.721 pages in 93 newspapers. The results are grouped by newspaper title which makes the search for the first occurrence and the overall distribution over time within the entire corpus rather laborious. Copyright-protected content is accessible for registered researchers with a MyKBR account. The results are presented as a list of newspaper titles sorted by the number of pages containing mentions of the keywords. When clicking on a result, a new page opens with a viewer allowing the user to see mentions of the keywords. The user can navigate through a list of other pages of this title. It is also possible from this page to switch to another title. There is apparently no relevance ranking for search results but there is the possibility to sort results by newspaper title, by date or by number of pages containing a keyword.

German Newspaper Portal

The German Newspaper Portal[37] has a very simple, "clean" interface that is available in German and English. It is not immediately clear if the search is also bilingual; testing reveals that this is not the case. The caption on the search page has a very minimal indication of the scope of the collection: one can search newspapers from 1671 to 1950. The first thing users see is a search box which invites for a direct keyword search. If users scroll down, three different browsing options are provided. Underneath those is a graph visualising the total amount of newspapers. At the bottom there is a display of a historical newspaper issue of the same date 100 years ago. The interface is clearly designed for a general audience, that is: users focused on encyclopedic use and browsing.

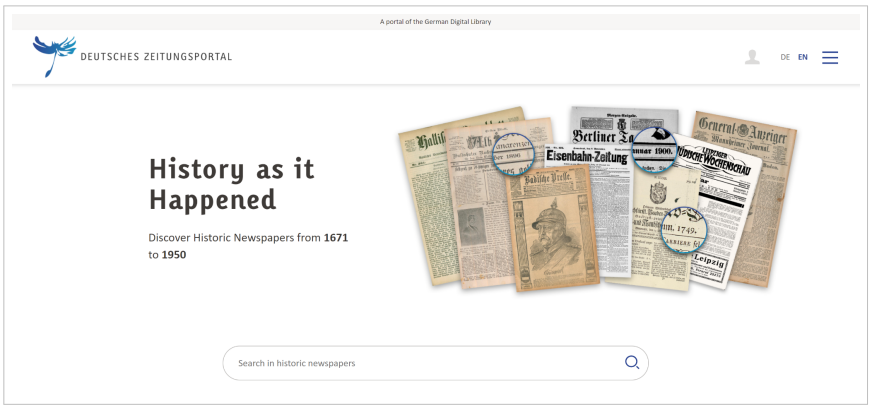
The "About" page indicates that it is a federated site that provides access to newspapers held at different German institutions. It provides data on the total number of newspapers: "The Deutsches Zeitungsportal was launched in October 2021 with 247 newspapers, 591,837 newspaper issues and a total of 4,464,846 newspaper pages from nine libraries. The offerings are being continually expanded and, in the long run, should comprise all digitised historical newspapers which are stored in German cultural and scientific institutions." They also indicate that it is not a representative selection[38], but how "not representative" it is, is not indicated. There is an alphabetical list of all the newspapers with information on their publication history, frequency of publication, and area of distribution.[39] Only 82% of the articles are full-text indexed, but it is unclear which parts of the collection it concerns. This makes it very hard to do source criticism on this collection.

A keyword query for "vaccination" in the search box generates a graph and result-list with snippets organised by titles: apparently 279 results from 26 June, 1802, until 5 June, 1950, were found. Results can be sorted by relevance, but it is unclear how that is defined. Alternatives are sorting functions by publication date (oldest first, newest first) or A-Z or Z-A, where results are apparently ranked by newspaper title.

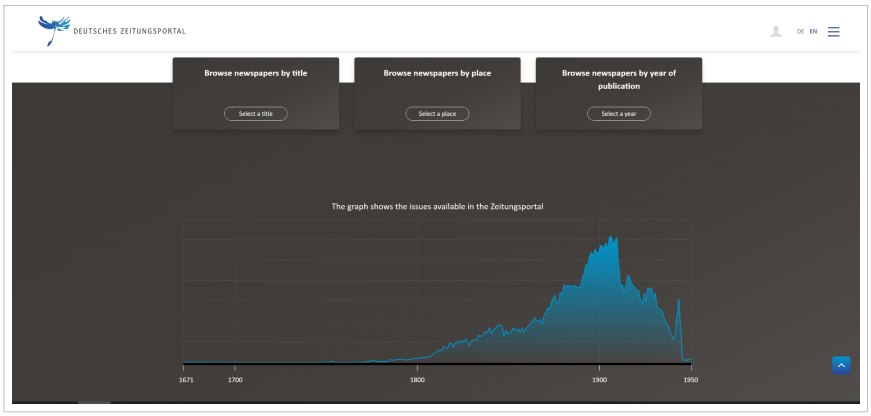
The earliest mention of "vaccination" is in the *Hallesches Tageblatt* of 26 June, 1802, where it is mentioned in a section on "Kuhpocken" ("cow pocks").

However, a wildcard search of "vaccin*" gives 759 results with the oldest in the *Gülich und bergische wöchentliche Nachrichten* of 20 May 1783, but there it mentions the Latin

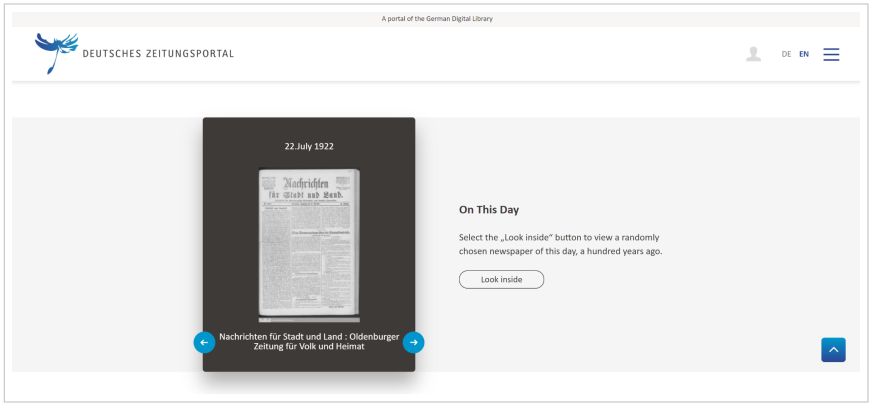
[37] https://www.deutsche-digitale-bibliothek.de/newspaper
[38] https://www.deutsche-digitale-bibliothek.de/content/newspaper/fragen-antworten
[39] https://www.deutsche-digitale-bibliothek.de/newspaper/select/title



(a) Search landing page of the Deutsches Zeitungsportal (© DDB, Deutsche Digitale Bibliothek).



(b) Timeline showing the newspaper issues by year on the Deutsches Zeitungsportal (© DDB, Deutsche Digitale Bibliothek).



(c) Example newspaper page on the Deutsches Zeitungsportal (© DDB, Deutsche Digitale Bibliothek).

Figure 9 User interface of the Deutsches Zeitungsportal.

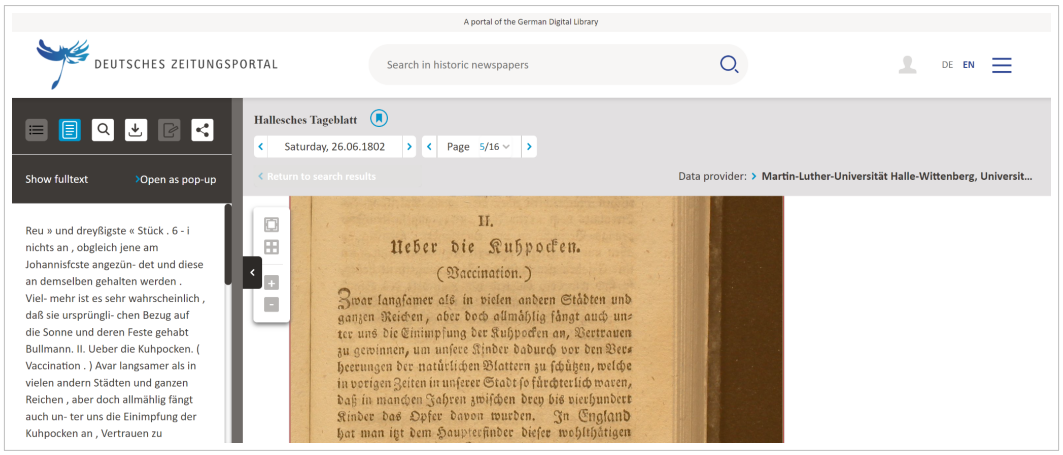


Figure 10 Earliest example of "Vaccination" as shown on the Deutsches Zeitungsportal (© DDB, Deutsche Digitale Bibliothek).[40]

"Vaccinium" which refers to a blueberry[41] – considering that the word vaccination was invented by Jenner in 1796 this result clearly is off topic. The earliest mention from 22 February 1802, is in the *Karlsruhe Zeitung*, the newspaper that most often contains the term (107 articles, 14% of the total). The results page contains a result hit timeline that reveals peaks in 1871-1874 (coinciding with the smallpox pandemic of 1870-1874), 1884 (perhaps a late response to Pasteur's publication on vaccination of 1880?), one around 1890 (perhaps new vaccinations found) and a final one in 1913 (with reference to the use of vaccinations at war time), after which the references decline.

The portal allows users to filter by newspaper title or distribution area (or period), but the functionalities are too limited for putting together a research corpus for our question. The FAQ section points to the well-documented API[42] where the portal allows digitally literate and registered users to extract data.

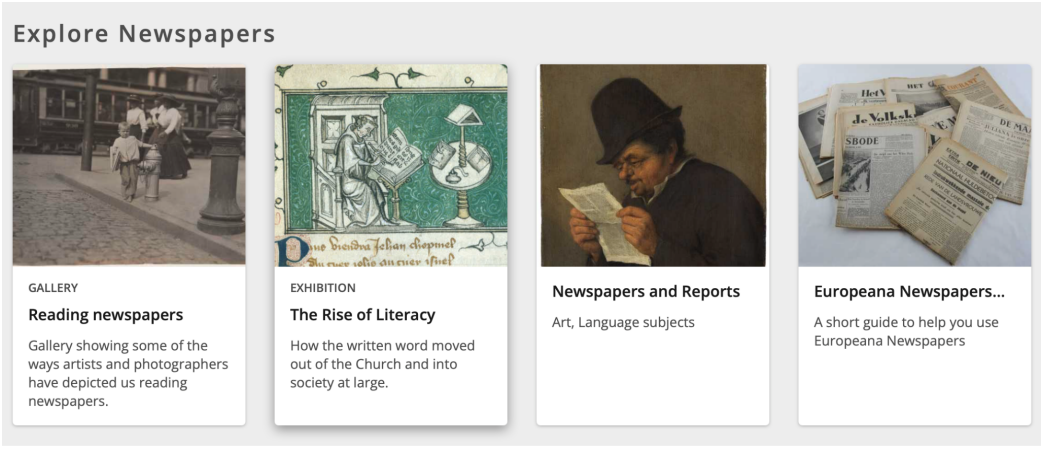
To conclude: the portal allows for exploratory search but is not suited for building a research corpus due to a lack of transparency on the scope and quality of the underlying collections and their processing. The API should be used to extract a corpus and for quality assessments, but this requires technical expertise most historians do not have.

Europeana Newspapers

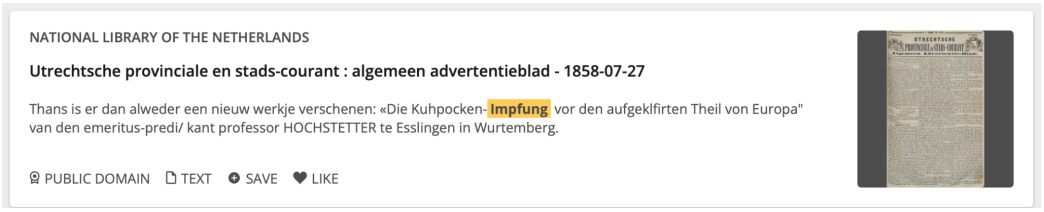
The Europeana portal includes a "Newspapers Theme"[43]. The title of the theme is "Explore the headlines, articles, advertisements, and opinion pieces from European newspapers from 20 countries, dating from 1618 to the 1980s." It includes 887,607 items from ten European countries (Austria, Estonia, Finland, Germany, Italy, Latvia, Luxembourg, The Netherlands, Poland and Serbia). However, it is not possible to see a listing of which titles are included.

[40] https://www.deutsche-digitale-bibliothek.de/newspaper/item/N5UFNO5HCR36P7BJK3TYEI5XM4IR6UUK?issuepage=5
[41] https://www.deutsche-digitale-bibliothek.de/newspaperitem/R2LPDFW7YX27WTEOLNLEBIE4PCDKF66Q?issuepage=4
[42] https://labs.deutsche-digitale-bibliothek.de/app/ddbapi/
[43] https://www.europeana.eu/en/collections/topic/18-newspapers

Additional content is provided at the end of the page, including a gallery on Reading Newspapers, Exhibition on the Rise of Literacy, teaching information on Newspapers, Reports, and a short guide to the use of Europeana Newspapers.



(a) Landing page of the Europeana newspapers portal (© Europeana).



(b) Example of search results for the query "vaccination" as shows on the Europeana newspaper portal (© Europeana).

Figure 11 User interface of the Europeana newspaper archive.

For the "vaccination" research questions, since it is a multilingual portal it is important first to assess which search terms could be used to find relevant newspaper articles on vaccines. Based on this author's language skills a search was undertaken on "vaccine" (EN: 4,626 results), "Impfung" (DE: 12,647 results) and "vaccin" (FR: 4,607 results). Using a wildcard, e.g. "vaccin*" (10,734 results), articles in other languages (e.g., Italian) were included. It was not possible to sort the results. However, a range of result filters are available, e.g., by language or providing country. Additionally there was a "date issued" filter which, however, was not easy to use. Each of the search results provided an overview of the providing institution, the title of the newspaper, as well as a text snippet including the search term, and a thumbnail of the newspaper in question (see below).

Relevant articles can be saved by the user in a personal "gallery" (following the creation of a Europeana account). This gallery can be kept "private" or made "public". It is possible to share a public gallery on social media, e.g. on Twitter[44]. It is possible to download individual search results as page images (jpeg). There does not seem to be an advanced search function. There is some filtering of search results, however, they are not sufficient enough to answer our research questions.

[44] `https://www.europeana.eu/en/set/7143`

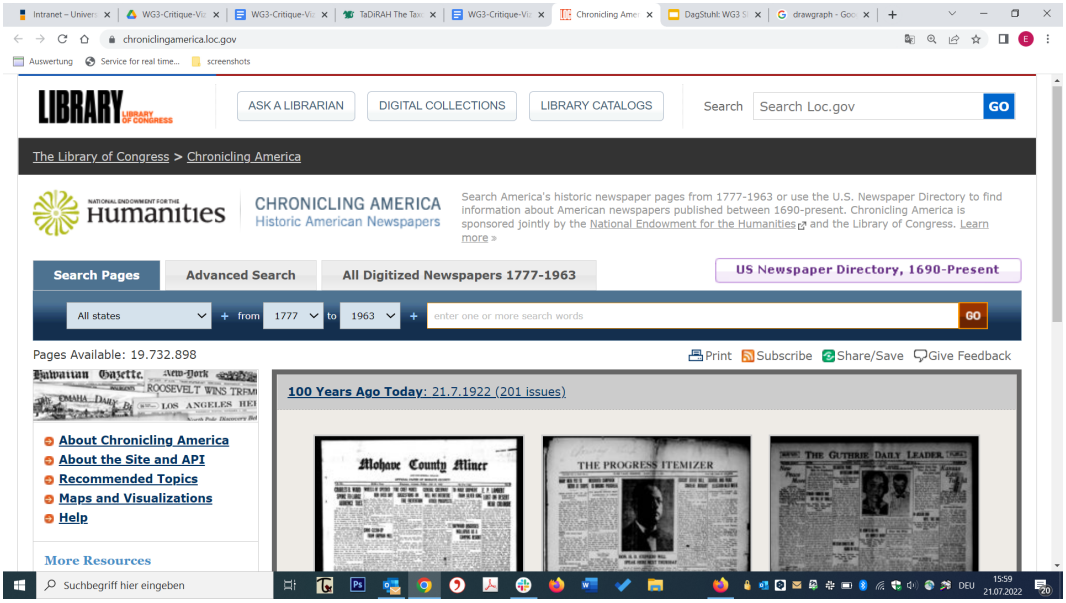


Figure 12 Screenshot of the landing page of Chronicling America (© Library of Congress).

Chronicling America

The landing page of the Library of Congress collection of historical US-newspapers[45] offers several facets and has tabs to give access to the collection and offers links to information pages, APIs, as well as help files, thematic corpora, maps, and visualisations. A search bar and tabs in the background indicate that there are more search options available for advanced search and more complex investigations of the collection. The attention of users is drawn to the centre of the page where a selection of newspaper front pages are displayed under the heading "100 Years Ago Today", today's date and the number of newspaper issues collected.

The collection is a composition of "historic US-newspapers from 1690 to the present". The interface is the result of a collaboration between the Library of Congress and the National Endowment for the Humanities. It includes 3,758 newspapers with 19,7 million digitised pages. The APIs[46] enable expert users to perform the following tasks: search, auto-suggest from newspaper titles, link to stable URLs, linked data views of the collection, JSON view of data, bulk data to use with external services, and CORS- and JSONP-support for JavaScript applications. For all APIs explanations on use and examples are given. Under the heading "Recommended Topics" thematic features in Chronicling America are collected. These corpora are arranged alphabetically, by category, and by date range. They cover a growing number of different themes, time-spans, and genres. The section heading "Maps and Visualizations" leads to a number of graphs and data visualisations of the collection. These pages are updated on a regular basis. So, while the landing page and the simple search bar may give the impression that this collection is meant for a general audience, both the sub-sites and the accessible design of the "Advanced Search" function are clearly intended for expert users.

[45] https://chroniclingamerica.loc.gov/
[46] https://chroniclingamerica.loc.gov/about/api/

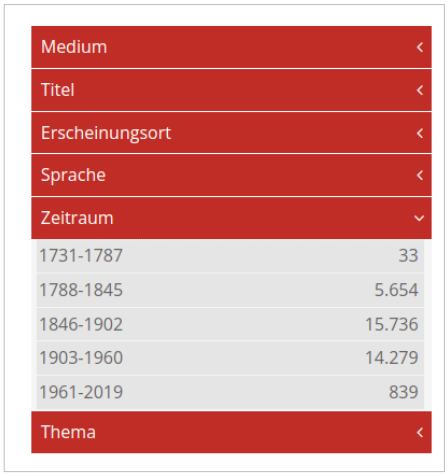
The collection is composed of newspapers in 19 languages: English is the dominant language (with 18,7 mio pages), followed by German (500,000), and Spanish (330,000). At the end of the scale Hebrew (830) and Arabic (2,000) can be found.

With regard to transparency and fairness we wish to highlight dedicated visualisations on the distribution of ethnic press coverage within the corpus. A keyword query for "vaccine" using the basic search bar leads to 208,360 results. The wildcard search for "vaccin*" to capture also results for "vaccination" or "vaccinated" led to slightly over 207,000 results. This apparently wrong output was quickly resolved by reading the help files which indicated that wildcards, as well as upper-/lower-case search, and simple Boolean operators are not implemented. However, the search engine utilises language specific dictionaries which use stemming to include word variants. In order to limit (or increase) the search results, combinations of words or the features offered in the "Advanced Search" should be used (here filters on states, titles, years, front pages, language, combination of words, phrase search, and distance search are implemented). The search for "vaccine fear" produces 69,762 results. A quick scan of the results showed, however, that the two terms often do not occur in the same news item so that this keyword search does not produce usable results. A distance search of the two terms (with a distance of 10 words) produced 1,344 results which did not prove more appropriate (corresponding to sentences similar to "I fear that ..."). Finally, the combination of the terms "vaccine" and "effect" in a distance search of 10 words led to 5,634 results that could be used to study newspaper reporting of this topic. However, it remains uncertain if the word "effect" really covers what a user was looking for in the context of discourses on vaccination. Bulk downloads are not possible at this level. Another point of "granular access" (as opposed to bulk download) is the Chronicling America API. Programmatic access is often preferable for computational analysis, as retrieving and processing data can be easily integrated into one workflow. However, the API functionality is in many ways similar to keyword search. The main functionality is search defined by a query term and refined by a few additional parameters. As can be gathered from the online documentation, the API is especially useful when a researcher wants to retrieve documents related to a specific topic in bulk. Of course, additional filtering can happen downstream in custom-made scripts, but it does not seem to be part of the API functionality (or at least is not very well publicised on the main page).[47]. Having said that, the API is undoubtedly easy to use and the examples are easily adaptable. We successfully used the search endpoint to retrieve articles that mention "vaccination" as a starting point for further processing.

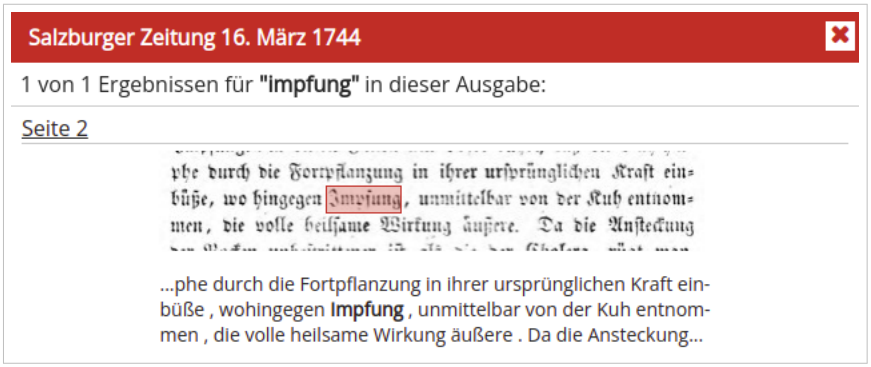
ANNO

AustriaN Newspapers Online (ANNO)[48] is a digitisation project of the Austrian National Library for Austrian historical newspapers and magazines. The project was launched with 15 newspapers in August 2003, and now, more than 25 million pages of more than 1,500 newspapers and magazines can be read and downloaded free of charge and in full text from the portal. The oldest editions date back to 1568. Like other newspaper portals, ANNO offers users to browse and read digital newspapers, and to search for articles based on keyword or an advanced search with additional filters (publication place, date, language, and topic). First hits for "vaccin" are found in the Italian paper *Il Corriere ordinario* from 1679, however referring to cows, a common false positive result we have also observed in other portals.

[47] `https://chroniclingamerica.loc.gov/about/api/`
[48] `https://anno.onb.ac.at/`



(a) Search results by period on the ANNO interface. The strategy on how the temporal facets shown on the left were defined is intransparent (© österreichische Nationalbibliothek).



(b) Screenshot of an individual result (© österreichische Nationalbibliothek).

Figure 13 User interface of the ANNO portal.

ANNO does not provide visual cues that summarise metadata of the retrieved results such as, e.g., *impresso* does. The results are displayed in a faceted browser environment, and facets, for which numerical information are provided, can be selected and deselected. The default ranking of results is by relevance, however, it is not traceable how relevance is defined. Ordering of results by other metadata like date is also possible. Clicking a result opens a popup that juxtaposes scan and OCR transcript, and highlights the search term(s) in both views. ANNO also includes the option to use Boolean operators. Alongside this common search and filter features, ANNO supports filtering by language and themes such as "science" or "agriculture" but it remains unclear, how these filters and the underlying data were generated. A Help and FAQ sections explains available functionalities but do not include information about the technical processing.

NewsEye

The NewsEye portal[49] includes newspaper data from various countries. The data are from different time periods, which makes a comparative analysis difficult. Keyword search in combination with filters can be used to to search through the data. Results are presented as snippets, allowing to quickly assess the relevance of the results, and thus the appropriateness of the keywords. The portal allows users to create and store a custom research dataset by selecting articles or newspaper issues from the search results page. This increases transparency for peers with regard to which data a researcher used for their analysis. The portal does not support transparency with respect to the methods used to select data. A systematic method could be, for example, to fix a set of keywords/facets, and include the resulting articles.

The portal contains an interface to create and store experiments, i.e., sequences of data processing steps (Figure 14a). This functionality increases transparency of the analysis step: not only can the pipeline be stored for future use, it also makes it easy to compare output of different pipelines.

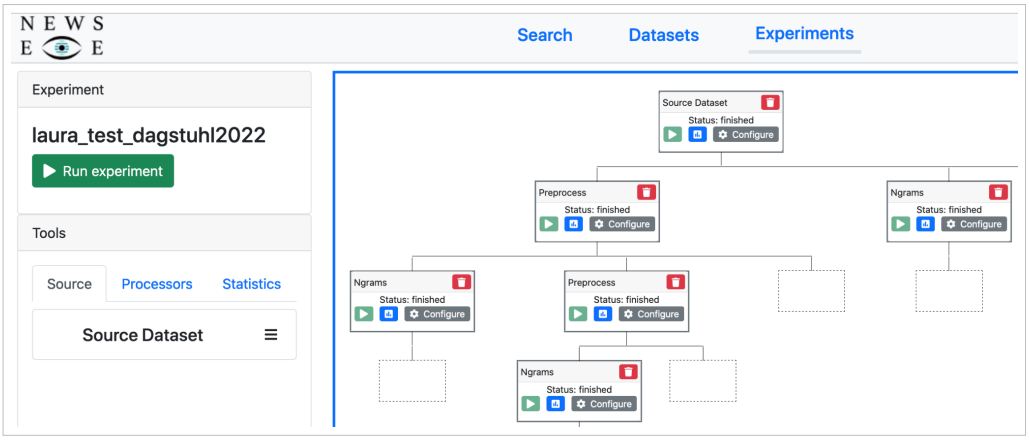
The NewsEye platform used for the current exploration was the experimental platform of the NewsEye project. So, some functionalities were not implemented in this interface yet: the help-button did not work yet; some of the facets still produced unexpected results; only a small number of simple data processing tools were included in the experiment interface (e.g., stopword removal).

The search for the truncated word "vaccin*" produced 25,255 search results in Swedish, French, English, German, and Finnish (Figure 14b). A random check of documents in the different languages confirmed that these were indeed related to vaccinatation. This result is not surprising as in many languages the word vaccination was derived from the Latin "vaccinus" (from the cow). A graph gives an overview of the distribution of the search hits over time. Several facets allow researchers to dig deeper into the search results. Results can additionally be sorted by date or relevance score and the function "random sample" gives a quick overview of what the reader can expect to find in the results. The first mention of "vaccin*" in the newspapers aggregated in NewsEye is in Swedish from the Finnish title *Abo Underrattelser* from 13 March, 1824, where the distribution of vaccines in Finland is discussed.

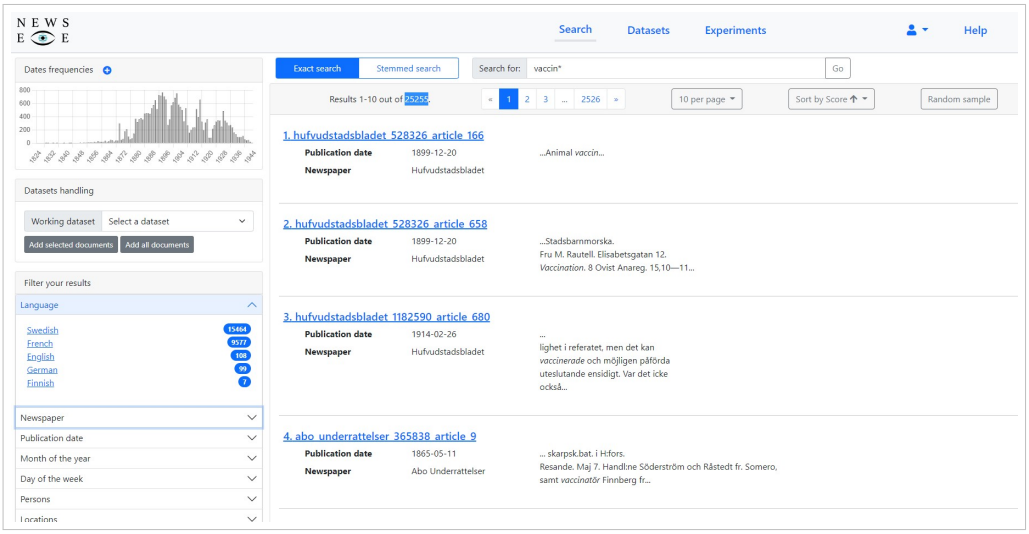
Trove

The Trove portal[50] aggregates a wide range of textual and visual digitised and born-digital resources (books, newspapers, websites, images), hosted by Australian cultural heritage institutions. Trove Newspapers offers a clean interface with common (advanced) search and filtering options alongside the notebook-based GLAM-workbench[51]. An informative About[52] section gives a concise overview of the whole "Trove ecosystem", its construction and guiding principles and is accompanied by a Research Guide[53] which offers basic insights into the availability and legal status of the collection. User expectations are managed effectively through additional documentation, e.g. on a variety of errors and instructions for correction

[49] https://platform2.newseye.eu/
[50] https://trove.nla.gov.au/newspaper/
[51] https://mybinder.org/v2/gh/GLAM-Workbench/trove-newspapers/master?urlpath=lab/tree/index.ipynb
[52] https://trove.nla.gov.au/about
[53] https://www.nla.gov.au/research-guides/australian-newspapers



(a) Experiment workflow on the NewsEye interface. Screenshot of NewsEye interface to interactively create and store experiments (© NewsEye).



(b) Search results for "vaccin*" in the NewsEye platform (© NewsEye).

Figure 14 User interface of the NewsEye portal.

for volunteers.[54] Noteworthy is also optional information concerning cultural sensitivity which users are free to en- or disable throughout their interaction with the portal. During this limited testing we were however not able to see it in action but learned that users are encouraged to ammend DublinCore and MARC metadata of affected articles with the reference "Culturally sensitive". Users are furthermore able to filter content by region, content type, media, and content length with the notable absence of language.

Our case study on vaccination reveals the tremendous added value of crowd sourcing and its effective implementation in Trove. The earliest reference can be found with a query for "vaccin*" and retrieves an article published in the *Sydney Gazette and New South Wales Advertiser* in 1803. The article covers an experimental treatment of orphans with early vaccines against cow pocks including the assertion that "It is believed, that it never has been

[54] `https://trove.nla.gov.au/help/become-voluntrove/text-correction`

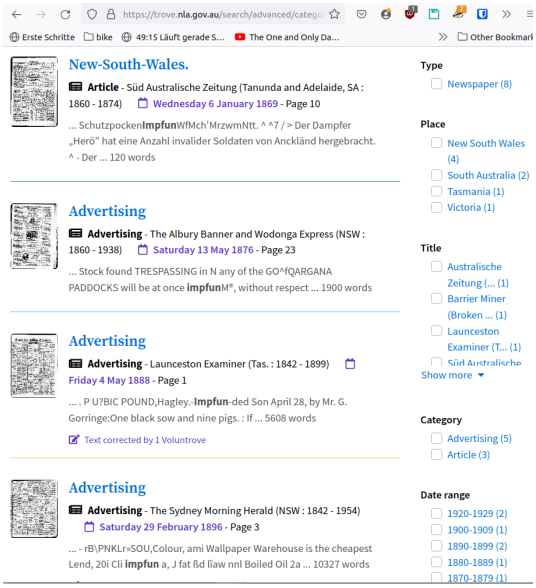


Figure 15 Ranking of search results for "Vaccination" as shown in the Trove interface (© National Library of Australia and Partner Institutions).

fatal, and never will be". The query term "vaccin*" does not occur in the text which has been manually transcribed by volunteers. Instead, the article has been tagged manually with the term "Vaccination" alongside other helpful yet anachronistic tags such as "Bioethics" or "Clinical trial". Trove search includes such tags as well and thereby helped retrieve this article. A distribution over time of search results is possible via hits per year counts but this feature is basic. Information e.g. regarding the breadth of content type detection across the corpus is missing as is information concerning the overall representativity of the corpus for the Australian press. The interface supports crowdsourced OCR correction and informs about the number of "Voluntroves" who worked on a given article. Overall, Trove stands out regarding audience-integration: Crowdsourcing and -annotation features, notebook-infrastructure and cultural sensitivity are well integrated and cater to the needs of different user groups.

4.3.5 Consolidation: Issues and recommendations with respect to fairness and transparency in each stage of the workflow

We identified distinct stages in a typical digital humanities workflow for the analysis of digitised historical newspapers:

1. Research corpus creation and selection,
2. Processing and enrichment,
3. Data analysis,
4. Visualisation of results,
5. Training,
6. Acknowledgements.

Figure 16 illustrates how the stages interrelate. The stages are often performed in an iterative fashion. In this section, we discuss issues with respect to fairness and transparency in each of the stages, and present recommendations for how to deal with them.

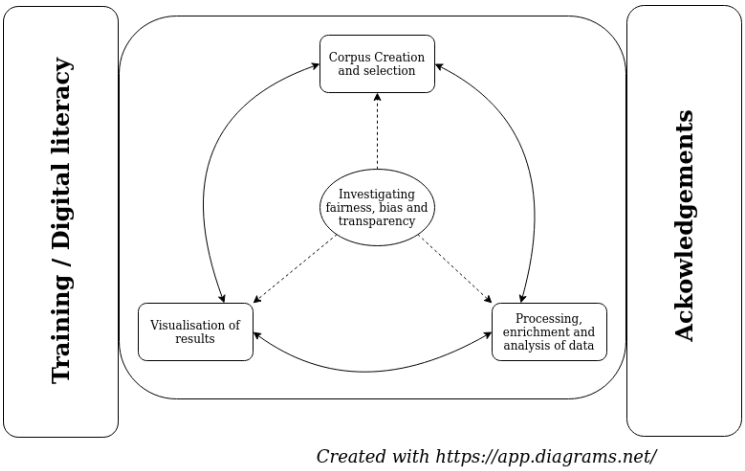


Figure 16 Typical workflow of a researcher. It is essential to investigate fairness, bias and transparency at various stages and in a continuous fashion. It is to be noted that if the training is a prerequisite to the creation of a research dataset, it never really stops (© Axel Jean-Caurant).

4.3.6 Research corpus creation, selection and sharing

▶ Focus areas for transparency and fairness

Awareness of the digitisation and preservation policies. The "digital sample" constitutes a subset of the totality or population of newspapers which existed at some point in time. This population is unknown, but can be approximated using contextual resources such as newspaper press directories or library catalogues [2]. Both are simultaneously useful and problematic. Catalogues record mainly what sources are preserved and are close to providing an overview of the complete newspaper record especially for countries with legal deposit. Newspaper press directories are a useful historical source, but also come with their own issues: classification of what "is" a newspaper changes over time and varies by directory. Ultimately, the population or "newspaper landscape" remains unknown, but we can describe those that have been recorded and/or preserved using metadata derived from catalogue or contextual resources [2].

A rich description of the newspaper landscape helps to contextualise and situate the "sample" of digitised newspapers. In other words, it enables us to at least approximate the "representativeness" of the collection, bearing in mind that the latter concept is more complex than simple proportionality and is always defined in relation to the research questions and ethical values or priorities of the researcher. When it comes to the composition of a corpus, we as researchers and content providers will never be able to get rid of biases and unwanted over- and under-representations. Our goal must rather be to identify, understand, acknowledge them, to infer how they may influence the research outcomes and to make them clearly visible within portals.

When using newspapers at scale and-to repeat the metaphor-as a "mirror" of the past, assessing diversity of collections emerges as a critical issue alongside the processes of media production which heavily influences how the presence was reflected. Researchers need to acknowledge whose perspectives or voices are absent in the data and which social categories dominate a collection.

The question of representativeness is closely intertwined with the issue of diversity and inclusion: which (social) perspectives are present in our data, which are missing? Coming back to the metaphor of newspapers as a "mirror" of the past, we can not simply trust the

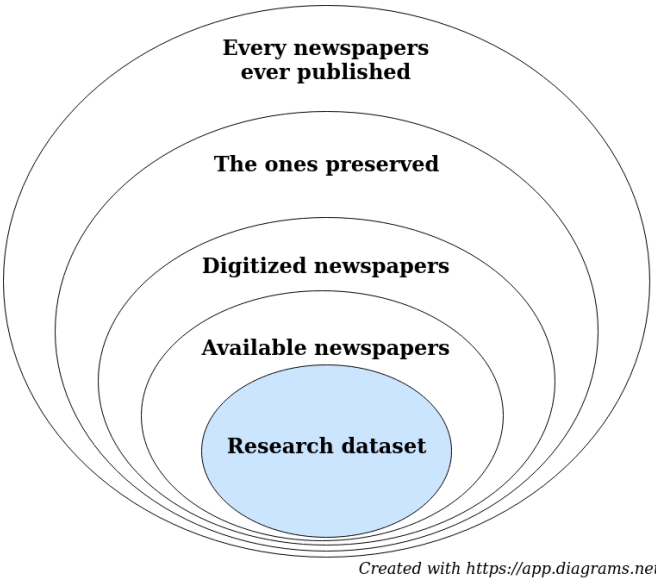


Figure 17 A research dataset is inherently biased, as it is impossible to create a complete dataset because of missing or unavailable sources (© Axel Jean-Caurant)

reflection but need to assess how it potentially distorts our image of the past. With rich descriptions of the sample (and population) we can situate data historically and socially. Of course these will always be rough and approximate descriptions, but nonetheless a crucial part of contextualising (the results derived from) big data. Finally, digitisation does not automatically mean accessibility which depends on the institutional policies (e.g. paywalls) and legal restrictions.

Diverse user needs. Keyword search may satisfy a large group of researchers (and laymen), but others may want to go beyond simply retrieving and reading newspaper content. Interfaces simultaneously provide and restrict access, i.e., the inbuilt functionalities set the limits to how users can navigate and analyse historical materials. They provide the tools and heuristics via which content becomes visible and users can create their research corpus or data set (see below).

However, while such type of access generally works for humanities' scholars like historians, it does not necessarily meet the needs of those who follow more data-driven approaches, such as computational humanities researchers, computational historians, or NLP researchers. The latter often wish to process larger datasets for automatic enrichment and filtering, among other tasks. While most libraries or platforms provide access via search, accessing "newspaper collections as data" (i.e., at scale) is becoming more prevalent, but contemporary portals remain limited in their support for such interactions with the data.

Toxicity and cultural bias. As newspapers are embedded in specific spatial and temporal contexts, their content also contains traces of historical biases, both in text and image. In the most extreme cases, historical newspapers contain "toxic" content, to use a term common in today's research on language models and ethical AI. The textual (or visual depiction) of people, especially the more marginalised and underprivileged, articulate attitudes which are considered offensive within contemporary norms.

But not all biased language is "toxic". A more neutral term would be "overrepresentation" of specific textual patterns among certain subsets of the data, for example conservative newspapers may mention words such as "agriculture" more frequently than newspapers of other political leaning.

▶ Measures to help achieve transparency and fairness

Related to the activity of "research corpus creation, selection and sharing", there are a number of measures that could help or improve transparency and fairness. These measures are both at the level of the cultural heritage institutions providing the digitised newspaper collections as well as at the level of the researchers who create their research corpora.

Collection documentation. Providers of digitised historical newspapers, such as cultural heritage institutions and specific newspaper portals (e.g., ANNO, BelgicaPress, Chronicling America, Delpher, impresso, NewsEye, etc.) can provide detailed information regarding the collection. For example: list of newspaper titles, dates of publication, how much of a newspaper title has been digitised. It could also be useful to provide whatever contextual information about a newspaper and the entire collection is available, e.g., concerning selection criteria, geographical scope, number of editions, print runs, publishers and editors. Information such as political orientation of the newspaper titles, even when imperfect and tied to specific time periods, will be useful here. Ideally, such contextual information is accompanied by sources such as bibliographic references. The question of who is responsible for providing this information was raised, e.g., the cultural heritage institution or the researcher undertaking the research. Perhaps a partnership between these two actors would be most valuable.

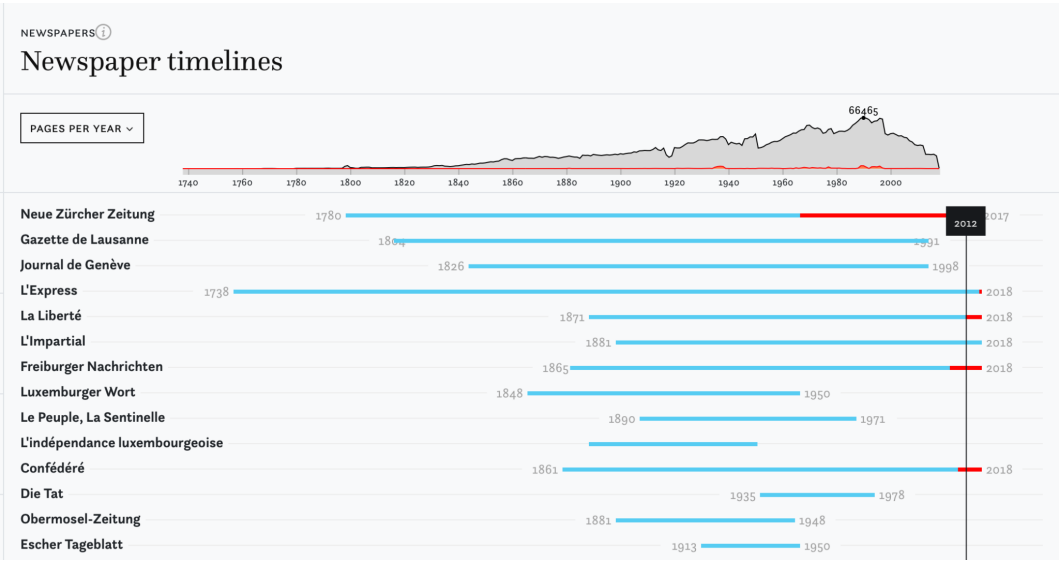
Figures 18 illustrates how contextual information could be displayed: firstly, the Newspaper Timelines[55] from the *impresso* project, and secondly, the Press Picker[56] from the Living with Machines project.

The provision of explicit information regarding digitisation quality (e.g., Optical Character Recognition, OCR) would also be useful, ideally provided at a number of levels: for the whole newspaper title, for an issue, or per article.

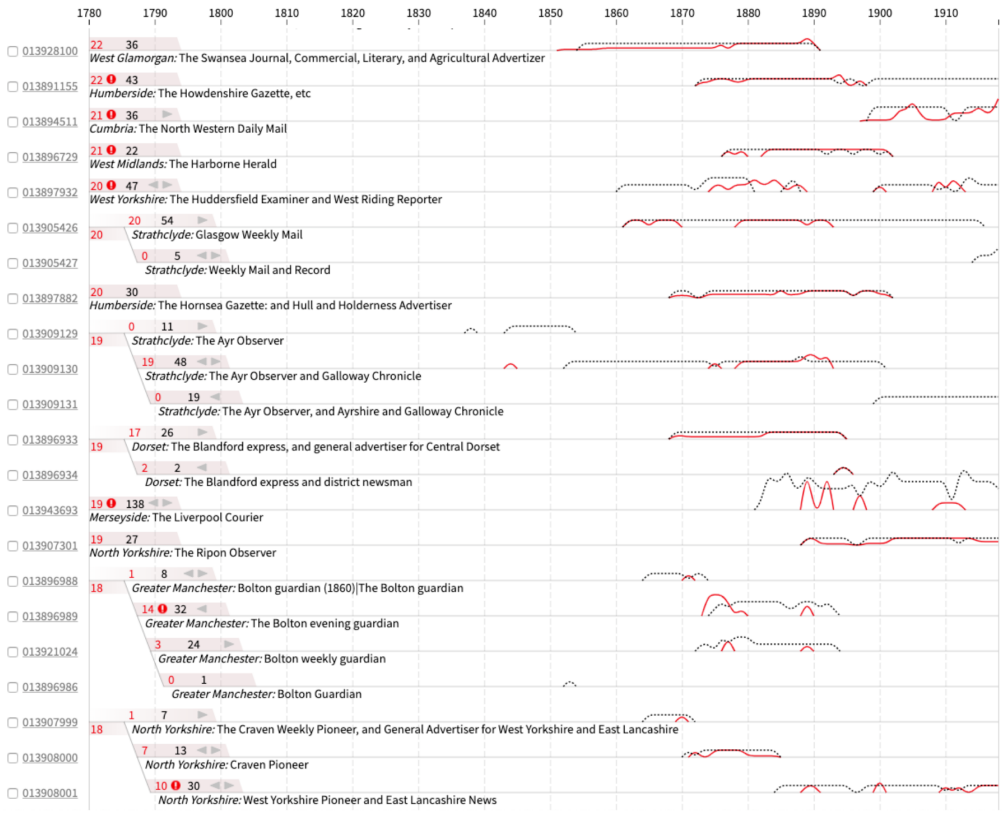
Terms of use. Digitised newspaper providers should provide explicit guidelines regarding terms of use, particularly in terms of legal consideration. For example, the *impresso* platform requires users to sign a Non-Disclosure Agreement (NDA)[57] before access to full collection is granted. Furthermore, it would be useful for cultural heritage providers to provide information about what percentage of the total collection has been digitised. This helps to provide transparency on the "missingness" in a collection, ideally at the level of each of the newspaper titles.

Contested terms. When considering measures to ensure transparency and fairness of research corpus creation, selection and sharing, it is important to consider diversity, equality, equity. To assist both researchers and cultural heritage institutions with this, an equity monitor could be developed. A number of aspects could be considered; for example, the identification of contested terms in a corpus (see [11, 6], as well as [3] and Conconcor[58]). If collection holders are aware that their corpora include contested terms, a disclaimer

[55] https://impresso-project.ch/app/newspapers/
[56] https://livingwithmachines.ac.uk/press-picker-visualising-formats-and-title-name-changes-in-the-british-librarys-newspaper-holdings/
[57] https://impresso-project.ch/assets/documents/impresso_NDA.pdf
[58] https://www.cultural-ai.nl/conconcor



(a) Timeline overview of available newspaper on the *impresso* platform (© impresso).



(b) Timeline overview of British newspapers on the PressPicker tool (© Living With Machines).

Figure 18 Top: Timeline overview of available newspaper on the *impresso* platform (© impresso); Bottom: Timeline overview of British newspapers on the PressPicker tool (© Living With Machines).

alerting users to this could be added to the website. This could be particularly relevant when contentious terms are used for query expansion or are used in a visualisation. For researchers whose corpus includes contested terms, it is advisable to explicitly acknowledge this in their publications.

4.3.7 Processing and enrichment

▶ Focus areas for transparency and fairness

In almost all cases, a raw, digitised newspaper corpus is processed in several ways before it is suitable for analysis. This could include, for example, Optical Character Recognition (OCR), Optical Layout Recognition (OLR), lexical processing such as part-of-speech tagging, named entity recognition (NER), linking to knowledge graphs, topic detection, or sentiment analysis. It could also include manual annotation of, for example, topics, people, or viewpoints. Each processing and enrichment step introduces bias or unwanted over- and under-representations into the data. When data is retrieved via a search engine or recommendation system, the ranking algorithm of that system also plays a role.

Tool performance. Regarding "low" level processing tasks (OCR/OLR), bias is mainly related to the quality of the tools. Do they work equally well on each part of the collection? How does OCR/OLR quality impact the retrievability and accessibility of each (type of) document? The quality of the tools on each part of the collection will depend on the data that they were originally trained on or developed for. They will likely work best on data that resembles the training/development set.

Similarly, for higher level, automated enrichment tasks (part-of-speech tagging, NER, linking to knowledge graphs, topic detection, sentiment analysis), we can ask: how well do they work for each part of the collection? What data were the tools trained on or developed for, and to what extent is this different from the data currently under investigation? NER tools may show a higher performance on some entities than others. Knowledge graphs may not cover all relevant entities.

Ingrained bias. In some cases, bias is ingrained in the collection [3]. As a product of their times, historical newspapers will also reflect the norms, values and language of e.g. colonising nations and their perspectives on their colonies. The use of automated enrichment tools may lead to unwanted side effects with respect to colonial or otherwise outdated terminology. Words may be taken out of context. Consider, for example, that a topic detection algorithm may define a geographically-focused topic as a list of terms including racist references to people.

Posterior annotation. Manual annotation is prone to bias that relates to the viewpoints, background and knowledge of the annotator. In some cases, these highly personal characteristics will be unknown, such as when making use of crowdsourcing. In some cases, we might not want to expose anonymous crowd workers to bias that is ingrained in the collection, such as when offensive, colonial terminology is present.

Ranking. Search engines typically rank documents based on a combination of the following factors: a matching score between a query and the content of the document, usage data in the form of previous queries and clicks, and an importance score of the document, e.g., using a PageRank-like algorithm. Whether a document will appear high in the ranking will therefore be influenced by its popularity (if usage data is taken into account), connectedness (if PageRank is used), and document properties such as length, which impact the matching score [20]. This may introduce distorted perceptions of search results. Therefore ranking-algorithms should be made transparent to users who rely on them.

▶ **Measures to help achieve transparency and fairness**

Fine-grained OCR performance metrics. Detailed information about OCR and segmentation quality helps a user to decide not only whether the quality is good enough, but also whether bias towards certain parts of the research corpus is to be expected. This requires fine-grained performance metrics, for example at the level or articles, newspaper titles or time periods.

Access to "raw" data. Another solution to mitigate or at least understand bias due to OCR errors is to provide access to the original "raw" scans, i.e., the images of pages.

Systematic documentation of tools and training sets. For automated enrichment tools, documentation of how, for what purpose, and on which training set they were created, helps a user to assess whether bias is to be expected when these tools are applied to their data. Several documentation approaches have been proposed in recent years, the most notable being Datasheets for Datasets [6] for documentation of training sets, and Model Cards [10] for documentation of trained models. Also, the research on provenance is relevant here, i.e. a formal representation of the consecutive processes involved in the creation of the enrichments. PROV[59] is an approach to formally capture provenance information on the semantic web.

Scanning for contentious terms. The content of historic newspaper collections will often be "biased" in the sense that the articles display the perspectives of the time in which they were created. Removing this type of bias will mostly not be feasible or desirable in the context of historical research. We recommend to include an "equity monitor" as part of a research design, where a user critically assesses whether contentious terminology is present in the corpus, and whether this is problematic. As noted above, contentious terminology could be problematic when used as input to automated enrichment tools, or when presented to crowd workers. In these cases, a user could decide to not include certain articles in their research. Note that detection of contentious terminology is not trivial and automation of this task is still in its infancy [3].

Disclaimer about contentious language. A user may include a disclaimer as part of the dataset, to warn (other) users and/or annotators that there may be offensive content. This is especially recommended when sharing the corpus for reproducibility or future research.

Representative annotators. We consider human annotators to be always biased. A diverse or representative group of annotators helps to avoid annotations that are skewed towards one background or viewpoint.

Transparency about annotators. Explicit information about who created the annotations helps users to assess whether an (unwanted) bias in the annotations is to be expected. This could consist of age, gender, country of citizenship, (native) language, level of expertise, and way of recruitment of the annotator(s). Note that this information is often not available when using crowdsourcing.

Multiple relevance rankings. Bias introduced by the ranking algorithm of a search engine may be explicated and mitigated by providing multiple rankings. Many search engines already include additional rankings next to relevance ranking, such as a chronological order. However, specifically the inclusion of multiple relevance rankings would allow a user to understand to what extent the ranking algorithm impacts their goals.

[59] `https://www.w3.org/2001/sw/wiki/PROV`

4.3.8 Data analysis

Once a research corpus has been selected and processed it is ready for analysis. Analysis already is an integral part of data selection, processing and enrichment. At first, we discussed this stage as part of the processing and enrichment stage. There are, however, tasks that clearly come after data processing and enrichment; for instance, the identification of named entities in the corpus has to be undertaken with NER software. Therefore, we have decided to identify it as a separate step in the research workflow.

Depending on the research question and the skill set of researchers, the analysis may be performed on the entire collection (e.g., downloading all the data and analysing it with a Jupyter notebook) or of a subset of the collection generated via the search and filter options in a portal. It also may be performed qualitatively, with a scholar browsing and reading specific articles, or quantitatively, applying computational approaches and tools.

▶ Focus areas for transparency and fairness

Traceability. In order to make the research traceable, researchers have to be explicit about the methods and tools they use including, for the latter, the version, the used settings, and why they were appropriate for the task in question.

Often, the analysis of a newspaper corpus involves a set of tools organised in a pipeline. In order to obtain transparency, users should have insight into the composition and performance of the various components of the pipeline. In the case of machine learning tools, it should be clear which versions are used and on which dataset they have been trained.

In order to be fully transparent, ideally all these things are documented, and the tools or queries stored alongside the data. This raises the question what level of documentation is required to make the research traceable or repeatable for others. Some researchers provide tools to document the settings, such as the Gephi Fieldnotes plugin developed by [22]. Others have proposed strategies for tracing all the data handling steps [7]. We see, however, the risk that the effort to produce such documentation may take a disproportionate amount of time and effort. Tool standardisation may make this need less urgent (e.g., the role of SPSS[60] in Social Science research) and therefore reduce this burden.

▶ Measures to help achieve transparency and fairness

Access to facsimiles. To facilitate qualitative research, where users explore the corpus at object level, an interface should present research results in the form of scans next to the OCR and metadata (which most portals currently afford). For quantitative analysis, tools will be used both inside and outside the portals. In order to improve the traceability of the research, researchers should have the ability to store tools and their settings alongside the datasets, perhaps on publicly accessible platforms such as GitHub and Zenodo, or using tools specifically designed for this purpose, such as the Gephi Fieldnotes plugin. "How to cite" text blocks with detailed and multimodal information on the tools and their settings could be helpful here.

Comparative perspectives. The transparency of analysis pipelines can be supported by interfaces that allow researchers to compare the performances of different algorithms on a specific task. An example is the NEWSGAC platform[61] that allows users to compare different

[60] https://www.ibm.com/de-de/analytics/spss-statistics-software
[61] https://github.com/newsgac/platform

algorithms for automatic genre detection in newspapers. In order to increase the transparency of machine learning tools, references to publications on models used and documentation on training datasets is provided (e.g., in the form of "datasheets for datasets"[6]).

Replication. Ideally, users should have the possibility to save, export and reuse their own analysis pipeline and the results. This output should ideally connect to the changing publication and presentation modes for the research results, that allow researchers to include data, code and narratives alongside each other (e.g., the Journal of Digital History[62] that publishes the data alongside the narrative and a description of the methodological issues, or the ESWC conference[63] where linking to data, code and other resources is a review requirement); this is further discussed in the Acknowledgements section below.

4.3.9 Visualisation of data, results and bias thereof

Visualisations have the added value of showing complex matter e.g. in graphs and images that help users to get a better overview of collections, corpora, data sets and also the content of these. Visualisations are important because they support the exchange between data and users since they can help to contextualise the collection on the one hand and research on the other. Using graphs, timelines, charts, maps, word-clouds, bubbles, and similar transparency concerning the collection and the research method is offered. The possibilities of how to support transparency by visualising collections and data span a wide range: e.g., research questions and methods are made explicit, topics can be contextualised, a classification of genres and faults or missing/biased data in the collection can be made visible, OCR-/layout-quality and research approaches can be identified, and comparisons of topics (and many other similar things) are possible. As a consequence interfaces can be designed to offer possibilities for visualisation.

▶ Focus areas for transparency and fairness

General Guidelines for Visual Design. Visual interfaces are in many contexts suitable, necessary means to make patterns inherent in the data set in question salient to the observer. However, visualisations are abstract representations of (typically) numerical data, and the visual mapping of the underlying information always imposes a level of distortion because numbers are rather easier to compare when they are served in textual form than when our brain has to approximate them when they appear in the form of visuals such as bars in a bar chart or dots in a scatter plot. The complexity of comprehending data in textual form increases with the size of the data set, but visualisations help to arrange the data in a way that users get a quick, understandable overview even for vast data sets. In order to limit the level of distortion, visual representations of data have to be carefully designed.

Accurate representations of data . Following Edward Tufte's guidelines for graphical excellence, first of all, visualisations should "show the data", make it coherent and avoid distorting what the data has to say [19]. An appropriate indicator for good visual design is when viewers are induced to think about the substance rather than about methodology, graphic design or the technology of graphic production. Moreover, visualisations should not "lie," i.e., the size of effect shown in the visual display needs to correspond to the size of effect in the data.

[62] https://journalofdigitalhistory.org/en/about
[63] https://2022.eswc-conferences.org/call-for-papers-research-track/

Choice of colour. Of particular importance for visual design is the selection of appropriate colour maps. Qualitative colour maps (a set of different hues) should be used to display categorical data, and continuous colour maps (sequential or diverging) to communicate quantitative data (colour gradients). Although powerful, a general advice is not to encode the most important feature with colour ("get it right in black and white"). Visualisations should also be colorblind-safe, i.e., one should not mix diverse shades of green and red. Several online tools support defining accurate colour maps, e.g. ColorBrewer[64].

Clarity. Next to choosing inappropriate colour maps, visualisation designers should avoid visual clutter that reduces the readability of the displayed data and conceals occurring patterns. Clutter can also occur when choosing 3D over 2D representations, which are the means of choice when visualising data that does not inhere 3D structures. Textures that cause visual stress (moiré vibrations) should furthermore be avoided.

Visual Exploration. In order to support Information seeking, visualisation tools should implement Shneiderman's mantra "Overview first, zoom and filter, then details on demand" [17]. Whereas the overview corresponds in digital humanities terminology to distant reading, details on demand refers to close reading. Thus, visual interfaces should support gradual zooming and filtering of the data to be analysed.

Visualising uncertainty. Especially, data in the context of humanities applications often embody uncertainty of different kinds (imprecision, inhomogeneity, incompleteness). Visualisations are suitable means to communicate these uncertainties, for example through transparency or grey glyphs, indispensable for increased reliability of visual display of information.

▶ Measures to help achieve transparency and fairness

Participatory design. To ensure a transparent visual interface with a minimised level of data distortion, we suggest conducting a participatory visual design process that involves visualisation experts on the one hand, and domain experts that ensure the suitability of the visual design for its intended purpose on the other. An exhaustive overview of visual design principles can be found in [12].

Collection visualisation. There are some good examples on how interfaces can offer transparency on the collections. The CLARIAH-NL Media Suite[65] , and the Sound and Vision Archive[66], offer explanations and graphs to contextualise the audiovisual collection of the Netherlands Institute for Sound and Vision (NISV), explaining both the role of the archive in the archival field in the Netherlands as well as the time the collection spans, what the digital archives cover, what kind of media is included, when updates happen, what part or the collection is digitised, how the collection is enriched, where additional information and help can be found and how it can be searched (as exemplified in Figure 19). It also indicates what the differences between the metadata of two media management systems are and offers links as well as downloads to the description of the metadata fields.

Digital newspaper collections also contextualise their datasets to a certain extent although additional information like the one offered for the Sound and Vision archive might be added. The *impresso* interface indicates the provenance of its collection in the detail view of

[64] https://colorbrewer2.org/
[65] https://mediasuitedata.clariah.nl/dataset/nisv-catalogue
[66] https://mediasuitedata.clariah.nl/dataset/audiovisual-collection-daan

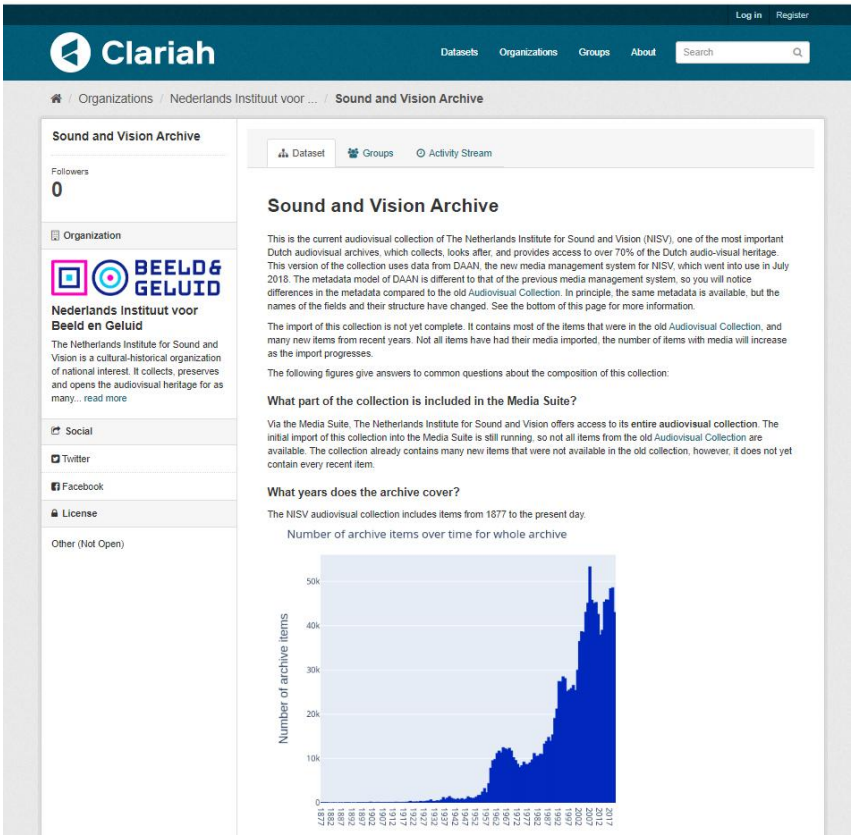


Figure 19 Landing Page of the Sound and Vision Archive which is one of the datasets of the Nederlands Instituut voor Beeld en Geluid within CLARIAH (© Nederlands Instituut voor Beeld en Geluid).

individual titles. A simple graph visualises the overall temporal distribution of the newspaper collection and a bar chart enables users to get a quick overview of the time span and amount of data each newspaper contributes to the collection (see Figure 17). Due to the efforts of the National Digital Newspaper Program[67] which "is a long-term effort to develop an Internet-based, searchable database of U.S. newspapers with descriptive information and select digitization of historic pages" Chronicling America is currently also adding information on its dataset. Figure 20 shows that contextualisation regarding the collection is made in the "Maps and Visualisation" section of the portal and it can be seen that most of this information is of very recent date (mostly updated February and June 2022) and that the "Map and Timeline" of the collection are updated on a weekly basis using the ArcGIS Instant App (see example in Figure 21). The interactive map visualisation has the added value that it supports a scalable reading of the collection, which is often required by humanities researchers.

These examples show that awareness for the necessity of transparency and biases is growing constantly. In this context visualisations can be helpful to support the communication of complex issues at a glance. The visualisation in Figure 16 frames the issue at hand very

[67] `https://www.loc.gov/ndnp/`

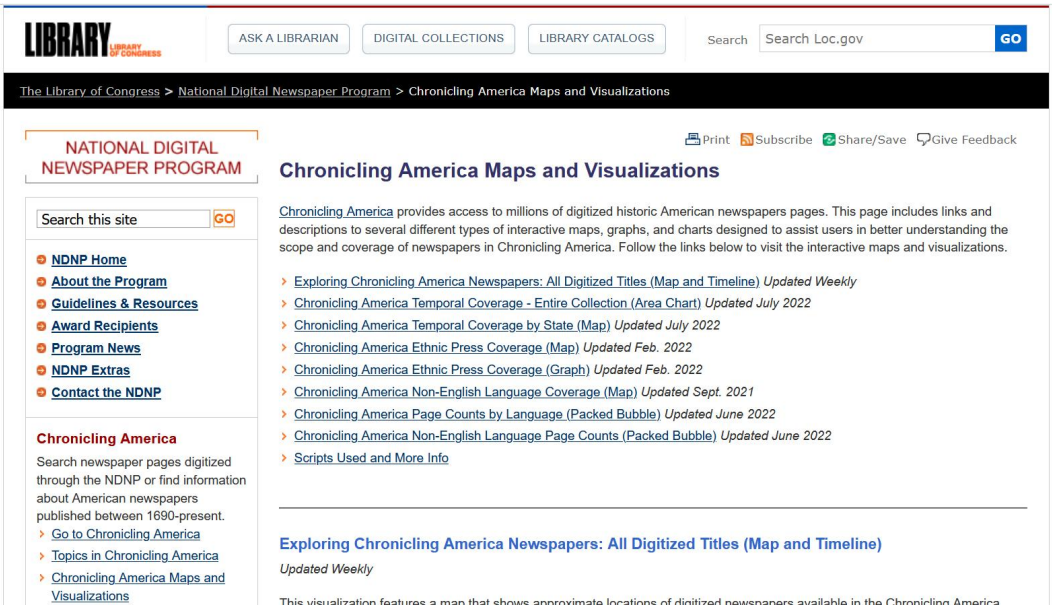


Figure 20 Screenshot of "Chronicling America Maps and Visualizations" where information about the collection can be found (© Library of Congress).

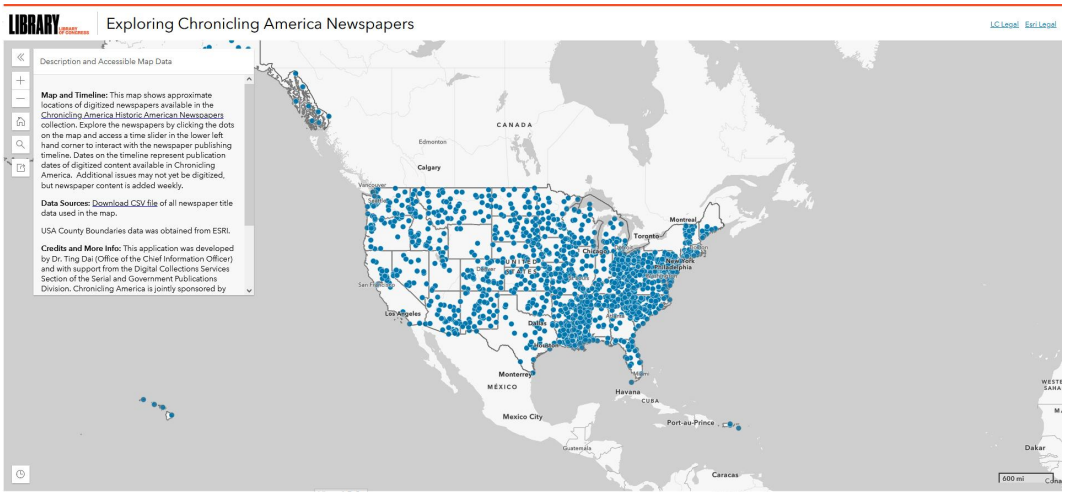


Figure 21 Screenshot of the visualisation of Chronicling America Newspapers, an interactive map created using the ArcGIS Instant App which enables an in depth exploration of the newspapers available in the collection ranked by the primary place of publication (© Library of Congress).

clearly. It is a good example of how well visualisations are able to communicate difficult and complex information. This is also true for analyses that are being done by researchers (see Section Data analysis).

4.3.10 Acknowledgement

▶ Focus areas for transparency and fairness

Reveal hidden labour in digitisation. There is a considerable amount of hidden labour in the process of generating digital newspaper collections. It would be ideal if this hidden labour could be made visible and all the actors in the process could be acknowledged for their work. This would help increase the level of transparency and fairness of the whole digitised historical newspaper ecosystem.

In order for historical newspaper collections to exist, they have to have been acquired by libraries, archives or museums, often in physical form. They then need to be digitised, processed (e.g. metadata generation, creation of lower resolution images for public display), and then enriched (e.g., OCR and article segmentation). It is not feasible to undertake the whole digitisation workflow at once, and therefore selection or prioritisation is needed. All these steps involve the expertise of cultural heritage professionals, computer and data scientists and software engineers, as well as (digital) humanities researchers who would like to use the digitised newspapers as historical sources for their research.

Reveal hidden labour in data curation and enrichment. In addition, the creation of research data sets or corpora, which requires sustained intellectual effort, is often not recognised or acknowledged as formal research output. This both discourages researchers to spend time and properly document this crucial step in the research workflow, but also devalues this work as a necessary evil before the "real" research work can begin. Both cultural heritage professionals and computer scientists contribute significantly to this phase, e.g., by providing historical context information about the historical newspaper collections or by working on information extraction methods to computationally facilitate the corpus building process. It is therefore important that this valuable work is uncovered and made visible.

Finally, such work is often made possible by the financial contribution of (public) funding agencies, the acknowledgement of which is also important.

▶ Measures to help achieve transparency and fairness

Acknowledgement in portals. We recommend the formal and visible acknowledgement of all contributing partners at each step of the digitisation process. This includes the contributions by cultural heritage professionals, software engineers, researchers, and (public) funding agencies. For example, when building a platform for the exploration and analysis of digitised historical newspapers, both the funding agency can be acknowledged, such as is the case with the impresso project ("impresso. Media Monitoring of the Past. Supported by the Swiss National Science Foundation under grant CR- SII5 173719, 2019." [68]) or the NewsEye project ("This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770299." [69]). Additionally, when building a research dataset or corpus, it is important to acknowledge all contributors, such as "Biltereyst, Daniël, Philippe Meers, Dries Moreels, Julia Noordegraaf and Christophe Verbruggen. Cinema Belgica: Database for Belgian Film History." [70]. Not only does this publicly acknowledge the work of people involved in the development of the platform or

[68] `https://impresso-project.ch`
[69] `https://www.newseye.eu/about/`
[70] `https://www.cinemabelgica.be`, all consulted on July 27, 2022

dataset in question, but it also enables it to be cited in articles or other research outputs that have made use of it. Being able to demonstrate the impact of such platforms becomes increasingly important for their sustainability.

Acknowledgement in publications. Acknowledgement in publications is another key method to make all parties who contributed to the development of data and the design of platforms visible. For instance, the domains in which each actor has contributed to such a dataset or platform could be stated explicitly, as is already required with regard to authorship by some journals. The Journal of Open Humanities Data Author Guidelines[71] for example provides a number of recommendations based on the ICMJE (Internal Committee of Medical Journal Editors)[72], outlining criteria for authorship. An area where there is still room for further development is the (academic) recognition of more innovative digital research outputs. Innovative Journals such as the Journal of Open Humanities Data (JOHD)[73], which focuses on the publication of "peer reviewed publications describing humanities data or techniques with high potential for reuse"[74]; the Journal of Digital History which intends "serve as a forum for critical debate and discussion in the field of digital history by offering an innovative publication platform and promoting a new form of data-driven scholarship and of transmedia storytelling in the historical sciences"[75] and the Journal of Data Mining & Digital Humanities (JDMDH)[76] which is situated at "the intersection of computing and the disciplines of the humanities, with tools provided by computing such as data visualisation, information retrieval, statistics, text mining by publishing scholarly work beyond the traditional humanities."[77]

4.3.11 Training

Training is essential to raise awareness of the complexity of enriched historical sources and variability in the quality of available data. In addition, and as we have outlined above, historical newspaper data may reproduce biases present in past societies. Training can provide the necessary understanding and tools to help deal with this.

Digital literacy. Training already exists on various levels. More and more, digital literacy is a part of the curriculum of (digital) humanities students. Some newspaper portals also offer domain-specific training. For example, *impresso* offers 1) General training on research using digitised historical newspapers, 2) Training on how to use the *impresso* processing tools and 3) Platform specific training about the functionality of the interface, linked to the FAQ, which contains references to literature.[78] *NewsEye* provides access to various training materials for schools and universities as well as material targeted at a general audience.[79] The *Ranke 2* platform, for example, offers a series of lessons on Digital Source criticism that can be helpful not only for students.[80]

[71] https://openhumanitiesdata.metajnl.com/about/submissions/
[72] https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html
[73] https://openhumanitiesdata.metajnl.com
[74] https://openhumanitiesdata.metajnl.com/about/
[75] https://www.degruyter.com/journal/key/jdh/html
[76] https://jdmdh.episciences.org
[77] https://jdmdh.episciences.org/page/editorial-policies
[78] https://impresso-project.ch/theapp/usage/
[79] https://www.newseye.eu/
[80] https://ranke2.uni.lu/

Table 5 Stages of digital humanities workflows for historical newspapers including focus areas and measures to implement.

	Focus area	Measures
Research corpus	Awareness of dig. & pres. policies	Collection documentation
	Diverse user needs	Multiple data access points
	Toxicity and cultural bias	Terms and conditions
		Contested terms
Processing & Enrichment	Tool performance	Performance metrics
		Access to "raw" data
		Doc. of tools and training sets
	Ingrained bias	Scanning for contentious terms
		Disclaimers
	Posterior annotation	Representative annotators
		Transparency about annotators
	Ranking	Offer multiple relevance rankings
Data analysis	Traceability	Access to facsimiles
		Replication
		Comparative perspectives
Visualisation	Visual design guidelines	
	Accurate data representations	
	Colour choices	Collection visualisation
	Clarity	Participatory design
	Visual exploration	
	Uncertainty	
Acknowledgement	Reveal hidden labour in dig.,	Ackn. in portals
	data curation and enrichment	and publications
Training	Digital literacy	Publications with best practices
		Code examples
		Example workflows and use cases
		Platform-specific training
		API training

We identified the following types of training as contributing to the skills and knowledge of researchers with respect to transparency and fairness when studying digitised historical newspapers:

1. Publications with best practices;
2. Code examples, for example in the form of Jupyter notebooks;
3. Example workflows and example use cases;
4. Example lesson plans and course material;
5. Platform specific training, in the form of interface walk-throughs;
6. API documentation.

Table 5 offers a high-level overview of the proposed focus areas and measures to achieve transparency and fairness.

In the following last section we undertake a review of the *impresso* portal and assess to which extent it fulfils the above-mentioned criteria for transparency and fairness.

4.3.12 Application: Analysis of *impresso* portal

We revisited the vaccination case study and its underlying questions to apply it to the *impresso* interface for historical newspapers which was developed in close cooperation between historians, computer scientists and designers and with special emphasis on transparency. We revisited the focus areas and accompanying measures we identified above and concentrated on opportunities for improvement of the current interface.

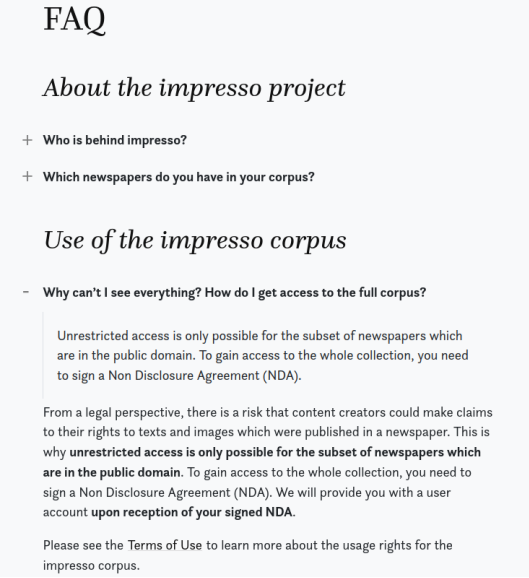


Figure 22 FAQ page on the impresso website (© impresso).

Corpus creation and selection. The Newspapers overview page offers a good overview of the print runs of the newspapers in the collection. Information about missing pages and issues (mismatches between available data and expected data based on library metadata) is available but could be explained more clearly. Information on the origins of the collections from different partners is available albeit scattered across the interface and could be further improved by links to the respective digitisation policies of the partnering institutions. *impresso* offers rich metadata for individual titles (e.g. name, print run, number of pages, orientation, regional focus) and links to their respective pages on the websites of the institutions from which they originate. The value of this information could be increased by allowing users to use them as filters in the search component and by offering a download of the data for individual processing and analysis outside the interface. More detailed information about the alignment between *impresso*'s collection and the collection of the partnering institutions should be added. The hitherto unsolved problem to relate the "tip of the iceberg" of the available digital content to the rest of it, i.e., the total record of newspapers in circulation in the past persists in the *impresso* interface as well. Search results are sorted by "relevance" by default, the underlying settings are not explained. Access is granted via a browser interface and following approval after signing an Non-Disclosure-Agreement. Users are able to export metadata for articles and, depending on legal agreements, also full text. The corresponding FAQ entry should be expanded to explain the content of the export file.

Processing, enrichment, and analysis of data. The FAQ section of the interface collects important information about the project, corpus and interface. Semantic enrichments such as topic modelling, named entity recognition or text reuse are well explained as is the entry

on OCR quality. Filtering by OCR quality should be enabled. The FAQ entries for data export and legal restrictions are valuable but could more concisely explain the legal status of exported data and link directly to project publications on system architecture and the processing pipeline.

Visualisation of results. The application makes use of data visualisations in multiple forms which are accompanied by corresponding FAQ entries. This includes frequently distributions over time e.g. for search results, graphs to represent overlaps between topics. The Inspect&Compare[81] component uses small multiples of bar charts to reveal overlaps and dissimilarities between two queries or article collections and can e.g. be used to evaluate search strategies [1]. The interface is clearly designed for research purposes. Basic knowledge about data and interactive visualisations is a prerequisite.

Training and digital literacy. From a user perspective, the portal supports both scholars with low digital literacy (as tools for analysis are built in) and for more advanced skilled researchers (because data and metadata can be exported as csv). The project has compiled and integrated educational materials for researchers ranging from beginner to advanced level which cover digitised newspapers per se as well as the functionalities of the interface.[82] Tutorials could be placed more prominently in the interface and specific support for visually impaired users is missing.

Acknowledgements. As stated above, the project indicates the source of funding by the Swiss National Science Foundation together with an overview of the full project consortium[83] and the contributions of individuals.

4.3.13 Conclusions

It was our goal to compose a set of recommendations for content providers such as libraries and archives as well as developers of research interfaces, in order to help individual researchers in the field to gain as much transparency and fairness as is required for the analysis of digitised and enriched historical newspaper collections. We did so by focusing on aspects with a potentially high impact on the outcome of research. We found that efforts to increase transparency and fairness have to be made on all stages of the workflow. The stages we identified were 1) corpus creation and selection; 2) processing, enrichment, and analysis of data; 3) visualisation of results; 4) training and digital literacy; and 5) acknowledgements. These stages interrelate and are dependent on each other. It was therefore not always possible to make clear distinctions. We discussed issues of transparency and fairness in each of the stages and also reflected on some measures that can help mitigate biases, lacks of transparency and fairness.

Overall it can be summarised that we found a lot of variability in the current landscape of digital newspapers. This applies within and across newspaper collections and includes differences in metadata standards such as METS/ALTO, the quality of OCR with older processing tending towards lower quality, but also in regard to the scope of enrichment: whereas some content providers invested in the correct identification of even small news items (e.g., obituaries), others only offer PDFs of scanned images. These variations have a high impact on research but are still poorly communicated in contemporary interfaces.

For researchers, therefore, some challenges remain and some of these challenges can be countered by content and interface providers. For example, it remains crucial to understand the "digital sample", i.e., researchers have to be given the ability to assess the (non-)

[81] `https://impresso-project.ch/app/inspect`
[82] `https://impresso-project.ch/theapp/usage/`
[83] `https://impresso-project.ch/consortium/people/`

representativity of the small number of digitised and available newspapers against the background of all past and potentially not archived newspapers. Also, providing detailed information – such as metadata, information on the digitisation process, distribution over time, political orientation of newspapers, links to historical contextual information, etc. – of the nature of the collections that can be found via the interfaces, can be of great help to researchers. Interfaces should therefore provide intuitive guidelines for and explanation of the collection. Also, information regarding diversity, equality, and equity (such as a notification of contested terminology) should be found.

Since the quality of OCR and layout recognition as well as classification issues remain a challenge, it also remains crucial for researchers to be able to always have access to the "raw" data, i.e., the images. If automated tools or training sets are available, a proper documentation has to be provided. The experimental nature of the analysis tools still poses challenges regarding replication and sustainability (e.g., the query storage facility, notebooks for re-training on the same corpora). For example, the CLARIAH Media Suite allows users to rerun queries but updates to the underlying collections cause this feature to break. Again, the progress in automation creates a need for extensive documentation (e.g., the Gephi Fieldnotes Plugin) which could be diminished once there is some standardisation of the tools (e.g., as in the case of SPSS that is more integrated in a methodological framework and also generally accepted as reliable). Another option could be to allow researchers to compare the results of different algorithms. For many research questions, however, it also remains important to work with external tools and adaptable methods. The possibility to download data sets and analyse on (or publish about) them outside the interfaces is one more important feature we identified. Overall, a higher degree of transparency in visual interfaces can be a real asset for humanities research.

References

1. Düring, M., Kalyakin, R., Bunout, E. & Guido, D. Impresso Inspect and Compare. Visual Comparison of Semantically Enriched Historical Newspaper Articles. *Information*. **12**, 348 (2021,9), `https://www.mdpi.com/2078-2489/12/9/348`, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute
2. Beelen, K., Lawrence, J., Wilson, D. & Beavan, D. Bias and representativeness in digitized newspaper collections: Introducing the environmental scan. *Digital Scholarship In The Humanities*. (2022)
3. Brate, R., Nesterov, A., Vogelmann, V., Van Ossenbruggen, J., Hollink, L. & Van Erp, M. Capturing Contentiousness: Constructing the Contentious Terms in Context Corpus. *Proceedings Of The 11th On Knowledge Capture Conference*. pp. 17-24 (2021)
4. Ehrmann, M., Bunout, E. & Düring, M. Historical Newspaper User Interfaces: A Review. *Proceedings Of The 85th International Federation Of Library Associations And Institutions (IFLA) General Conference And Assembly*. pp. 24 (2019), `https://infoscience.epfl.ch/record/270246?ln=en`
5. Fry, B. Visualizing data. (" O'Reilly Media, Inc.",2008)
6. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J., Wallach, H., Iii, H. & Crawford, K. Datasheets for datasets. *Communications Of The ACM*. **64**, 86-92 (2021)
7. Hoekstra, R. & Koolen, M. Data scopes for digital history research. *Historical Methods: A Journal Of Quantitative And Interdisciplinary History*. **52**, 79-94 (2019)
8. Linhares Pontes, E., Cabrera-Diego, L., Moreno, J., Boros, E., Hamdi, A., Doucet, A., Sidere, N. & Coustaty, M. MELHISSA: a multilingual entity linking architecture for historical press articles. *International Journal On Digital Libraries*. **23**, 133-160 (2022)
9. McGillivray, B., Poibeau, T. & Ruiz Fabo, P. Digital humanities and natural language processing "Je t'aime ... Moi non plus". (Alliance of Digital Humanities,2020)

10 Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. & Gebru, T. Model cards for model reporting. *Proceedings Of The Conference On Fairness, Accountability, And Transparency*. pp. 220-229 (2019)
11 Modest, W. & Lelijveld, R. Words Matter, Works in Progress I. National Museum of World Cultures. (2018)
12 Munzner, T. Visualization analysis and design. (CRC press,2014)
13 Neudecker, C., Baierer, K., Gerber, M., Clausner, C., Antonacopoulos, A. & Pletschacher, S. A survey of OCR evaluation tools and metrics. *The 6th International Workshop On Historical Document Imaging And Processing*. pp. 13-18 (2021)
14 Oberbichler, S., Boroş, E., Doucet, A., Marjanen, J., Pfanzelter, E., Rautiainen, J., Toivonen, H. & Tolonen, M. Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. *Journal Of The Association For Information Science And Technology*. **73**, 225-239 (2022)
15 Hechl, S., Langlais, P., Marjanen, J., Oberbichler, S. and Pfanzelter, E., Digital interfaces of historical newspapers: opportunities, restrictions and recommendations. *Journal Of Data Mining & Digital Humanities*. (2021)
16 Schneider, P. Rerunning OCR: A Machine Learning Approach to Quality Assessment and Enhancement Prediction. *ArXiv Preprint ArXiv:2110.01661*. (2021)
17 Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. *The Craft Of Information Visualization*. pp. 364-371 (2003)
18 Van Strien, D., Beelen, K., Ardanuy, M., Hosseini, K., McGillivray, B. & Colavizza, G. Assessing the impact of OCR quality on downstream NLP tasks. (SCITEPRESS-Science,2020)
19 Tufte Edward, R. The visual display of quantitative information. (Cheshire, Connecticut: Graphic Press,2001)
20 Traub, M., Samar, T., Van Ossenbruggen, J., He, J., Vries, A. and Hardman, L. Querylog-based assessment of retrievability bias in a large newspaper corpus. *2016 IEEE/ACM Joint Conference On Digital Libraries (JCDL)*. pp. 7-16 (2016)
21 Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D. and Helberger, N. Recommenders with a mission: assessing diversity in news recommendations. *Proceedings Of The 2021 Conference On Human Information Interaction And Retrieval*. pp. 173-183 (2021)
22 Wieringa, M., Geenen, D., Es, K. and Nuss, J. The Fieldnotes Plugin: Making Network Visualization in Gephi accountable. *Good Data*. **14** pp. 277 (1988)

4.4 Towards an International Historical Newspaper Infrastructure

Clemens Neudecker (Staatsbibliothek zu Berlin, DE)
Maud Ehrmann (EPFL – Lausanne, CH)
Matteo Romanello (EPFL – Lausanne, CH)
Martin Volk (Universität Zürich, CH)
Lars Wieneke (C2DH – Esch-sur-Alzette, LU)
Dario Kampkaspar (TU Darmstadt, DE)

Portals and platforms that aggregate digitised newspapers from multiple sources and institutions have added great value for researchers, as they e.g. allow the comparative study of newspaper data from different countries and in multiple languages from within a uniform

environment. With few exceptions, digitised newspapers are commonly made available online via national portals and collections, or even fragmented across numerous online repositories, with each offering different features and functionalities for searching and accessing the data. Together, this makes working with digitised newspapers very tedious for researchers, and raises the need for standardised and modular information flows between systems [6]. Additionally, new tools and services are required for the accommodation of scholarly research requirements for content discovery and management, and to reflect their iterative, exploratory research workflows.

Different types of actors act and collaborate in the field of digitised newspapers. Libraries are predominantly interested in the continued digitisation, preservation and online presentation of their newspaper holdings, often providing only very basic ways to browse the digitised newspapers by title or date, and to a lesser extent, perform keyword searches when full text has been produced using Optical Character Recognition (OCR). Researchers, on the other hand, are in need of more advanced ways to access the data, build corpora from it, or explore it through quantitative aspects. Computer scientists, research software engineers and designers, finally, work on developing robust text and image processing approaches and scholarship-oriented (re)search interfaces. All actors contribute to advancing many aspects of historical newspaper access, processing and research, some focusing on breadth (e.g., libraries taking care of full collections), others on depth (e.g., research projects venturing into innovative prototypes on small or medium-sized collections).

Research interests are as diverse as the newspaper's contents, with use cases from different disciplines such as the humanities, social sciences, computational linguistics, computer science, or digital humanities. Similarly, a multitude of methods are currently used in the analysis and exploration of digitised newspapers either as images, text or a combination thereof. Furthermore, specific approaches are typically required for document and language processing to deal with challenges due to the primarily historical and multilingual nature of the newspaper content.

In an ideal world, portals like Europeana Newspapers[84] and *impresso*[85] would continue to aggregate additional newspaper data, and add more specialised functionalities for computational approaches to digitised historical newspapers.

Looking at the current situation, due to resource constraints, libraries struggle to offer more specialised functionalities as would be desired by researchers. The main focus of cultural heritage institutions involved with newspaper digitisation remains on increasing the volume of content available digitally for the general public, and providing long-term access to these digitised resources. Digital infrastructures that give access to digitised newspapers from multiple countries and in multiple languages are faced with the challenge of sustaining continued development, aggregation of additional newspaper content, and the integration of specific functionalities and interfaces for computational analysis of digitised newspapers, once funding runs out.

What could be simple and lightweight alternatives to better support a diverse research community interested in the computational analysis of digitised newspapers? And how could the cost and effort to start new research projects and collaborations around digitised newspapers be reduced? What is the essential required infrastructure, what improvements are achievable with reasonable time and effort and what could be a more ambitious, long-term vision for the international newspaper digitisation and research ecosystem?

[84] `https://www.europeana.eu/en/collections/topic/18-newspapers`
[85] `https://impresso-project.ch/app/`

Several initiatives have looked into cost-efficient ways to improve the provision of digitised newspapers for computational analysis. Already in 2013, Tim Sherratt coined the phrase "From portals to platforms" in his presentation at the LIANZA conference in New Zealand [1], suggesting that libraries should focus more on ways to publish their digitised data in machine-readable formats and via an Application Programming Interface (API), rather than building portals with sophisticated user interfaces. A main argument is that user interfaces will hardly ever be able to fully cater to all the diverse use cases, especially from researchers, as they are always constrained – pre-determined by a set of design decisions about what invites further exploration, or what is deemed necessary, relevant and useful, also for the greater public. Platforms that put a focus on ways for distributing data put the decisions back into the hands of the users, enabling them to obtain, interact with and analyse the data in their own preferred environment, and with the tools and methods of their own choice.

Another important perspective here is that of the US project *Collections as Data* [2]. The Principal Investigator of *Collections as Data* project, Thomas Padilla, was invited to participate in this Dagstuhl Seminar, but unfortunately could not attend. But the main ideas and concepts of *Collections as Data* were nevertheless present throughout the whole seminar, and especially in Sally Chambers evening lecture "Newspapers as Data: Challenges and Solutions". A core ambition of *Collections as Data* lies in the use of practical and cost-efficient means to better support computational research of cultural heritage data. This also includes the call to provide "actionable" collections, via e.g. Jupyter notebooks and the use of APIs [4], or making the data available in ways that allow using it for machine learning purposes [3].

An intermediate (ideal?) solution that retains the best of each approach would be to develop and maintain interfaces that offer both capabilities, with innovative, powerful, and appropriate functionalities for content search, discovery, and comparison on the one hand, and on the other (or on the back-end of the former) user-facing data dumps and APIs.

To summarise, even without cross-national newspaper portals offering advanced search and exploration tools for researchers (or prior to their development), stakeholders, that is to say libraries, computer scientists and digital humanists, can still contribute to improving digitised newspaper collections for computational analysis by researchers in multiple, simple and practical, and sustainable ways. The below recommendations aim to provide some guidance on best practices for further improving the possibilities for computational approaches to digitised newspapers based on simple and practical means:

1. libraries should strive to always expose digitised newspaper content (metadata, images and full text) via API, preferably using the APIs from the International Image Interoperability Framework (IIIF)[86]; additionally, modalities to download bulk data (dumps) with different selections (only images, only OCR, OCR as plain text, by newspaper title or language etc) in simple, machine-readable formats (CSV, JSON, TXT) are seen as very useful by researchers across multiple disciplines. As these datasets are often substantial in size, delivery mechanisms – as established for the exchange of data in biology and physics – should be implemented;
2. the choice of (meta)data formats in libraries should follow de-facto standards in newspaper digitisation like the Library of Congress maintained XML-based standards METS/ALTO[87] or IIIF manifests;

[86] `https://iiif.io/`
[87] `http://www.loc.gov/standards/mets/`; `https://www.loc.gov/standards/alto/`

3. information on the provenance (selection, digitisation and processing) of digitised newspaper collections should be documented and made publicly available (e.g. via the *Atlas of Digitised Newspapers*[88]) to identify and assess biases or to find contextual information that is required to understand the background of how the data was produced and its implications on the use of it for scholarly research;
4. when new newspaper digitisation projects are conceived, they should always include layout analysis and OCR; as a next step, standardised ways for encoding information about image content in digital newspapers could be investigated;
5. the full text of digitised newspapers should be annotated with language labels based on automatic language identification. This is particularly important in multilingual countries like Belgium, Luxembourg, Italy, or Switzerland, as it has a potential impact on downstream natural language processing (NLP) tasks as well as on how contents are made searchable via user-facing exploration interfaces;
6. whenever possible, the full text of digitised newspapers should be enriched with named entity recognition (NER) and entity disambiguation and linking (EL) to a multilingual knowledge base such as e.g. Wikidata;
7. for all automatic data processing and enrichments like OCR, NER, EL etc., there should be information made available with the newspaper content that allows users to quickly assess the performance quality of this automatic processing to aid in selection of newspaper content for dataset and corpus building;
8. notation and interchange formats for tool processes and semantic enrichments should be standardised with the aim to achieve interoperability between tools, text and annotations (this could building on existing initiatives, e.g. the Distributed Text Services[89] and the know-how of libraries and historical newspaper research projects;
9. article segmentation, the detection of reading order and layout evaluation are difficult open questions that will require more work and research, and the development of metrics and tools for quality control which find community agreement;
10. a cross-national meta-search engine, catalogue or registry could help researchers find out more easily what newspapers are already digitised and with what granularity (e.g. scans only vs. OCR full text vs. article separation). Such additional metadata and contextual information could be added into established and sustainable newspaper catalogues and repositories (e.g. ZDB[90]);
11. to support the widest possible range of use and reuse scenarios, open licences such as the Public Domain Mark or Creative-Commons-Zero should be advertised and used for digitised newspapers, as even CC-BY-SA-NC or CC-BY-SA can be prohibitive for some relevant usages (e.g. the training and distribution of machine learning models [5]).

Step by step, and in collaboration with all stakeholders, the implementation of the above recommendations could form the basis for a "Newspaper Network Notation Framework" (NNNF), similar to and following the same process as IIIF, from basic design decisions in the wider community to concrete specifications, and eventually standardised implementations of interoperable newspaper collections that are fit for computational approaches.

[88] https://www.digitisednewspapers.net/
[89] https://distributed-text-services.github.io/specifications/
[90] https://zdb-katalog.de/index.xhtml

References

1 Sherratt, T. From portals to platforms. Building new frameworks for user engagement (2013)
2 Padilla, T. Responsible Operations: Data Science, Machine Learning, and AI in Libraries. *OCLC Research Position Paper* (2019)
3 Lee, B.C.G. The "Collections as ML Data" Checklist for Machine Learning & Cultural Heritage (2022)
4 Candela, G., Saez, M.-D., Escobar, P., Marco-Such, M. Reusing digital collections from GLAM institutions. *Journal of Information Science*, 48, 251-267 (2022, 2)
5 Lassner, D., Neudecker, C., Coburger, J., Baillot, A. Publishing an OCR ground truth data set for reuse in an unclear copyright setting. *Zeitschrift für digitale Geisteswissenschaften* (2021)
6 Romanello, M., Ehrmann, M., Clematide, S. & Guido, D. The Impresso System Architecture in a Nutshell. *EuropeanaTech Insights*. (2020), `https://pro.europeana.eu/page/issue-16-newspapers#the-impresso-system-architecture-in-a-shell`, `https://infoscience.epfl.ch/record/283595`

Participants

- Kaspar Beelen
The Alan Turing Institute – London, GB
- Estelle Bunout
University of Luxembourg, LU
- Sally Chambers
Ghent University, BE & KBR, Royal Library of Belgium – Brussels, BE
- Simon Clematide
Universität Zürich, CH
- Mariona Coll-Ardanuy
The Alan Turing Institute – London, GB
- Mickaël Coustaty
University of La Rochelle, FR
- Marten Düring
University of Luxembourg, LU
- Maud Ehrmann
EPFL – Lausanne, CH
- Laura Hollink
Centrum Wiskunde & Informatica – Amsterdam, NL
- Stefan Jänicke
University of Southern Denmark – Odense, DK
- Axel Jean-Caurant
University of La Rochelle, FR
- Dario Kampkaspar
TU Darmstadt, DE
- Jana Keck
German Historical Institute Washington, US
- Yves Maurer
National Library of Luxembourg, LU
- Clemens Neudecker
Staatsbibliothek zu Berlin, DE
- Julia Noordegraaf
University of Amsterdam, NL
- Eva Pfanzelter
Universität Innsbruck, AT
- David A. Smith
Northeastern University – Boston, US
- Martin Volk
Universität Zürich, CH
- Lars Wieneke
C2DH – Esch-sur-Alzette, LU

Remote Participants

- Antoine Doucet
University of La Rochelle, FR
- Matteo Romanello
University of Lausanne, CH

Report from Dagstuhl Seminar 22301

Algorithmic Aspects of Information Theory

Phokion G. Kolaitis[*1], Andrej E. Romashchenko[*2], Milan Studený[*3], Dan Suciu[*4], and Tobias A. Boege[†5]

1 **University of California – Santa Cruz, US & IBM Research, US.**
kolaitis@ucsc.edu
2 **University of Montpellier – LIRMM, FR & CNRS, FR.**
andrei.romashchenko@lirmm.fr
3 **The Czech Academy of Sciences – Prague, CZ.** studeny@utia.cas.cz
4 **University of Washington – Seattle, US.** suciu@cs.washington.edu
5 **MPI für Mathematik in den Naturwissenschaften – Leipzig, DE.**
post@taboege.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22301 "Algorithmic Aspects of Information Theory".

Constraints on entropies constitute the "laws of information theory". These constraints go well beyond Shannon's basic information inequalities, as they include not only information inequalities that cannot be derived from Shannon's basic inequalities, but also conditional inequalities and disjunctive inequalities that are valid for all entropic functions. There is an extensive body of research on constraints on entropies and their applications to different areas of mathematics and computer science. So far, however, little progress has been made on the algorithmic aspects of information theory. In fact, even fundamental questions about the decidability of information inequalities and their variants have remained open to date.

Recently, research in different applications has demonstrated a clear need for algorithmic solutions to questions in information theory. These applications include: finding tight upper bounds on the answer to a query on a relational database, the homomorphism domination problem and its uses in query optimization, the conditional independence implication problem, soft constraints in databases, group-theoretic inequalities, and lower bounds on the information ratio in secret sharing. Thus far, the information-theory community has had little interaction with the communities where these applications have been studied or with the computational complexity community. The main goal of this Dagstuhl Seminar was to bring together researchers from the aforementioned communities and to develop an agenda for studying algorithmic aspects of information theory, motivated from a rich set of diverse applications. By using the algorithmic lens to examine the common problems and by transferring techniques from one community to the other, we expected that bridges would be created and some tangible progress on open questions could be made.

Seminar July 24–29, 2022 – http://www.dagstuhl.de/22301

2012 ACM Subject Classification Mathematics of computing → Information theory; Information systems → Database design and models; Mathematics of computing → Discrete mathematics; Mathematics of computing → Probabilistic inference problems

Keywords and phrases Information theory, Information inequalities, Conditional independence structures, Database query evaluation and containment, Decision problems

Digital Object Identifier 10.4230/DagRep.12.7.180

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Algorithmic Aspects of Information Theory, *Dagstuhl Reports*, Vol. 12, Issue 7, pp. 180–204
Editors: Phokion G. Kolaitis, Andrej E. Romashchenko, Milan Studený, and Dan Suciu

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Phokion G. Kolaitis
Andrej E. Romashchenko
Milan Studený
Dan Suciu

The goal of this seminar was to bring together researchers from several communities who share an interest in the methods and the uses of information theory. Participants included experts in information theory, databases, secret sharing, algorithms, and combinatorics. There were four tutorials, two from the information theory community and two from the database community, that helped define a common language and a common set of problems. There were several contributed talks, from experts in all these fields. The proof of one of the major open problems in information theory was announced at the workshop by not one, but, by *two* researchers, namely Cheuk Ting Li and Geva Yashfe, who used quite different techniques to independently solve this open problem. Overall, the workshop was a success.

Organization of the Seminar

The seminar was held between July 25-29, 2022 (Monday to Friday), and had 25 on-site participants, and 8 remote participants. Since the participants represented quite diverse communities, we started the first day with an introduction of each participant. The four tutorials were scheduled during the first two days: two tutorials on information inequalities and conditional independence were given by László Csirmaz and Milan Studený, and two tutorials on different aspects of database theory were given by Marcelo Arenas and Hung Ngo. All four tutorials were very well received, with many questions and lively discussions during and after the tutorials. There were 18 contributed talks in total, spread over all 5 days of the seminar. We scheduled two sessions to discuss open problems: one on Tuesday afternoon, and one on Thursday afternoon. The seminar concluded with an hour-long discussion assessing the seminar and contemplating future directions. Our collector, Tobias Boege, recorded all open problems, and later typed them for inclusion in this report.

Outcomes of the Seminar

There are several major outcomes:

- Having participants with very diverse backgrounds enabled us to exchange interesting ideas and problems. Information theorists became inspired by problems that arise in database research, while database theoreticians learned tools and techniques from information theory; almost all talks raised algorithmic questions that inspired people from the algorithms community.
- We have assembled a list of open problems, which we included here, and we also plan to publish independently. We hope that this list will help define the community interested in the algorithmic aspects of information theory, and will also inspire young researchers to contribute to this emerging area.

- At the end of the workshop the participants expressed a lot of interest in continuing to have some organized forum for discussing problems in information theory. One of us (Andrei Romashchenko) is planning to organize regular talks, to be made publicly available online, via Zoom.
- Everyone was happily surprised that one major problem in information theory was essentially settled during this seminar. The problem asks whether the implication problem for conditional independence statements is decidable. This problem has been studied since at least the early 80's, and has resisted any prior attempts at settling it. Cheuk Ting Li announced a proof of the undecidability of this problem, and presented the high-level structure of the proof; he had posted on arXiv a paper describing the proof just a few weeks before the seminar. Geva Yashfe had solved a different open problem: he showed that it is undecidable whether a given $(2^n - 1)$-dimensional vector is an almost entropic vector. Through discussions at the seminar, he realized that his proof can be extended to also prove that the implication problem for conditional independences is undecidable. He gave a presentation of his proof on the blackboard, during the seminar.

Acknowledgements

We are grateful to the Scientific Directorate and to the staff of the Schloss Dagstuhl – Leibniz Center for Informatics for their support of this seminar. We also wish to express our sincere thanks to Dr. Tobias Boege for collecting the abstracts of the talks and compiling the list of open problems.

2 Table of Contents

3 Overview of Talks

3.1 Open Problems on Information-Theoretic bounds for Database Query Answers

Mahmoud Abo Khamis (relationalAI – Berkeley, US)

Joint work of Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu
Main reference Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?", in Proc. of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017, pp. 429–444, ACM, 2017.
URL https://doi.org/10.1145/3034786.3056105

Information theory has been used to derive tighter bounds on the output sizes of database queries as well as to devise matching algorithms that can answer these queries within the derived bounds. Such bounds are typically derived by translating database statistics into constraints over (conditional) entropies. The query output size is then bounded by the maximum joint entropy subject to these constraints as well as Shannon inequalities. The most general form of this class of bounds is called the polymatroid bound [1].

In this talk, we present two open problems related to the polymatroid bound.

1. The first problem asks whether we can make the polymatroid bound stronger by adding conditional independence constraints that can be inferred from the structure of the database query.
2. The second problem asks whether we can utilize the query structure to infer constraints on the multivariate mutual information, in the same way we utilize (conditional) independence to infer a zero-constraint on the (conditional) mutual information between two sets of variables.

References

1 Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another? In Emanuel Sallinger, Jan van den Bussche, Floris Geerts (editors): Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, pp. 429–444 (2017)

3.2 Tutorial: a brief introduction to database theory

Marcelo Arenas (PUC – Santiago de Chile, CL)

Main reference Marcelo Arenas, Leonid Libkin: "An information-theoretic approach to normal forms for relational and XML data", J. ACM, Vol. 52(2), pp. 246–283, 2005.
URL https://doi.org/10.1145/1059513.1059519

In this talk, we will give an overview of some fundamental concepts in database theory: relational schemas, queries, data dependencies and normal forms. Besides, we will present a connection between normalization theory and information theory.

3.3 Approximate Implication for PGMs and Relational DBs

Batya Kenig (Technion – Haifa, IL)

Joint work of Batya Kenig, Dan Suciu
Main reference Batya Kenig, Dan Suciu: "Integrity Constraints Revisited: From Exact to Approximate Implication", Log. Methods Comput. Sci., Vol. 18(1), 2022.
URL https://doi.org/10.46298/lmcs-18(1:5)2022
Main reference Batya Kenig: "Approximate implication with d-separation", in Proc. of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI 2021, Virtual Event, 27-30 July 2021, Proceedings of Machine Learning Research, Vol. 161, pp. 301–311, AUAI Press, 2021.
URL https://proceedings.mlr.press/v161/kenig21a.html

The implication problem studies whether a set of conditional independence (CI) statements (antecedents) implies another CI (consequent), and has been extensively studied under the assumption that all CIs hold exactly. In this work, we drop this assumption, and define an approximate implication as a linear inequality between the degree of satisfaction of the antecedents and consequent. More precisely, we ask what guarantee can be provided on the inferred CI when the set of CIs that entailed it hold only approximately. We use information theory to define the degree of satisfaction, and prove several results. In the general case, no such guarantee can be provided. We prove that such a guarantee exists for the set of CIs inferred in directed graphical models, making the d-separation algorithm a sound and complete system for inferring approximate CIs. We also prove an approximation guarantee for independence relations derived from marginal and saturated CIs.

3.4 Recent advances in secret sharing

Amos Beimel (Ben Gurion University – Beer Sheva, IL)

Main reference Benny Applebaum, Amos Beimel, Oded Nir, Naty Peter: "Better secret sharing via robust conditional disclosure of secrets", in Proc. of the Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020, pp. 280–293, ACM, 2020.
URL https://doi.org/10.1145/3357713.3384293

A secret-sharing scheme allows to distribute a secret s among n parties such that only some predefined "authorized" sets of parties can reconstruct the secret s, and all other "unauthorized" sets learn nothing about s. For over 30 years, it was known that any (monotone) collection of authorized sets can be realized by a secret-sharing scheme whose shares are of size $2^{n-o(n)}$ and until recently no better scheme was known. In a recent breakthrough, Liu and Vaikuntanathan [1] have reduced the share size to $2^{0.994n+o(n)}$, and this was further improved by several follow-ups accumulating in an upper bound of $1.5^{n+o(n)}$ [2]. In this talk will survey the known results on secret-sharing schemes and present some ideas of the new constructions.

References

1 Tianren Liu, Vinod Vaikuntanathan: Breaking the circuit-size barrier in secret sharing. STOC 2018: 699-708
2 Benny Applebaum, Oded Nir: Upslices, Downslices, and Secret-Sharing with Complexity of $1.5n$. CRYPTO (3) 2021: 627-655

3.5 Universality of Gaussian conditional independence models

Tobias Andreas Boege (MPI für Mathematik in den Naturwissenschaften, DE)

Main reference T. Boege: "The Gaussian conditional independence inference problem", PhD thesis, Otto-von-Guericke-Universität Magdeburg, 2022.
URL https://dx.doi.org/10.25673/86275

We study statistical models of jointly normal random variables defined by conditional independence (CI) constraints. These models are semialgebraic sets. In this talk I present a number of so-called "universality theorems" for these models:

1. all real algebraic numbers are necessary to witness that a given set of conditional independence constraints is consistent;
2. the problem of deciding consistency (or, equivalently, solving the conditional independence implication problem for Gaussians) is complete for the existential theory of the reals; and
3. all homotopy types of semialgebraic sets are attained by oriented Gaussian CI models.

These results parallel the celebrated universality theorems in matroid theory due to MacLane, Mnëv and Sturmfels.

3.6 Tutorial on information inequalities

László Csirmaz (Alfréd Rényi Institute of Mathematics – Budapest, HU)

The information content of the marginals of (finitely many) jointly distributed random variables reveals many important properties of the distribution, such as functional dependency or (conditional) independence. The information content is measured by the entropy, and information inequalities compare these marginal entropies. The entropy region is the collection of the entropy vectors of these distributions indexed by the non-empty subsets of the variables. The main focus of the talk is on discrete distributions, but many of the methods and concepts apply, with some modification, for linear, continuous, Gaussian, or quantum distributions. Points in the entropy region satisfy the basic Shannon inequalities [3], and the entropy region is bounded by collection of these inequalities, known as the Shannon bound. Points within the Shannon-bound are the also called polymatroids. The first non-Shannon entropy inequality was discovered by Zhang and Yeung [4]. The method obtaining this inequality was formalized and generalized by [1]. The general idea is to find an operation which preserves entropic points, but does not preserve polymatroids in general. Typical operations are restricting, factoring, conditioning, tightening, etc. Unfortunately they preserve both entropic points [2] and polymatroids, but might help in reducing the computational complexity of obtaining bounds on (or parts of) the entropy region. The two-step process of an adequate operation, dubbed as Copy Lemma, can be phrased as follows: first extend the distribution by adding identical copies of some of the variables (this step does not work in the quantum setting because of the no-cloning theorem), and then redefine the distribution so that the new and old variables become conditionally independent given the remaining variables. A known set of inequalities for the larger set of variables (e.g., the basic Shannon inequalities) might imply additional constraints on the old variables. The Copy Lemma can be considered to be

a special case of the Maximal Entropy Method. Fix some marginal distributions on $M > N$ random variables to be identical with certain marginal distributions on N random variables, and take the distribution on M with the maximal possible entropy. This distribution will have strong structural properties: depending on the fixed marginals several conditional independences hold. Harvesting them might yield new constrains on the entropies of the original distribution. The presentation concludes with several computational challenges.

1. Obtaining additional information inequalities requires embedding a distribution on N variables to a distribution on $M > N$ variables, and then applying known information inequalities on M variables. For M larger than 15 even working with the Shannon bounds is prohibitively expensive. Devise a method which simplifies this treatment.
2. Look systematically for limitations of the above methods with small number of added variables.
3. Study how the above technique can be applied for quantum information, and obtain new quantum-information inequalities.
4. A repository of (discrete) distributions with four variables whose convex combination approximates the complete entropic region might be extremely useful in answering practical / theoretical questions.

References

1 R. Dougherty, C. Freiling, K. Zeger, Non-Shannon information inequalities in four random variables, *ArXiv:1104.3602* (2011)
2 F. Matúš, Two constructions on limits of entropy functions, *IEEE Trans. Inform. Theory*, vol 53(1) (2007) pp. 320–330
3 R. W. Yeung, *A first course in information theory*. Kluwer Academic Publishers, New York, 2002
4 Z. Zhang, R. W. Yeung, On characterization of entropy function via information inequalities, *Proc IEEE Trans. Inform. Theory*, vol 44(4) (1998) pp. 1440–1452

3.7 Linear Programming Technique in the Search for Lower Bounds in Secret Sharing

Oriol Farràs (Universitat Rovira i Virgili – Tarragona, ES)

Main reference Oriol Farràs, Tarik Kaced, Sebastià Martín Molleví, Carles Padró: "Improving the Linear Programming Technique in the Search for Lower Bounds in Secret Sharing", in Proc. of the Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I, Lecture Notes in Computer Science, Vol. 10820, pp. 597–621, Springer, 2018.
URL https://doi.org/10.1007/978-3-319-78381-9_22

Secret sharing scheme is a method by which a dealer distributes shares to parties such that only authorized subsets of parties can reconstruct the secret. The information ratio is an indicator of the efficiency of a secret sharing scheme; it is the size in bits of the largest share of the scheme divided by the size of the secret.

This talk is focused on the following optimization problem: Given a family of subsets of parties F, find the infimum of the information ratio of all secret sharing schemes whose authorized subsets are the ones in F. Lower bounds on this optimal value can be computed by solving linear programming problems involving information inequalities.

We present improvements of this linear programming technique that use the Ahlswede-Körner lemma and the common information of random variables, avoiding the use of explicit non-Shannon information inequalities. Moreover, we show results of the application of this technique to the classification of representable matroids. The results presented in this talk were published in references.

References

1 Michael Bamiloshin, Aner Ben-Efraim, Oriol Farràs, Carles Padró: Common information, matroid representation, and secret sharing for matroid ports. Des. Codes Cryptogr. 89(1): 143-166 (2021).
2 Oriol Farràs, Tarik Kaced, Sebastià Martín Molleví, Carles Padró: Improving the Linear Programming Technique in the Search for Lower Bounds in Secret Sharing. IEEE Trans. Inf. Theory 66(11): 7088-7100 (2020).

3.8 Information Complexity

Yuval Filmus (Technion – Haifa, IL)

Main reference Mark Braverman, Ankit Garg, Denis Pankratov, Omri Weinstein: "From information to exact communication", in Proc. of the Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013, pp. 151–160, ACM, 2013.
URL https://doi.org/10.1145/2488608.2488628

This talk surveys work by IMU Abacus Medal winner Mark Braverman and others.

The basic question asked by information theory is how many bits of communication are needed to transmit information from A to B. In contrast, communication complexity studies the amount of communication needed between two parties, A and B, who want to compute a joint function of their inputs. Information complexity is an approach to communication complexity using information theory, specifically the notion of Information Complexity which is analogous to entropy. Using this approach, it is possible to compute asymptotically tight bounds on communication complexity. For example, Braverman, Garg, Pankratov and Weinstein [1] considered the fundamental problem of Set Disjointness, in which A and B each hold a subset of $\{1, \ldots, n\}$, and their goal is to decide whether the two subsets are disjoint. They showed that the exact communication complexity of this function (with vanishing error) is roughly $0.4825\ldots n$, where 0.4825... is an explicitly computable constant.

References

1 Mark Braverman, Ankit Garg, Denis Pankratov, Omri Weinstein: From information to exact communication. STOC 2013: 151-160

3.9 Dependencies in team semantics

Miika Hannula (University of Helsinki, FI)

Joint work of Miika Hannula, Jonni Virtema
Main reference Miika Hannula, Jonni Virtema: "Tractability frontiers in probabilistic team semantics and existential second-order logic over the reals", Ann. Pure Appl. Log., Vol. 173(10), p. 103108, 2022.
URL https://doi.org/10.1016/j.apal.2022.103108

According to the traditional Tarski's truth definition the semantics of first-order logic is defined with respect to an assignment of values to the free variables. In team semantics, truth is defined with respect to a set (or a probability distribution) of such assignments. This allows modeling concepts that inherently arise only in the presence of multitudes. Examples of concepts available in team semantics, but not in the Tarski semantics, include concepts of dependence and independence. In this talk we will take a brief look at how in team semantics one can analyze the relationships between relational and probabilistic dependencies as well as their interplay with logical operations.

3.10 Entropy Inequalities, Lattices and Groups

Peter Harremoës (Niels Brock Copenhagen Business College, DK)

Main reference Peter Harremoës: "Entropy Inequalities for Lattices", Entropy, Vol. 20(10), p. 784, 2018.
URL https://doi.org/10.3390/e20100784

The notion of entropy inequalities of Shannon and non-Shannon type have mostly been studied for formal power sets of random variables. Such power sets form Boolean lattices with inclusion as ordering. For applications in database theory and the study of Bayesian networks and similar graphical models of independence it is also relevant to study other lattices than the Boolean lattices. In general an element in a lattice should correspond to a set of variables, and one element in the lattice dominates another point if and only if the first corresponding set of variables determine the corresponding second set of variables. For any lattice one may ask which entropy inequalities that will hold for variables that are related in a way determined by the lattice. For certain classes of lattices all entropy inequalities are of the Shannon type and one goal of this research is to identify these lattices. There is also a link between entropy inequalities and inequalities for subgroups of a group. It turns out that this is related to the question of whether a specific lattice can be represented as the lattice of subgroups of a given group. This relation can be used see how certain codes used in cryptography and channel coding can be realized by the algebraic structure of certain groups.

3.11 On the undecidability of conditional independence implication

Cheuk Ting Li (The Chinese University of Hong Kong, HK)

Main reference Cheuk Ting Li: "The Undecidability of Conditional Affine Information Inequalities and Conditional Independence Implication with a Binary Constraint", in Proc. of the IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021, pp. 1–6, IEEE, 2021.
URL https://doi.org/10.1109/ITW48936.2021.9611489

The conditional independence implication problem is to decide whether several statements on the conditional independence among random variables implies another such statement. In this talk, we show that this problem is undecidable if we also allow imposing cardinality constraints (e.g., "X is a binary random variable"). This is proved via a reduction from the domino problem about tiling the plane with a set of tiles. We will also briefly discuss a recent preprint which establishes the undecidability of the original conditional independence implication problem (without cardinality constraints) and related results, e.g., the undecidability of conditional information inequalities and network coding.

3.12 Tutorial on an Information Theoretic Approach to Estimating Query Size Bounds

Hung Ngo (relationalAI – Berkeley, US)

Joint work of Mahmoud Abo Khamis, Sungjin Im, Hossein Keshavarz, Phokion Kolaitis, Ben Moseley, Long Nguyen, Hung Ngo, Kirk Pruhs, Dan Suciu, Alireza Samadian Zakaria
Main reference Mahmoud Abo Khamis, Hung Q. Ngo, Dan Suciu: "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?", in Proc. of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017, pp. 429–444, ACM, 2017.
URL https://doi.org/10.1145/3034786.3056105

Cardinality estimation is one of the most important problems in database management. One aspect of cardinality estimation is to derive a good upper bound on the output size of a query, given a statistical profile of the inputs. In recent years, a promising information-theoretic approach was devised to address this problem, leading to robust cardinality estimators which are used in practice.

The information theoretic approach led to many interesting open questions surrounding optimizing a linear function on the almost-entropic or polymatroidal cones. This talk briefly introduces the problem, the approach, summarizes some known results, and lists open questions.

3.13 Term Coding

Søren Riis (Queen Mary University of London, GB)

In this presentation, I introduce Term Coding (TC), which can be seen as an interface between Universal Algebra and Coding Theory. The work grew out of research related to Information flows and Information bottlenecks and the relationship to multiuser information

theory [1, 2, 3, 4]. A large class of (finite) mathematical structures, e.g. universal algebras, can be defined by various equations. Traditionally, the main focus is on systems that satisfy the equations for all points in the domain. The concern in TC is finding structures (codes) that meet the defining equations for many but not necessarily all points. TC provide a general framework for (dynamic) network coding, index coding, and graph guessing games and is related to non-Shannon information inequalities and other advances in Information Theory [4].

References

1 Riis, S., 2007. Reversible and irreversible information networks. IEEE Transactions on Information Theory, 53(11), pp.4339-4349.
2 Gadouleau, M. and Riis, S., 2011. Graph-theoretical constructions for graph entropy and network coding based communications. IEEE Transactions on Information Theory, 57(10), pp.6703-6717.
3 Riis, S. and Gadouleau, M., 2011, July. A dispersion theorem for communication networks based on term sets. In 2011 IEEE International Symposium on Information Theory Proceedings (pp. 593-597). IEEE.
4 Baber, R., Christofides, D., Dang, A.N., Vaughan, E.R. and Riis, S., 2016. Graph guessing games and non-Shannon information inequalities. IEEE Transactions on Information Theory, 63(7), pp.4257-4267.
5 Riis, S. and Gadouleau, M., 2019. Max-flow min-cut theorems on dispersion and entropy measures for communication networks. Information and Computation, 267, pp.49-73.

3.14 A couple of unusual information inequalities and their applications

Andrej E. Romashchenko (University of Montpellier – LIRMM, FR & CNRS, FR)

Joint work of Emirhan Gürpinar, Andrej Romashchenko
Main reference Emirhan Gürpinar, Andrei E. Romashchenko: "Communication Complexity of the Secret Key Agreement in Algorithmic Information Theory", CoRR, Vol. abs/2004.13411, 2020.
URL https://arxiv.org/abs/2004.13411

It is known that the mutual information of a pair of objects x and y is equal to the size of the largest shared secret key that two parties (holding as their inputs x and y respectively) can establish via a communication protocol with interaction on a public channel. We discuss communication complexity of this problem and show show that a tight lower bound on the communication complexity can be proven with help of the expander mixing lemma combined with information inequalities.

References

1 Emirhan Gürpinar, Andrej Romashchenko. *Communication Complexity of the Secret Key Agreement in Algorithmic Information Theory*. arXiv:2004.13411

3.15 Conditional Ingleton inequalities

Milan Studený (The Czech Academy of Sciences – Prague, CZ)

Main reference Milan Studený: "Conditional Independence Structures Over Four Discrete Random Variables Revisited: Conditional Ingleton Inequalities", IEEE Transactions on Information Theory, Vol. 67(11), pp. 7030–7049, 2021.
URL https://doi.org/10.1109/TIT.2021.3104250

Linear information inequalities valid for entropy functions induced by discrete random variables play an important role in the task to characterize discrete conditional independence structures [3, 4, 5]. Specifically, the so-called conditional Ingleton inequalities in the case of 4 random variables are in the center of interest: these are valid under conditional independence assumptions on the inducing random variables. The four inequalities of this form were earlier revealed: by Yeung and Zhang in 1997 [7], by Matúš in 1999 [5] and by Kaced and Romashchenko in 2013 [2]. In a recent 2021 paper [6] the fifth inequality of this type was found. These five information inequalities can be used to characterize all conditional independence structures induced by four discrete random variables. One of open problems in that 2021 paper was whether the list of conditional Ingleton inequalities over 4 random variables is complete: the analysis can be completed by a recent finding of Boege [1] answering that question.

References

1 T. Boege: No eleventh conditional Ingleton inequality. A manuscript (2022) available at `https://arxiv.org/abs/2204.03971`.
2 T. Kaced, A. Romashchenko: Conditional information inequalities for entropic and almost entropic points. *IEEE Transactions on Information Theory 59* (2013), 7149–7167.
3 F. Matúš and M. Studený: Conditional independences among four random variables I. *Combinatorics, Probability and Computing 4* (1995), 269–278.
4 F. Matúš: Conditional independences among four random variables II. *Combinatorics, Probability and Computing 4* (1995), 407–417.
5 F. Matúš: Conditional independences among four random variables III: final conclusion. *Combinatorics, Probability and Computing 8* (1999), 269–276.
6 M. Studený: Conditional independence structures over four discrete random variables revisited: conditional Ingleton inequalities. *IEEE Transactions on Information Theory 67* (2021), 7030-7049.
7 Z. Zhang, R.W. Yeung: A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. on Inform. Theory 43* (1997), 1982–1986.

3.16 Tutorial on conditional independence implication problem

Milan Studený (The Czech Academy of Sciences – Prague, CZ)

URL https://dblp.dagstuhl.de/pid/34/5033.html

The beginning of the tutorial was a brief overview of classic results on conditional independence inference. Then basic concepts were introduced: discrete random vector and conditional independence concept. Basic observation about discrete structures is that they form a finite lattice and can be characterized in terms of the so-called Horn clauses. After that the concept of a semi-graphoid was introduced and relevant partial axiomatizability results recalled: this

concerns marginal and saturated independence [4], functional dependence [6] and relative two-antecedental completeness of semi-graphoidal inference [9]. The inspiration from the theory of relational databases for these results was explained [1, 8]. Basic information-theoretical measures were defined and their relation to the entropic function and the multiinformation functions recalled. A substantial part of the talk was devoted to algorithmic aspects of conditional independence inference, where the concept of a structural semi-graphoid plays the crucial role [7, 3]. The last part of the talk dealt with special conditional independence implications valid in case of Gaussian conditional independence structures [5, 2].

References

1 W. W. Armstrong (1974). Dependency structures of database relations. In Information Processing 74, North Holland, 580-583.
2 T. Boege, A. D'Alì, T. Kahle, B. Sturmfels (2019). The geometry of gaussoids. Foundations of Computational Mathematics 19 (4), 775-812.
3 R. Bouckaert, R. Hemmecke, S. Lindner, M. Studený (2010). Efficient algorithms for conditional independence inference. Journal of Machine Learning Research 11, 3453-3479.
4 D. Geiger, J. Pearl (1993). Logical and algorithmic properties of conditional independence and graphical models. Annals of Statistics 21 (4), 2001-2021.
5 R. Lněnička, F. Matúš (2007). On Gaussian conditional independence structures. Kybernetika 43 (3), 327-342.
6 F. Matúš (1991). Abstract functional dependency structures. Theoretical Computer Science 81 (1), 117-126.
7 M. Niepert (2009). Logical inference algorithms and matrix representations for probabilistic conditional independence. In 25th UAI conference, AUAI Press, 428-435.
8 Y. Sagiv and S. F. Walecka (1982). Subset dependencies and completeness result for a subclass of embedded multivalued dependencies. Journal of Association for Computing Machinery 29 (1), 103-117.
9 M. Studený (1997). Semigraphoids and structures of probabilistic conditional independence. Annals of Mathematics and Artificial Intelligence 21(1), 71-98.

3.17 Max-Information Inequalities and the Domination Problem

Dan Suciu (University of Washington – Seattle, US)

Joint work of Dan Suciu, Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Ngo
Main reference Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, Dan Suciu: "Bag Query Containment and Information Theory", ACM Trans. Database Syst., Vol. 46(3), pp. 12:1–12:39, 2021.
URL https://doi.org/10.1145/3472391

A max-information inequality is an inequality involving linear expressions of entropic terms, and one occurrence of max. We consider the problem: given a max-information inequality, check if it holds for all entropic vectors. E.g. $\max(H(XY), H(YZ), H(ZX)) \geq {}^2/_3 H(XYZ)$ is valid. It is open whether this problem is decidable.

We say that a structure B "dominates" a structure A, if, for any other structure C, the number of homomorphisms $A \to C$ is less than or equal to the number of homomorphisms $B \to C$. We consider the problem: given two structures A, B, where B is acyclic, check whether B dominates A. The domination problem is equivalent to the "conjunctive query containment problem under bag semantics", which is of interest in database theory. It is open whether this problem is decidable.

We prove that these two problems are computationally equivalent. In particular, any progress on the decidability or undecidability of one of these problems will automatically carry over to the other problem.

3.18 A Conditional Information Inequality and Its Combinatorial Applications

Nikolay K. Vereshchagin (NRU Higher School of Economics – Moscow, RU)

Joint work of Tarik Kaced, Andrej Romashchenko, Nikolay K. Vereshchagin
Main reference Tarik Kaced, Andrei E. Romashchenko, Nikolai K. Vereshchagin: "A Conditional Information Inequality and Its Combinatorial Applications", IEEE Trans. Inf. Theory, Vol. 64(5), pp. 3610–3615, 2018.
URL https://doi.org/10.1109/TIT.2018.2806486

We show that the inequality $H(A \mid B, X) + H(A \mid B, Y) < H(A \mid B)$ for jointly distributed random variables A, B, X, Y, which does not hold in general case, holds under some natural condition on the support of the probability distribution of A, B, X, Y. This result generalizes a version of the conditional Ingleton inequality: if for some distribution $I(X : Y \mid A) = H(A \mid X, Y) = 0$, then $I(A : B) < I(A : B \mid X) + I(A : B \mid Y) + I(X : Y)$.

We present one application of this result. Assume that a family $\mathcal{F}$ of pair-wise disjoint "squares" $S \times S \subset U \times V$ is given (U, V are fixed finite sets). Assume that for each $u \in U$ there are at least L squares in $\mathcal{F}$, whose first projection covers u, and similarly, for each $v \in V$ there are at least R squares in F, whose second projection covers v. Then $|\mathcal{F}| \geq LR$.

3.19 When are Exhaustive Minimal Lists of Information Inequalities Scalable?

John MacLaren Walsh (Drexel University – Philadelphia, US)

Joint work of Yirui Liu, John MacLaren Walsh
Main reference Yirui Liu, Ph.D. Dissertation – Drexel University, October 2021.
Main reference Yirui Liu, John MacLaren Walsh: "Linear Complexity Entropy Regions", in Proc. of the IEEE International Symposium on Information Theory, ISIT 2021, Melbourne, Australia, July 12-20, 2021, pp. 1642–1647, IEEE, 2021.
URL https://doi.org/10.1109/ISIT45174.2021.9518030

Exhaustively determining the entropy region, and the information inequalities that describe it, form a tantalizingly fundamental problem in multi-terminal information theory. Among other equivalences, determining all information inequalities has been shown to be equivalent to determining the capacity regions of all networks under network coding, as well as determining all inequalities linking sizes of intersections of subgroups of a common group. Faces of the entropy region, in turn, dictate fundamental possible conditional independence relations among a series of random variables.

A key observation, however, is that many of these ultimate uses of the entropy region, both fundamental and applied, need only study the relationship between entropies of subsets that lie within a very small subfamily of the powerset of all subsets of the inscribed random variables. That is to say, were one able to exhaustively characterize the projection of the

entropy region onto only the (sub)family of exclusively subsets involved in a problem of interest, one may provide the fundamental limits in that problem while the harder problem of determining the entire entropy region on the powerset remains unsolved. A natural question one may then ask is, which types of systems of subsets enable one to provide an exhaustive characterization of the associated projection of the entropy region onto only these subsets? Furthermore, which types of such systems of subsets yield the closure of the projected entropy region to be polyhedral? Which of those systems of subsets yield polyhedra enable a description complexity, as measured by the total number of involved inequalities, which scales nicely, even linearly, in the size of the problem, measured as the number of random variables?

In this talk, we set about characterizing some of these families of subsets with scalable complexity through the idea of pasting the entropy regions of small, overlapping, subsets to obtain bounds on associate projections of the entropy region on their union. A straightforward argument shows that this pasting construction easily yields outer bounds, and as such, attention shifts to when these outer bounds are tight. Examples of infinite families of subsets where the pasted entropy regions exhaustively characterize the associated projection of the larger entropy region are detailed. Among these are included cases where the associated projection of the entropy region has a number of required minimal inequalities that scales linearly with the number of random variables. Moreover, a construction proves cases where such pasted outer bounds are guaranteed to be loose.

Bearing these cases where pasting small entropy regions together only yields a loose outer bound for the true entropy region in mind, attention then shifts to finding inner bound constructions whose inner bound property is preserved under pasting. Of particular interest is in the inner bound to the entropy region formed by the set of inequalities linking linear ranks which dictates the part of the entropy region reachable by time sharing linear codes, as well as those linked with quasi-uniform distributions. A first inner-bound preserving technique pastes together integral polyhedral quasi-uniform bound on a chain of sets in the overlap of their ground set. A second inner-bound preserving pasting technique based on requiring the existence of consistent common informations is then also provided for these types of inner bounds. These constructions, together with the constructions that correctly characterize the associated projections of the entropy region, form a substantial family of composable constructions that can be used to create inner bounds for the entropy region of controllable complexity.

3.20 Graph Information Ratio

Lele Wang (University of British Columbia – Vancouver, CA)

Joint work of Ofer Shayevitz, Lele Wang
Main reference Lele Wang, Ofer Shayevitz: "Graph Information Ratio", SIAM J. Discret. Math., Vol. 31(4), pp. 2703–2734, 2017.
URL https://doi.org/10.1137/16M1110066

We introduce the notion of information ratio $\mathrm{Ir}(H/G)$ between two (simple, undirected) graphs G and H, defined as the supremum of ratios k/n such that there exists a mapping between the strong products G^k to H^n that preserves non-adjacency. Operationally speaking, the information ratio is the maximal number of source symbols per channel use that can be reliably sent over a channel with a confusion graph H, where reliability is measured w.r.t. a

source confusion graph G. Various results are provided, including in particular lower and upper bounds on $\mathrm{Ir}(H/G)$ in terms of different graph properties, inequalities and identities for behavior under strong product and disjoint union, relations to graph cores, and notions of graph criticality. Informally speaking, $\mathrm{Ir}(H/G)$ can be interpreted as a measure of similarity between G and H. We make this notion precise by introducing the concept of information equivalence between graphs, a more quantitative version of homomorphic equivalence. We then describe a natural partial ordering over the space of information equivalence classes, and endow it with a suitable metric structure that is contractive under the strong product. Various examples and open problems are discussed.

3.21 On entropic and almost-entropic representability of matroids

Geva Yashfe (The Hebrew University of Jerusalem, IL)

Joint work of Lukas Kühne, Geva Yashfe
Main reference Lukas Kühne, Geva Yashfe: "On entropic and almost multilinear representability of matroids", CoRR, Vol. abs/2206.03465, 2022.
URL https://doi.org/10.48550/arXiv.2206.03465

This talk discusses some recent results obtained jointly with Lukas Kühne and announces one new theorem. There is no algorithm which, given a matroid M,

1. Decides whether M is entropic.
2. Decides whether M is multilinear.
3. Decides whether M is almost-multilinear.
4. Decides whether M is almost-entropic.

(The last theorem was proved during the conference after an inspiring discussion with Janneke Bolt, Andrej Romashchenko, and Alexander Shen.)

Here a matroid M is a polymatroid with values in the natural numbers (including 0) and satisfying that every singleton has rank at most 1. It is entropic if, as a polymatroid, its ray intersects the entropic cone. It is almost-entropic if it is in the closure of the entropic cone. The multilinear variants are about the analogous cones of linear rank functions.

A corollary of these theorems is that the conditional independence problem and its "approximate" variant are undecidable. Closely related results in the non-approximate setting have been obtained by Cheuk-Ting Li (preceding ours by some weeks) and have also been presented at this conference.

3.22 Machine-Proving of Entropy Inequalities

Raymond W. Yeung (The Chinese University of Hong Kong, HK)

Joint work of Laigang Guo, Raymond W. Yeung
Main reference Laigang Guo, Raymond W. Yeung, Xiao-Shan Gao: "Proving Information Inequalities and Identities with Symbolic Computation", in Proc. of the IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022, pp. 772–777, IEEE, 2022.
URL https://doi.org/10.1109/ISIT50566.2022.9834774

The entropy function plays a central role in information theory. Constraints on the entropy function in the form of inequalities, viz. entropy inequalities (often conditional on certain Markov conditions imposed by the problem under consideration), are indispensable tools for proving converse coding theorems. In this talk, I will give an overview of the development of machine-proving of entropy inequalities for the past 25 years. To start with, I will present a geometrical framework for the entropy function, and explain how an entropy inequality can be formulated, with or without constraints on the entropy function. Among all entropy inequalities, Shannon-type inequalities, namely those implied by the nonnegativity of Shannon's information measures, are best understood. We will focus on the proving of Shannon-type inequalities, which in fact can be formulated as a linear programming problem. I will discuss ITIP, a software package originally developed for this purpose in the mid-1990s, as well as some of its later variants. In 2014, Tian successfully characterized the rate region of a class of exact-repair regenerating codes by means of a variant of ITIP. This is the first nontrivial converse coding theorem proved by a machine. At the end of the talk, I will discuss some recent progress in speeding up the proving of entropy inequalities.

4 Open problems

The following problems have been collected from discussions, talks, and open problem sessions. The problems have been grouped into several different a themes and are followed by references to the literature. The person posing each problem is indicated in square brackets after the statement of the problem.

Secret sharing and cryptography

(1.1) The share size of a perfect secret sharing scheme with n participants can be bounded by $\mathcal{O}(2^{0.525n}) \cap \Omega(n/\log n)$. Can the upper bound be improved $\mathcal{O}(2^{cn})$ with $c < 1/2$? [Amos Beimel]

(1.2) Is there a secret sharing scheme which can be realized by an abelian group but not by a field? [László Csirmaz]

(1.3) Two parties want to compute the `OR` of their random bits. What is the minimal amount of information either of them has to disclose about their bit? How does this number depend on the number of rounds? [Alexander Shen]

Peculiarities of the entropy region

(2.1) Consider the set $\mathbf{S}_{ij}$ inside of Γ_4^* defined in [11, Section VII]. Theorem 5 in the same paper shows that the infimal Ingleton score in Γ_4^* is attained in this region and Example 2 exhibits a polymatroid in it which refutes the 4-atom conjecture as formulated in [6]. By how far was the 4-atom conjecture off? That is, what is the infimal Ingleton score? Is it an algebraic number? Which distributions reach it? [László Csirmaz]

(2.2) According to the experiments by Csirmaz [4] and in the coordinate system chosen there, the face $\{\,\beta = 0\,\}$ of $\mathbf{S}_{ij}$ contains entropic points whose distributions have at most 8 states per variable but there is a gap in his visualization between the interior and the face. Can this face be approximated arbitrarily well at all from the interior with only a bounded alphabet size? [László Csirmaz]

(2.3) Let Γ_n^* denote the entropy region of discrete random variables with finite support and Γ_n^∞ the entropy region of discrete random variables with countable support. Clearly $\Gamma_n^* \subseteq \Gamma_n^\infty \subseteq \overline{\Gamma_n^*} = \overline{\Gamma_n^\infty}$. Are the containments strict? [Tobias Boege]

(2.4) Are there "holes" on the boundary of $\overline{\Gamma_n^*}$? More concretely, is there a ray on one of its faces and four numbers $x < y < y' < z$ such that on this ray the intervals (x, y) and (y', z) parametrize entropic points and the segment parametrized by (y, y') contains no entropic point? [László Csirmaz]

(2.5) Is there an extreme ray of $\overline{\Gamma_n^*}$ which contains no entropic point? [John MacLaren Walsh]

(2.6) Fix n discrete random variables and add another one. Which entropy profiles arise? Specifically, look at the triples $H(W)$, $H(A|W)$, $H(B|W)$ for fixed A, B and arbitrary W. [Alexander Shen]

(2.7) Are the interior of the entropy region and its complement effectively open sets? [Alexander Shen]

Information quantities

(3.1) Let A and B be jointly distributed and consider the optimization problem $\sup I(A : B|X)$ for X jointly distributed with (A, B). Is the supremum attained? How to compute it as a function of the joint probability table for A and B? What if $A \perp\!\!\!\perp B$? [Nikolay Vereshchagin]

(3.2) Suppose the joint distribution of A, B, C factors into $p(a, b, c) = p_1(a, b) \cdot p_2(a, c) \cdot p_3(b, c)$. What is the minimal value of $I(A : B : C) = I(A : B) - -I(A : B \mid C)$? Which assumptions on the distribution guarantee non-negativity? [Mahmoud Abo Khamis]

(3.3) Let $\Delta(A, B, C) \coloneqq I(A : B \mid C) + I(A : C \mid B) + I(B : C \mid A)$, $\Delta'(A, B) \coloneqq \inf_C \Delta(A, B, C)$ and $\Gamma(A, B) = \sup_{X,Y}[I(A : B) - -I(A : B|X) - -I(A : B|Y) - -I(X : Y)]$. Then $\Gamma(A, B) \le \Delta'(A, B)$. Can the inequality be strict? [Alexander Shen]

(3.4) Let S be a finite set of triples which is closed under rotation and $w : S \to \mathbb{R}$ a weight function. Consider rotation-invariant distributions of three discrete random variables A, B, C whose support are the triples in S and the following optimization problem:

$$\sup_{ABC} \left(\frac{1}{3}\Big(H(A) + H(B) + H(C)\Big) + H(ABC) - \sup_{XYZ \sim ABC} H(XYZ) + \mathbb{E}[w(A, B, C)] \right).$$

The constraint $XYZ \sim ABC$ above means all distributions X, Y, Z which have the same 1-marginals as A, B, C. How to compute the optimum? This problem comes up in fast matrix multiplication theory. See [10] for the background and a relaxation. [Yuval Filmus]

(3.5) Is there a zero-one law in information transmission? Suppose a transmission task with only one restricted link. Is there a rate threshold below which correct transmission is only possible with negligible probability and above which there exists an encoding that ensures correct transmission with high probability? [Alexander Shen]

Suppose that a pair of random variables A and B is given. Add a new, jointly distributed random variable X and record the conditional entropies $\big(H(A|X), H(B|X), H(AB|X)\big)$. This yields a point in $\mathbb{R}^3$. The collection of all such points over all X forms a closed convex subset of $\mathbb{R}^3$, the *extension profile* $\mathrm{Ext}_1(A, B)$. Adding k random variables $X_1, \ldots, X_k$ to the pair (A, B) and recording all the conditional entropies $\big(H(A|X_I), H(B|X_I), H(AB|X_I)$ for all non-empty subvectors X_I of $(X_1, \ldots, X_k)$ results in higher extension profiles $\mathrm{Ext}_k(A, B)$.

(3.6) Does $\mathrm{Ext}_1(A, B)$ determine $\mathrm{Ext}_k(A, B)$ for all $k \geq 1$? This is, in spirit, similar to Grothendieck's reconstruction principle in geometry. [Rostislav Matveev]

Information inequalities

We consider various entropy-like regions Θ and the collections of linear inequalities which are satisfied by all points in Θ. These are contained in the dual cone $\Theta^\vee$.

(4.1) Is $(\Gamma_4^*)^\vee$ semialgebraic? The missing piece in an attempt in [7] to answer this question negatively is that the following Ingleton inequality is essentially conditional:

$$\begin{aligned} &I(A : C \mid D) = I(A : D \mid C) = I(B : C \mid D) = I(B : D \mid C) = 0 \\ \Rightarrow\ &I(C : D) \leq I(C : D \mid A) + I(C : D \mid B) + I(A : B). \end{aligned}$$

[Andrej Romashchenko]

(4.2) Is the fifth conditional Ingleton inequality of [14] essentially conditional? Is it valid for almost-entropic points? [Milan Studený]

(4.3) By [8, Theorem 7.1] the validity of a max-linear information inequality is equivalent to the validity of an unconditional linear information inequality with existentially quantified non-negative coefficients. Can these coefficients always be chosen rational? If yes, this would imply Turing-equivalence of the problems. [Dan Suciu]

(4.4) Are the cones of linear rank inequalities for $n \geq 6$ polyhedral? [Alexander Shen]

(4.5) Are the inequalities for Shannon entropies valid for *prefix complexity* with precision $\mathcal{O}(1)$? This is known with $\mathcal{O}(\log n)$ precision for Kolmogorov complexity; cf. [13, Chapter 10]. [Alexander Shen]

The Gaussian differential entropy region Υ_n^* is (up to a scaling and an additive term) made of all vectors $(\log\det\Sigma_K : K \subseteq [n])$ with Σ a positive definite $n \times n$ matrix and Σ_K the diagonal submatrix with rows and columns indexed by K. A rational point in $(\Upsilon_n^*)^\vee$ corresponds to a *determinantal inequality* for positive definite matrices under the logarithm and hence its validity is decidable in the existential theory of the reals. A study of the relation between $(\Gamma_n^*)^\vee$ and $(\Upsilon_n^*)^\vee$ was initiated in [2]. The convex conic closure of Υ_n^* is contained in that of Γ_n^* after adding certain functions ϕ_i to the latter, which are defined by

$$\phi_i(I) \coloneqq \begin{cases} -1 & \text{if } i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

The functions ϕ_i have to be added because points in Υ_n^* can have negative entries. What if we instead consider the multiinformation functions of discrete and Gaussian random vectors? They are non-negative, increasing and supermodular. The multiinformation regions are linear images of the entropy regions and correspond to their tight parts.

(4.6) Is the multiinformation region of Gaussians contained in the one for discrete random variables (without ϕ_i's)? This would give a decidable subregion of the cone of linear information inequalities via determinantal inequalities for positive definite matrices. [Tobias Boege]

Let $\mathcal{C}$ be a set of finite groups closed under cartesian product and subgroups. Then given any $G \in \mathcal{C}$ and subgroups H_i, $i \in [n]$, define the vector $(\log[G : \bigcap_{i\in I} H_i] : I \subseteq [n])$, which is the entropy profile of a family of uniform distributions on cosets of $H_I = \bigcap_{i\in I} H_i$ in G. Let $\Gamma_n^*(\mathcal{C})$ be the region of all such vectors. The euclidean closure of each such region is a convex cone under the assumptions on $\mathcal{C}$. Note that the region for abelian groups satisfies the Ingleton inequality. The set of all finite groups yields the Shannon entropy region by [3].

(4.7) Study $\Gamma_n^*(\mathcal{C})$ for classes of finite groups with more structure theory, such as vector spaces over $\mathbb{F}_p$, abelian groups, nilpotent groups, solvable groups, ... Are the inclusions between their entropy regions all strict? [Rostislav Matveev]

Matroids in information theory

(5.1) By [9] it is undecidable if a matroid is entropic. Does this still hold for sparse paving matroids (which are conjectured to be almost all matroids)? [Geva Yashfe]

A $^k/_m$ *matroid approximation* of an integer polymatroid g is a restriction of the free expansion $E(m \cdot g)$ such that each factor X_i of the expansion retains at least a fraction of $^k/_m$ of its elements; see [15, Chapter 10] for the free expansion.

(5.2) If g is entropic, do there exist entropic matroid approximations with $^k/_m$ arbitrarily close to 1? A sufficiently good matroid approximation would violate the Ingleton inequality when g does and hence provide an example of a non-multilinear but entropic matroid. [Geva Yashfe]

Conditional independence

(6.1) How to define a conditional independence relation $\perp\!\!\!\perp$ on a general lattice so that the resulting CI structures fulfill the semigraphoid axioms? [Peter Harremoës]

(6.2) An information inequality of the form $I(X : Y \mid Z) \leq \sum c_i I(A_i : B_i \mid C_i)$ implies the conditional independence inference rule $\bigwedge_i [A_i \perp\!\!\!\perp B_i \mid C_i] \Rightarrow [X \perp\!\!\!\perp Y \mid Z]$. Consider an inference rule and all of its proofs from *Shannon-type* inequalities. What can be said about the size of the coefficients c_i of the "best" such proofs? [Batya Kenig]

(6.3) Is there a finite set of generalized conditional independence inference rules for structural semigraphoids in the sense of [1]? [Janneke Bolt]

Query size estimation

The talk by Hung Ngo introduced a number of bounds for query size estimation in databases. The following questions by Hung concern the relations of these bounds:

(7.1) Is the entropic bound computable? [All by Hung Ngo]

(7.2) What is the complexity of computing the polymatroid bound? Is it NP-hard?

(7.3) If so, what are instances with tractable parametrized complexity?

(7.4) Investigate the tightness of the various bounds, especially the gap between entropic and combinatorial bound.

(7.5) What are the worst-case lengths of proof sequences for Shannon flow inequalities?

Complexity and expressivity

(8.1) Is the validity of linear information inequalities (with rational coefficients), i.e., their containment in $(\Gamma_n^*)^\vee$, decidable? This is open even for $n = 4$. [Many participants]

(8.2) Are there valid inequalities for the linear rank region which cannot be proved by applying valid *entropic* inequalities to subspaces, their sums and intersections (i.e., common informations); cf. [5]? [Alexander Shen]

(8.3) How do linear rank inequalities depend on the field size? How do they depend on the characteristic? [Alexander Shen]

(8.4) The Copy lemma can be used to derive new inequalities from the Shannon inequalities via projection. How does the proof strength of this method increase with dimension or number of copies? [Alexander Shen]

(8.5) Is the entropy maximization principle strictly stronger than the Copy lemma? (The inequalities provable by a single use of the Copy lemma can also be proved by a single use of the MEP. Iterated applications of Copy lemma can be used to simulate a single use of MEP.) [László Csirmaz]

(8.6) Construct an explicit ternary function on a (large) set X that cannot be represented by a circuit of 10 arbitrary binary gates with inputs/ouputs on X. This may be seen as a discrete version of Hilbert's 13^{th} problem. [Alexander Shen]

Combinatorics and Kolmogorov complexity

(9.1) Given an n-bit string s of complexity m. What is the largest complexity obtainable from s by changing at most k bits? [Alexander Shen]

(9.2) Given an n-bit string of complexity m of which every bit flips with probability ε. Which complexity increase can be guaranteed with high probability? The analogue in probability is the increase in entropy of a vector of binary random variables subject to noise. [Alexander Shen]

(9.3) Is there a general procedure for obtaining parallel results in Shannon entropy and Kolmogorov complexity? For a negative result, see [12]. [Alexander Shen]

Let $S \subseteq \mathbb{N}^{\{a,b,c\}}$ be a finite set, $N_I(S)$, for $I \subseteq \{a, b, c\}$, the cardinality of the I-coordinate projection of S. Suppose that $V \cdot \ell \leq N_{ab}(S) \cdot N_{ac}(S)$ for some integers V and ℓ. Then S can be split into S_1 and S_2 such that $N_{abc}(S_1) \leq V$ and $N_a(S_2) \leq \ell$. This is a combinatorial analogue of the non-negativity of conditional mutual information; cf. [13, Section 10.7].

(9.4) Find such analogues for (non-Shannon) information inequalities. [Alexander Shen]

(9.5) How to interpret known *conditional* information inequalities for entropies combinatorially? [Alexander Shen]

References

1. Janneke H. Bolt and Linda C. van der Gaag. Generalized rules of probabilistic independence. In *Symbolic and quantitative approaches to reasoning with uncertainty. 16th European conference, ECSQARU 2021, Prague, Czech Republic, September 21–24, 2021. Proceedings*, pages 590–602. Springer, 2021.
2. Terence Chan, Dongning Guo, and Raymond W. Yeung. Entropy functions and determinant inequalities. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1251–1255, 2012.
3. Terence H. Chan and Raymond W. Yeung. On a relation between information inequalities and group theory. *IEEE Transactions on Information Theory*, 48(7):1992–1995, 2002.
4. László Csirmaz. Visualizing the entropy region. `https://github.com/lcsirmaz/entropy-rules/blob/089f64bb/visual/`.
5. Randall Dougherty, Chris Freiling, and Kenneth Zeger. Linear rank inequalities on five or more variables, 2009.
6. Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-shannon information inequalities in four random variables, 2011.
7. Arley Gomez, Carolina Mejia, and J. Andres Montoya. Defining the almost-entropic regions by algebraic inequalities. *Int. J. Inf. Coding Theory*, 4(1):1–18, 2017.
8. Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Bag query containment and information theory. *ACM Trans. Database Syst.*, 46(3), 2021.
9. Lukas Kühne and Geva Yashfe. On entropic and almost multilinear representability of matroids, 2022.
10. François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation, ISSAC 2014, Kobe, Japan, July 23–25, 2014*, pages 296–303. Association for Computing Machinery (ACM), 2014.
11. František Matúš and László Csirmaz. Entropy region and convolution. *IEEE Trans. Inf. Theory*, 62(11):6007–6018, 2016.
12. Andrej A. Muchnik and Nikolay K. Vereshchagin. Shannon entropy vs. Kolmogorov complexity. In *Computer science – theory and applications. First international computer science symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8–12, 2006. Proceedings.*, pages 281–291. Berlin: Springer, 2006.
13. Alexander Shen, Vladimir A. Uspensky, and Nikolay K. Vereshchagin. *Kolmogorov complexity and algorithmic randomness. Translated from Russian*, volume 220 of *Math. Surv. Monogr.* American Mathematical Society (AMS), 2017.
14. Milan Studený. Conditional independence structures over four discrete random variables revisited: conditional ingleton inequalities. *IEEE Trans. Inf. Theory*, 67(11):7030–7049, 2021.
15. Neil White, editor. *Theory of matroids*, volume 26 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2008.

Participants

- Marcelo Arenas
PUC – Santiago de Chile, CL
- Albert Atserias
UPC Barcelona Tech, ES
- Amos Beimel
Ben Gurion University – Beer Sheva, IL
- Tobias Andreas Boege
MPI für Mathematik in den Naturwissenschaften – Leipzig, DE
- Janneke Bolt
TU Eindhoven, NL
- Laszlo Csirmaz
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Kyle Deeds
University of Washington – Seattle, US
- Oriol Farras
Universitat Rovira i Virgili – Tarragona, ES
- Yuval Filmus
Technion – Haifa, IL
- Emirhan Gürpinar
University of Montpellier, LIRMM – Montpellier, FR
- Miika Hannula
University of Helsinki, FI
- Peter Harremoës
Niels Brock Copenhagen Business College, DK
- Batya Kenig
Technion – Haifa, IL
- Phokion G. Kolaitis
University of California – Santa Cruz, US & IBM Research, US
- Rostislav Matveev
MPI für Mathematik in den Naturwissenschaften – Lepzig, DE
- Fabio Mogavero
University of Naples, IT Hung Ngo, relationalAI – Berkeley, US
- Carles Padró
UPC Barcelona Tech, ES
- Andrei Romashchenko
University of Montpellier & CNRS, LIRMM – Montpellier FR
- Sudeepa Roy
Duke University – Durham, US
- Alexander Shen
University of Montpellier & CNRS – LIRMM, FR
- Milan Studený
The Czech Academy of Sciences – Prague, CZ
- Dan Suciu
University of Washington – Seattle, US
- John MacLaren Walsh
Drexel University – Philadelphia, US
- Lele Wang
University of British Columbia – Vancouver, CA
- Geva Yashfe
The Hebrew University of Jerusalem, IL

Remote Participants

- Mahmoud Abo Khamis
RelationalAI – Berkeley, US
- George Konstantinidis
University of Southampton, GB
- Cheuk Ting Li
The Chinese University of Hong Kong, HK
- Frederique Oggier
Nanyang TU – Singapore, SG
- Soren Riis
Queen Mary University of London, GB
- Yufei Tao
The Chinese University of Hong Kong, HK
- Nikolay K. Vereshchagin
NRU Higher School of Economics – Moscow, RU
- Raymond W. Yeung
The Chinese University of Hong Kong, HK

Report from Dagstuhl Seminar 22302

Educational Programming Languages and Systems

Neil Brown[*1], **Mark J. Guzdial**[*2], **Shriram Krishnamurthi**[*3], **and Jens Mönig**[*4]

1 **King's College London, GB.** dagstuhl@twistedsquare.com
2 **University of Michigan – Ann Arbor, US.** mjguz@umich.edu
3 **Brown University – Providence, US.** shriram@gmail.com
4 **SAP SE – Walldorf, DE.** jens@moenig.org

Abstract

Programming languages and environments designed for educating beginners should be very different from those designed for professionals. Languages and environments for professionals are usually packed with complex powerful features, with a focus on productivity and flexibility. In contrast, those designed for beginners have quite different aims: to reduce complexity, surprise, and frustration.

Designing such languages and environments requires a mix of skills. Obviously, some knowledge of programming language issues (semantics and implementation) is essential. But the designer must also take into account human-factors aspects (in the syntax, development environment, error messages, and more), cognitive aspects (in picking features, reducing cognitive load, and staging learning), and educational aspects (making the language match the pedagogy). In short, the design process is a broad and interdisciplinary problem.

In this Dagstuhl Seminar we aimed to bring together attendees with a wide variety of expertise in computer education, programming language design and human-computer interaction. Because of the diverse skills and experiences needed to create effective solutions, we learned from each other about the challenges – and some of the solutions – that each discipline can provide.

Our goal was that attendees could come and tell others about their work and the interesting challenges that they face – and solutions that they have come up with. We aimed to distill lessons from the differing experiences of the attendees, and record the challenges that we jointly face. The seminar allowed attendees to share details of their work with each other, followed by discussions, and finally some plenary sessions to summarize and record this shared knowledge.

Seminar July 24–29, 2022 – http://www.dagstuhl.de/22302
2012 ACM Subject Classification Applied computing → Education; Software and its engineering → Software notations and tools
Keywords and phrases computer science education research, errors, learning progressions, programming environments
Digital Object Identifier 10.4230/DagRep.12.07.205

1 Summary

Mark Guzdial

To the world at large, programming is one of the most visible and valuable aspects of computing. Yet learning to program has been well documented as challenging and a barrier to entry. The way that a computer interprets a program literally and the complex

[*] Editor / Organizer

Educational Programming Languages and Systems, *Dagstuhl Reports*, Vol. 12, Issue 07, pp. 205–236
Editors: Neil Brown, Mark J. Guzdial, Shriram Krishnamurthi, and Jens Mönig
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

interdependencies within programs are surprising to novices, who have a rather limited understanding of things that can go wrong in programs and what can be done to protect against these problems.

Thinking about computing (and specifically programming) education is particularly timely for several reasons:

- Numerous countries, states, and other geographic entities are making a big push to put computing into curricula for all students.
- Computing tools are now making a serious dent into several disciplines, not just traditional sciences (like physics and biology) but also in social sciences (like history and sociology). New hybrid areas like bioinformatics and data science are being created. People in numerous disciplines now would benefit from, and in some cases need to, learn how to program.
- People outside traditional academic structures are being given the power to program. Everything from spreadsheets to home automation systems are providing "scripting" interfaces that people can use to simplify or enrich their lives. At the same time, those who do not adapt to these trends risk being left behind in their jobs.

In response to the challenges that programmers face, computer scientists have created numerous programming languages and systems, such as interactive development environments (IDEs). To a professional, the more the tools, usually, the better. Tools that can compute advanced analyses, whose meaning may take a great deal of training to understand (as just one example, a worst-case execution time analysis to aid in building a real-time system, or a dependent type system that can statically analyze rich program invariants) can be well worth the investment.

Beginners, however, have a very different set of concerns. For instance:

Syntax Basic matters of syntax can be problematic. Beginners can struggle with the notion that the computer requires very precise utterances (something they have not had to deal with before in writing, math, and other disciplines, where a human reader is usually able to deal with ambiguity, and is forgiving). They may also struggle with basic tasks like typing.

Graduated Introduction A typical textbook will introduce concepts slowly and gradually: each chapter, for instance, might introduce one new concept. Our tools, however, do not typically offer this same graduation. Instead, a student put in front of (say) a Java IDE must confront all of Java: a typo could result in an indecipherable error, strange behavior, and so on. This can result in a very confusing, and unfriendly, learning environment.

Errors Computing is relatively unique in confronting learners with a large number of errors. But errors are intimidating to beginners (many of whom worry that they can "break the computer"), and can at any rate be a deflating experience. At the same time, errors can be viewed as a learning opportunity. How do we design errors – and more broadly, system feedback – in such a way that it is constructive, comprehensible, and encouraging?

Accessibility Computing has traditionally had a rather shameful relationship with users who have special needs. For example, the number of blind professional developers is negligible, and in not at all representative of the percentage of blind people in the population. However, spending a few minutes with our programming tools will make clear why this is not surprising at all: so much is oriented towards visual inspection and manipulation. Similarly, our tools are rarely tested against, for instance, the needs and capabilities of learners with cognitive impairments.

While it may take a long time before the entire computing "pipeline" can adjust to such needs, we can still start to make progress in this direction. Furthermore, the community of beginners who do not need much computing sophistication can immediately benefit from advances in this area.

Our seminar was successful in bringing together attendees with expertise in computing education, programming language design, and human-computer interaction. The presentation and discussions explored a wide range of issues, from language and environment design, to teaching methods and assessment issues. In particular, we studied and discussed:

Tools and Languages A wide variety of tools and languages were presented – from block-based to textual, from imperative to functional, to those embedded in games, and to those explicitly designed for programmer's whose native language is not English. Some of the tools were explicitly for learning and teaching of programming, like tutors and ebooks. We discussed tools for understanding program execution, such as debuggers.

Blocks Many of the attendees are exploring block-oriented programming in some way, so we had several sessions focused on novel uses of block-based programming and applying the blocks modality in different kinds of programming languages and paradigms.

Data Increasingly, we recognize that learning programming is not just about the language and the IDE, but also about *data* – what data students use and how data are described and structured. Data can be motivating for students. Carefully selected data sets can play an important role in supporting learning in contexts outside of computing, e.g., about social issues or about scientific phenomena). Data can be complex and messy, which can take more time to explain and "clean."

Learning Issues Our discussion included the cognitive issues when learning programming and the challenges in helping students to transfer knowledge to new contexts. We discussed the strategies to be taught to students to help them succeed at reading, writing, and debugging programs. We saw teaching techniques that are unique to programming, like Parsons problems. We particularly focused on the cognitive tasks of planning and identifying goals, which are critical to student success in programming. We discussed learning trajectories that explain how we might expect student learning to occur.

Process Computer science has been called the study of algorithmic processes, and computer science education research needs to also be concerned with students learning computational processes in the context of other processes. Some of the processes we discussed include learning processes (i.e., what has to happen to make sure that learning is successful and is retained and possibly transferred), classroom processes (i.e., how does programming fit into the classroom context, including how work is evaluated), and student programming processes (i.e., students are learning design, development, and debugging processes, and need scaffolding to help them succeed and develop their processes to be more expert-like).

Practical Issues of Sustainability Building software has always been difficult to fit into most national research infrastructures. Funding agencies are often reticent to pay for software development, let alone maintenance. Maintaining software over time is expensive but is critical to test educational hypotheses in ecologically-valid contexts (i.e., classrooms vs laboratories) and to gain the benefits of novel software implementations in education. A key area of maintenance of educational software is the fit between the curriculum, the teacher, the school context, and the software. Software may need to change as teaching goals and teachers change, which is a new and complex area for software maintenance.

2 Table of Contents

3 Overview of Talks

3.1 OCaml Blockly

Kenichi Asai (Ochanomizu University – Tokyo, JP)

Joint work of Kenichi Asai, Haruka Matsumoto

OCaml Blockly is a block-based environment for OCaml, built on top of Google Blockly. It has the following features:

- It produces no syntax error, once all the holes are filled with blocks.
- It produces no unbound variable error, since the scoping rules of OCaml are built in. Variable blocks can exist only at the legal places.
- It produces no type error, since the typing rules of OCaml are built in.

In a nutshell, OCaml Blockly behaves decently as the language designer expects. It has been used seriously since 2019 in a functional language course for CS-major students as well as an introductory game programming course for non-CS-major students and one-day seminars for high school students.

3.2 OCaml Stepper

Kenichi Asai (Ochanomizu University – Tokyo, JP) and Youyou Cong (Tokyo Institute of Technology, JP)

Joint work of Tsukino Furukawa, Hinano Akiyama, Kenichi Asai, Youyou Cong
Main reference Tsukino Furukawa, Youyou Cong, Kenichi Asai: "Stepping OCaml", in Proc. of the Proceedings Seventh International Workshop on Trends in Functional Programming in Education, TFPIE@TFP 2018, Chalmers University, Gothenburg, Sweden, 14th June 2018, EPTCS, Vol. 295, pp. 17–34, 2018.
URL http://dx.doi.org/10.4204/EPTCS.295.2

OCaml Stepper is an algebraic stepper for OCaml that supports the basic constructs of OCaml as well as exception handling, outputs, and modules. Since it produces each step incrementally, it can step execute non-terminating programs. It is used in a functional language course for CS-major students to help the students understand the behavior of programs, in particular, how recursion works, where an infinite loop occurs, and where programs exhibit unintended behaviors.

3.3 Snap! Tools for Customizing The Environment For Teachers and Learners

Michael Ball (University of California – Berkeley, US)

Joint work of Michael Ball, Jens Mönig, Brian Harvey

Snap! has broad support for making customizations to the programming environment that aid in teaching students. Most of these tools are designed for instructors to build Snap! Projects that they then give to students, rather than something students use themselves. Snap! has

always had the ability to make custom control structures which aid in creating "Domain Specific Languages" that may be easier for students to learn. Recently, we have expanded the suite of tools that allow instructors to create "microworlds" which direct the student's attention to a subset of blocks [1]. Finally, I'll show how we can use metaprogramming in Snap! to write code that dynamically changes the environment, such as adding or hiding blocks. This allows instructors to add "levels" or different modes to projects.

3.4 Columnal

Neil Brown (King's College London, GB)

URL https://www.columnal.xyz/

Data processing is an important use of computers. Spreadsheets have issues with usability and flexibility, while programming languages like R can be confusing for novices and can obscure the data which should be the central focus. Columnal is a tool that on the surface appears like a spreadsheet, but underneath has a pure functional programming model that manipulates tables in a processing pipeline. The transformations are made visible and fully reproducible in a visual system.

3.5 Structured expressions

Neil Brown (King's College London, GB)

Main reference Michael Kölling, Neil C. C. Brown, Amjad Altadmri: "Frame-Based Editing", Journal of Visual Languages and Sentient Systems, Vol. 3, pp. 40–67, KSI Research Inc., 2017.
URL http://www.ksiresearch.org/vlss/journal/VLSS2017/vlss-2017-kolling-brown-altadmri.pdf

Program source code can often be separated into statements and expressions. Statements are well-suited to being represented as draggable items in block-based programming. However, expressions are less well suited to mouse manipulation as blocks. In this talk I showed an alternative approach, from our "Stride" tool, which structures expressions more at the lexing level than the parsing level, and supports keyboard entry while maintaining the structured nature that eliminates many syntax errors.

3.6 Mio: A Block/Text Programming Environment

Youyou Cong (Tokyo Institute of Technology, JP)

Joint work of Youyou Cong, Junya Nose, Hidehiko Masuhara

The program design recipe is a sequence of steps for defining functions. It serves as an excellent guidance for beginning students, but students tend to skip intermediate steps and go straight to coding, either because they think those steps are not important, or because

they do not know what to write at each step. We propose Mio, an environment with built-in support for design-recipe-based programming. Mio provides blocks for steps before coding, and generates feedback on those steps. A preliminary experiment shows that the blocks and feedback can be effective in encouraging students to follow the recipe.

3.7 Ask-Elle and other model-based tutors

Bastiaan Heeren (Open University – Heerlen, NL) and Johan Jeuring (Utrecht University, NL)

Joint work of Bastiaan Heeren, Johan Jeuring, Alex Gerdes, Thomas Binsbergen, Hieke Keuning, Josje Lodder, Niek Mulleners
Main reference Alex Gerdes, Bastiaan Heeren, Johan Jeuring, L. Thomas van Binsbergen: "Ask-Elle: an Adaptable Programming Tutor for Haskell Giving Automated Feedback", Int. J. Artif. Intell. Educ., Vol. 27(1), pp. 65–100, 2017.
URL http://dx.doi.org/10.1007/s40593-015-0080-x

Ask-Elle is a tutor for learning the higher-order, strongly-typed functional programming language Haskell. It supports the stepwise development of Haskell programs by verifying the correctness of incomplete programs, and by providing hints. Programming exercises are added to Ask-Elle by providing a task description for the exercise, one or more model solutions, and properties that a solution should satisfy. The properties and model solutions can be annotated with feedback messages, and the amount of flexibility that is allowed in student solutions can be adjusted. The main contribution of our work is the design of a tutor that combines (1) the incremental development of different solutions in various forms to a programming exercise with (2) automated feedback and (3) teacher-specified programming exercises, solutions, and properties. The main functionality is obtained by means of strategy-based model tracing and property-based testing. We have tested the feasibility of our approach in several experiments, in which we analyse both intermediate and final student solutions to programming exercises. Similar techniques have been used in tutoring systems for code refactoring and for proofs with structural induction.

3.8 Towards Giving Timely Feedback to Novice Programmers

Johan Jeuring (Utrecht University, NL)

Joint work of Johan Jeuring, Hieke Keuning, Samiha Marwan, Dennis Bouvier, Cruz Izu, Natalie Kiesler, Teemu Lehtinen, Dominic Lohr, Andrew Petersen, Sami Sarsa

Every year, millions of students learn how to write programs. Learning activities for beginners almost always include programming tasks that require a student to write a program to solve a particular problem. When learning how to solve such a task, many students need feedback on their previous actions, and hints on how to proceed. For tasks such as programming, which are most often solved step by step, the feedback should take the steps a student has taken towards implementing a solution into account, and the hints should help a student to complete or improve a possibly partial solution. Research on feedback addresses questions about when feedback should be given, why, and how. This talk discusses how this research

is translated to when, why and how to give feedback on, or a hint at, a particular step a student takes when solving a programming task. We select datasets consisting of sequences of steps students take when working towards a solution to a programming problem, and annotate these datasets at those places at which we think an expert should intervene, why the expert intervenes, and how the expert wants to intervene. We use these datasets to compare feedback and hints given by learning environments for programming to the hints and feedback in the annotated datasets.

3.9 TigerJython and Its Debugger

Tobias Kohn (Utrecht University, NL)

Joint work of Tobias Kohn, Bill Manaris, Aegidius Plüss, Jarka Arnold
Main reference Tobias Kohn, Bill Z. Manaris: "Tell Me What's Wrong: A Python IDE with Error Messages", in Proc. of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE 2020, Portland, OR, USA, March 11-14, 2020, pp. 1054–1060, ACM, 2020.
URL http://dx.doi.org/10.1145/3328778.3366920

Development environments mediate between the programmer and the machine. Aimed at novice programmers, TigerJython is a Python environment that seeks to improve, in particular, the programmer's interaction with the executing interpreter. It separates I/O more clearly from program code, displays enhanced error messages and provides a debugger for program visualisation, but also deliberately refrains from more complex features such as project management. A key insight for the design was to base the various features on a consistent and novice-friendly mental model of the machine.

3.10 The Materials and Media of Programming

Shriram Krishnamurthi (Brown University – Providence, US)

Joint work of Shriram Krishnamurthi, Elijah Rivera, Rob Goldstone, John B. Clements, Greg Cooper, Kathi Fisler, Justin Pombrio

What materials and media can we reach for to practice and learn programming? Many present themselves to our attention. Over multiple talks, we sampled several of these:

- Program planning using a mix of code and text (thanks, Snap*!*)
- Multiple notional machines based on stacks and trees
- New models of reactive programming, especially ones that attend to a diversity of educational and intellectual needs beyond just the production of output
- The exploration of variation in semantics through bottom-up discovery followed by top-down theory formation

3.11 Cognitive skills and learning to program – are meta-analyses of transfer effect useful?

Eva Marinus (Pädagogische Hochschule Schwyz, CH)

There have been past (Liao & Brigth, 1991) and more recent (Scherer et al., 2019) attempts to conduct meta-analyses on the transfer effects of learning to program on the acquisition of near (e.g., programming, programming conceptions) and far cognitive skills (e.g., mathematics, reasoning, creativity). These meta-analyses have reported overall transfer effect sizes of 0.41 and 0.49 respectively. However, when taking a closer look at the results, it becomes clear that it is premature to draw such conclusions as many studies on which the meta-analyses are based suffer from methodological problems, such as small sample size, the absence of an alternative treatment control group, and the absence of a pretest before the intervention. In addition, a couple of unexpected findings are reported, such as very high effect sizes (>2.0 or even >3.0), no effect of treatment duration and the fact that the largest effect sizes were found for transfer to creative thinking. It is concluded that future meta-analyses should first evaluate the quality of the studies on which the analyses are based and consider if there are enough appropriate studies to conduct the meta-analysis on.

3.12 Research Potpourri

Janet Siegmund (TU Chemnitz, DE)

Joint work of Janet Siegmund, Norman Peitek, Sven Apel, André Brechmann

Combining eye tracking and fMRI measures to disentangle the effects of single pieces of code on comprehension showed that descriptive identifiers activate natural-language areas more than nonsense identifiers. Kids in elementary school can already learn basic structures of programming with an offline Scratch Junior version. Mental health in academia is a huge problem and needs more attention. Can we improve on that to a Star-Trek like future?

3.13 Conceptual transfer in students learning new programming languages

Ethel Tshukudu (University of Botswana – Gaborone, BW)

Joint work of Ethel Tshukudu, Quintin Cutts
Main reference Ethel Tshukudu, Quintin I. Cutts: "Understanding Conceptual Transfer for Students Learning New Programming Languages", in Proc. of the ICER 2020: International Computing Education Research Conference, Virtual Event, New Zealand, August 10-12, 2020, pp. 227–237, ACM, 2020.
URL http://dx.doi.org/10.1145/3372782.3406270

The research investigates how students transfer conceptual knowledge between programming languages (Python and Java) during code comprehension. A Model of Programming Language Transfer for relative novice programmers that is based on code comprehension was developed.

Through validating the model, the research concludes that similarities between programming languages play a significant role in semantic and conceptual transfer between programming languages. This research also shows how the model was used to shape the design of a transfer pedagogy in the classroom. It revealed how the pedagogy can lead to improved conceptual transfer and understanding.

4 Additional Talk Abstracts

4.1 Ebooks

Barbara Ericson (University of Michigan – Ann Arbor, US)

Interactive ebooks are increasingly replacing traditional textbooks. There is evidence that they improve student satisfaction and learning and can be used to identify struggling students. There are commercial ebook platforms like Zybooks and several free and open-source ebook platforms including Runestone, Open DSA, and CS Circles. Runestone has over 30 ebooks for computing and math. These ebooks can contain text, images, videos, editable and runnable code, and many other types of practice problems including: Parsons problems, clickable code, and fill-in-the-blank problems. Runestone also includes a spaced practice tool. Instructors can create a custom course, create assignments, create timed exams, automatically grade assignments, and visualize student progress.

4.2 Parsons Problems

Barbara Ericson (University of Michigan – Ann Arbor, US)

Parsons problems are fragments of mixed-up code that have to be placed in the correct order. They were originally designed to provide engaging practice that constrains the logic, allow common errors, model good code, and provide immediate feedback. I have conducted research that provides evidence that Parsons problems have lower cognitive load and are significantly faster to solve than writing the equivalent code or fixing the equivalent code as long as the solution to the Parsons problem matches the most common student written solution. Most teachers and undergraduate students find them useful for learning programming, but some would rather write the equivalent code. We added the ability to toggle from a Parsons problem to the equivalent write code problem and are also testing using Parsons problems to help students who struggle while writing code. I invented two types of adaptation for Parsons problems which change the difficulty of a problem based on the learner's performance. Learners are nearly twice as likely to correctly solve adaptive Parsons problems than non-adaptive ones. Parsons problems can also be used on exams. Scores on Parsons problems highly correlate with scores on write code problems and there is evidence that they are more sensitive to student learning.

4.3 Peer Instruction

Barbara Ericson (University of Michigan – Ann Arbor, US)

Peer Instruction is a pedagogical technique to improve learning in lecture through active and collaborative learning. In Peer Instruction the instructor may assign pre-reading with an assessment before or at the beginning of lecture and then during lecture stop every 10-15 minutes to display a hard question. Students answer individually, discuss their answer with peers, and then answer again. Instructors lead a discussion about the question. Peer Instruction has led to twice the learning gains over traditional lecture in Physics and has increased student engagement and understanding in many fields including computer science. We have been adding support for Peer Instruction to the Runestone ebook platform to make it easier to run Peer Instruction sessions, find peer instruction questions, test research questions, and improve peer instruction questions over time. The tool supports synchronous in-person and synchronous remote users (via a chat interface) and asynchronous users via a saved chat.

4.4 Task plans and implementation choices

Kathi Fisler (Brown University – Providence, US)

Programming problems generally admit multiple solutions, which can differ in choices of data structures, control structures, or the orders in which lower-level steps are performed. Research in other settings suggests that people develop more flexible design skills from seeing multiple solutions to the same problem. We suggest that subtasks can be a useful mechanism for helping students contrast different solutions to the same problem. We demonstrate this with problems that involve compositions of multiple list-manipulation tasks. We show how different mappings from each task to either being computed in a separate function or being tracked in a variable summarizes high-level differences in solution structures. These mappings highlight how different implementation choices lead to different solutions within the same task structure.

4.5 HOFs as abstractions over code and examples

Kathi Fisler (Brown University – Providence, US)

A higher-order function can take a function as an input parameter to another function. Common higher-order functions include filter, map, and sort (taking the sort criterion as an input). How do we teach students what HOFs are and how they come about? One approach is to show students two functions that perform the same core action (such as filtering), then abstract over the common code to define the filter function. We have been exploring a

different approach based on teaching students the features of different higher-order function. We present students with a set of input/output pairs and ask them to classify or cluster the pairs into ones that could be produced by the same function. This is designed to help students recognize the features of higher-order list functions (such as relative lengths of input/output elements, relative types of inputs/outputs, whether all input elements are preserved). We hypothesize that learning these features might help students better understand what different higher-order functions do and when they might apply.

4.6 2D tables for introductory programming

Kathi Fisler (Brown University – Providence, US)

What's the first data structure that we show students? Many courses use arrays, lists, or classes (as tuples of fields). We propose that 2D tables, akin to spreadsheets or CSV files, are a compelling first data structure for teaching computer science (whether to majors or non-majors). Tables are authentic and pervasive. Tables lend themselves to data from a wide range of domains, some of which raise challenges around socially-responsible computing. Programming languages that support tables provide operations such as sorting, filtering, and computing; these can be taught before showing manual iteration or recursion. This choice thus lets us focus on practical tasks and issues from early in a course. We hypothesize that this will help reach populations of students who care less about programming for programming's sake, but without sacrificing depth of computing content.

4.7 Learning Strategies

Diana Franklin (University of Chicago, US)

Joint work of Diana Franklin, Cathy Thomas, Jean Salac, and others

When languages and tools are developed, there is an assumption of the process that learners go through to complete tasks. However, there is great variation in this process, and there are more and less effective processes. Therefore, it is imperative that we identify what those processes / steps are and develop strategies and scaffolds for those strategies so that more students can be successful in CT. For example, our strategy, TIPP&SEE, is useful for learning from example code. We saw statistically significantly stronger performance in students at academic risk in the treatment group, and their performance was comparable to students not at academic risk in the control group. The variance in TIPP&SEE classrooms was significantly smaller than the variance in control classrooms. There are many areas still ripe for benefiting from strategies, such as planning, decomposition, and debugging.

4.8 Learning Trajectories

Diana Franklin (University of Chicago, US)

Joint work of led by Katie Rich, with Carla Strickland and others

Learning trajectories are possible intermediate learning goals, dependencies between them, and activities that can build knowledge from one goal to another. The creation of the learning goals themselves and their dependencies is influenced by pedagogical approaches and education theory, such as constructivism, pieces of knowledge, and spiral conceptual ordering. Our learning trajectories start with everyday understandings of CS concepts, proceed to computational thinking ideas that apply to computation, and end with coding-specific goals.

4.9 Concept-Annotated Examples for Library Comparison

Elena Glassman (Harvard University – Allston, US)

Joint work of Litao Yan, Miryung Kim, Bjoern Hartmann, Tianyi Zhang

Programmers often rely on online resources – such as code examples, documentations, blogs, and Q&A forums – to compare similar libraries and select the one most suitable for their own tasks and contexts. However, this comparison task is often done in an ad-hoc manner, which may result in suboptimal choices. Inspired by Analogical Learning and Variation Theory, we hypothesize that rendering many concept-annotated code examples from different libraries side-by-side can help programmers (1) develop a more comprehensive understanding of the libraries' similarities and distinctions and (2) make more robust, appropriate library selections. We designed a novel interactive interface, ParaLib, and used it as a technical probe to explore to what extent many side-by-side concepted-annotated examples can facilitate the library comparison and selection process. A within-subjects user study with 20 programmers shows that when using ParaLib, participants made more consistent, suitable library selections and provided more comprehensive summaries of libraries' similarities and differences.

4.10 Three Examples of Participatory Design

Mark Guzdial University of Michigan – Ann Arbor, US)

Participatory design is a process of asking those using (or participating in) a program or piece of software to help with the design task. We are challenged to figure out how to set the stage so that participants and leaders have a shared understanding of what is to be designed, and then to have participation from the range of possible participants. I describe three examples of where I have used participatory design: In re-designing an on-line MS in CS program to involve more women, to design a data visualization tool for social studies teachers, and to design new computing education courses.

4.11 HyperCard

Mark Guzdial University of Michigan – Ann Arbor, US)

Apple introduced HyperCard with every Macintosh sold in 1987. At one point, it was likely the most used end-user programming environment in the world. It is unusual for its wordy, phrase-based grammar, and a tight integration with a set of GUI-building tools.

4.12 Teaspoon Languages

Mark Guzdial University of Michigan – Ann Arbor, US)

A Teaspoon language is a task-specific programming (TSP) language with three characteristics: It is a specification of a process for a computational agent (i.e., it really is programming), it can be used for a task that a teacher finds useful, and is learnable by a teacher in less than 10 minutes. These requirements make them very simple, and inexpensive to use in a class (e.g., the whole process of learning and using can fit into a single one hour course period). I describe three example Teaspoon languages for secondary school classes in mathematics, social studies, and engineering. Teaspoon languages may be a way to develop early skills and self-efficacy in computing.

4.13 Hedy

Felienne Hermans (Leiden University, NL)

Teaching block-based languages is popular at the elementary school age, but from the middle school age (10 to 14) kids tend to want to learn textual languages, but Python is still quite tricky for them because its syntax can cause high cognitive load. Hedy aims to make the path to using Python easier, with a gradual approach, using different language levels, so learners do not have to learn all syntax rules at once. In level 1, there is hardly any syntax at all, for example, printing is done with: `print Hello Dagstuhl!`

In every level, new syntax and concepts are added, until kids are doing a subset of Python in level 18 with conditions, loops, variables and lists. In addition to the cognitive load of syntax, in this talk we also discuss other ways in which Hedy aims to reduce cognitive load, such as built-in lesson plans aiming at broad applications of programming.

Hedy was launched in 2020 and since its creation 2.5 million Hedy programs have been created by children worldwide. Try Hedy at `www.hedy.org`.

4.14 Localizing programming languages

Felienne Hermans (Leiden University, NL)

Most programming languages are built with English keywords, and often leaning on the assumption that users will only use latin characters. However, for non-English speakers, especially for non-Latin or right to left languages, using English programming languages can be a large barrier. Hedy (`www.hedy.org`) allows users to program in any natural language, for which a number of technical challenges needed to be addressed, including (but not limited to) allowing non-Latin variable names, calculating using non-Latin numerals, right-to-left parsing and rendering and using multilingual grammars. Open challenges remain, such as production rules that deviate from the traditional English structure, f.e. In Turkish, one would write the condition before a keyword (i==0 iken) rather than before it (while i==0).

4.15 Higher Order Functions in Snap!

Jadga Hügle (SAP SE – Walldorf, DE)

Snap! is a programming language which supports the use of the Higher Order Functions MAP, KEEP (filter) and COMBINE (fold) and allows users to build their own HOFs.

Since Snap! is a blocks-based programming language, Higher Order Functions as well as lambdas are represented visually. They are ovally shaped like any other function but all contain a grey ring – the lambda – as one of their input slots.

The plain rings are part of Snap!'s Operators category and turn functions into data. To get to the return value again, they can be used as an input to the CALL block.

Functions in the rings have implicit parameters meaning all the input slots are left empty. When running the function, all empty input slots in the lambda function will be filled with the current data. In this example, the numbers from 1 to 10 are filled into both empty input slots in the + function.

```
map ( _ + _ ) over (numbers from 1 to 10)
```

In this talk, we showed how we represent lambda with the grey rings, how to recursively build MAP and then apply it to different forms of data like a table of Titanic passengers to find out the percentage of survivors per class or samples of a recording to create an Echo effect.

4.16 PLTutor

Amy J. Ko (University of Washington – Seattle, US)

Learning programming languages is hard, partly because the semantic rules that govern how programs execute in a language are often invisible. We present PLTutor, an approach to providing granular visualizations of these rules, helping learners infer them through causal

inference through a series of examples through forward and backward stepping through a program visualization, coupled with direct instruction about the rules. We evaluated learning gains among self-selected CS1 students using a block randomized lab study comparing PLTutor with Codecademy, a writing tutorial. In our small study, we find some evidence of improved learning gains on the SCS1, with average learning gains of PLTutor 60% higher than Codecademy (gain of 3.89 vs. 2.42 out of 27 questions). These gains strongly predicted midterms (R2=.64) only for PLTutor participants, whose grades showed less variation and no failures.

4.17 Ten Years of Teaching the World to Code with Gidget

Michael J. Lee (NJIT – Newark, US)

Teaching the world to code at scale is hard, in part due to the lack of access to learning materials and teachers. This talk describes Gidget – a free, online puzzle game designed to teach programming concepts by debugging faulty code – and major results from a selection of studies published over the past decade. In Gidget, learners help the eponymous robotic character complete its missions by modifying existing code that is nearly correct. The game also allows learners to run their code step-by-step to highlight how things change within the system at different levels of granularity. This encourages learners to inspect code structure and syntax at their own pace in a friendly environment. Additional features – such as frustration detection hints, embedded formative assessments, avoiding terms of violence (e.g., "remove" instead of "destroy"), choosing pro-social game motives (e.g., cleaning up a chemical spill), and automatically generated levels – have shown to be effective in greatly increasing novice programmers' self-efficacy and attitudes towards coding, and that it leads to measurable learning gains of introductory programming (CS1) concepts.

4.18 Block-Oriented Programming

Jens Mönig (SAP SE – Walldorf, DE)

Following an unwritten rule of sorts, that if something is important for a programming language it becomes (or should turn into) a first-class citizen *of* the language, programming language paradigms are often named after the "first-class citizens" they support, i.e. concepts that can be accessed as data. As "Object-Oriented" has first-class bundles of data-structured and behavior, "Functional" has reified and anonymous functions, an as of yet to be discovered "Block-Oriented" programming paradigm would feature blocks (syntax elements) as first-class citizens of the language. Snap! Now experimentally supports blocks that programmatically manipulate other blocks and can modify stacks of blocks at runtime.

5 Breakout Groups

5.1 Supporting Planning

Summary of questions we're trying to discuss / answer:

- How do we support the planning process? You get a complex problem statement. **How can you structure that thinking process before starting to code?**
- **How to teach the importance of planning?** Let them peer-review each other's designs/ programs/ documentation?
- **What parts of the verbal question gives you cues as to what data type, what control structure, etc?**
- **How do we generate subgoal labels that people would agree on?** Would it be easier to agree on functional programming labels?
- **What are the scaffolds for the different parts of the process?** Problem understanding phase. Algorithmic thinking to solve the problems. Implementation.
- **What are the phases of meta-cognitive processes: pre-, during, post (evaluating)? Can these processes be measured?**
- **How can we relate the theories we're talking about in this space to theories in HCI?**
- **How do we relate planning for program design / function design to planning for experience design or UI design or design thinking?**
- **How do we make planning a more valuable part of the planning process?** Bidirectional with coding. Planning becomes much more valuable in group projects because there, you need to identify the tasks and split up the group equitably. If you teach planning processes as part of the debugging process, then they hit the wall and then do it.
- **How can we create educational experiences so that students see the value of planning / documentation?** A long-term course in which students need to keep reusing their code. At some point in the course, you inherit someone else's code and documentation.
- **Can we extract a plan from examples that are close and modify them?** What can you extract out of your plan to find code that does pieces or something close. Kelleher has worked on creating examples, help students figure out what example is relevant, and help them modify it.
- **What tool features should we have to support these educational interventions and make it so much easier to complete if you do the planning features we've provided?**

  Potential features:
  - Define input and output pairs, provide feedback on their accuracy
  - Collaboration support – use planning as a communication mechanism for groups.

    Input / output types (and comments)

    Input / output pairs (for tasks and subtasks)

    Three categories: Complete, Incomplete, Incorrect
  - Scaffolding:
    - If they get it wrong how do we scaffold them?
    - highlight text description, model how to translate from text description to input/output and types, what are the common mistakes in that process (given the input give them the output)
    - Give them simpler inputs and have them give outputs

  - Understanding what is/ went wrong (and led them self-correct) vs. leading them to the solution.
  - Monitoring-scaffolds asking if what they are doing still aligns with their goals
  - Interface that allows them to go back and forth between planning (design), execution (code) and evaluation (testing & debugging?) phases so that they can adapt the planning if need be.
  - Detect plans from code and tests and give them a hint (e.g., that there is a misfit?) or can we detect that parts haven't been debugged/ tested?
  - Specify solutions (subtasks) and tests upfront and check if they appear in the student environment?
  - Challenge of matching tasks and solutions – how to cope with different solutions – would ML help here?
  - Monitoring if they got stuck: Finding/Distinguishing sink states (i.e., where did they get stuck) -
  - How to scaffold the design of test cases – especially the boundary conditions, error conditions,
  - Properties of valid / invalid inputs
  - Properties of valid / invalid outputs

  Perhaps other criteria – like complexity or run time – compare with optimal solution
  Task breakdown and division of labor (enter student names, then they divide tasks here)
  Input / output pairs for the subtasks
  Define function names and comments that all can see
  If you have a bug, could we have a debugger that steps through only a subset of the code – only the code that is relevant to a specific subset of the code. Can we automatically identify which is the relevant or the irrelevant code? Then we could grey out working code and highlight code that has not been verified correct yet.
  Can we ask them what the intermediate values are?
  When are they wrong about the values?
  When is the code wrong about the values?
  "Stepper 2.0", that skips the stuff that does not need to be checked :)
- Strategies
  What parts of the verbal question gives you cues as to what data type, what control structure, etc? What are the common mistakes?

5.2 Interaction between Learning Content/Curriculum with Tool-PL

Participants: **Mike Lee, Kenichi Asai, Felienne Hermanns, Jadga Huegle, Neil Brown, Jens Mönig, Mark Guzdial**

5.2.1 Re-designing the tools for the curriculum

- Should the tools help to make clearer how the program executes?
- What are the projects that we're asking students to do, and how do we structure the tool to support them?
- Do we start with the PL and then build the learning for the PL, or do we build the tasks first and build the PL to support the tasks?
- Snap (more free-form), Hedy (building for imagined tasks), Teaspoon languages (building the PL for the specific tasks)

- Building for specific purposes allows us to optimize the UI. BlueJ vs IntelliJ. How much do we have to be authentic/professional so that students buy-in?
- You don't want a game to look too professional. It has to be fun and playful. Learning is an afterthought. Does it make transfer harder later? Students seem to make the leap to Python easily. They're not surprised that the UI changes.
- Self-efficacy: They're confident in Gidget, so they think they can learn programming.
- In Germany, it's odd that you'd use a pedagogical version of anything.
- Explicitly, they're trying to make Snap into a presentation tool.

- Projects of projects, to do scenes separately and compose them.
- Make this about assembling things you like to do, not being about (serious voice) "PROGRAMMING"!

- When we teach Teaspoon languages, we say explicitly "This is programming" and "This is not." Works and not-works as you might predict. Teachers value having an explicit process that can be inspected and changed. Teachers are also scared about error messages and not being able to figure out what's wrong.
- Research Question (RQ): Can we bridge direct-manipulation and programming interfaces?

- RQ: Could the GUI tool explain itself in code with good abstraction, good semantics? Not just "click at 10,56" but grabbing an edge of an object, or clicking a button, or part of achieving a subgoal?

- Chaos game as a way to get emergent behavior. Start at a random place, pick another place, put a dot at a point halfway.

- Algorithm so simple.
- Goal: Make the programming simple to tell these kinds of simple but powerful stories.
- Can teachers do this? Education professor uses it. Works well with teacher workshop.
- Having fewer commands/blocks means fewer things will go wrong.
- Every block in Snap! Is a Teaspoon language.

- RQ: Need to collect these stories – how are the PLs changing to support specific learning scenarios?

5.2.2 Supporting Teachers and Classrooms

- Alphonse the Camel – doesn't really work in a computational timeline. Maybe some things just work better in their existing form.
- Programming tasks need to work at a level that is comfortable for students in a classroom discourse. Computational language needs to be at the right level of abstraction.

- How do we represent and talk about powerful ideas like emergence?

- How do we support teachers in doing interesting and innovative work? How do we give them the confidence to make changes, to explore, and to feel that they can succeed?
- As a teacher, you're more confident if you know ALL the surface features, all the available blocks.
- What friction do you have to pay to get started, to learn it?
- How much up-front costs are there?
- Trade-offs between limiting the space to be known and what's possible.

- Hedy reduces what is to be known, but then you can't build so much.

5.2.3 Building Curriculum

- About the stories more than the concepts.
- Western culture and educational systems value the abstract concepts over the concrete instantiations.
- We want to "Storm heaven" – change everything.
- "But does this help me pay my bills?"
- "Tyranny of the Status Quo" – trying to change everything means that someone is going to lose.
- Decompose the stories into the parts that the computer can/should do and what we should do in natural language.
- What are the possibilities for what we can do in the computer?

5.2.4 Maintaining Content and PL connections

- Keeping BlueJ updated with Java – invalidates teacher's materials.
- Building unit tests from the examples that are useful in Columnal
- It's a good example. Make sure it always work.
- Harder to update curriculum content, e.g., screenshots.
- Teachers can upload their own content into the curriculum at Hedy. Can we class-specific or shared with others.
- Easier for them.
- Less cognitive load.
- AND all examples are automatically indexable and updatable. If Felienne updates semantics, she can automatically refactor their examples and keep them updated.
- We alert them. "Warning: This is the change."
- Creates sense of community. Ownership.
- Mike Lee supports teachers in Gidget customizing and adding levels AFTER they finish the 37 built-in levels.
- Use-driven programming language design.
- Can look at teachers' materials to figure out where they want to go.
- Do teachers want to change order, e.g., variables, conditionals, and looping? Can we support that in the language?
- What if teachers spent more units on given concepts? Does that tell us what concepts are harder than we expected?
- How do you deal with privacy (student identity)? What if they say something vile or evil? How to avoid spam and manipulation of page rank?
- Snap games about shooting about schools
- Hedy tries to moderate in all their different languages.

- Snap forum is a small slice of the community. Totally slanted.

- "How to hack in JavaScript in Snap"
- "Why isn't Jens paying attention to me?"
- It's hard to give all of this the attention the community needs.

- Supporting the Computing At School in UK community on-line.

- Didn't want to turn away industry people, but focused on teachers.
- A couple of people were very negative. Wanted to ban them. It's our forum just kick them off. But you want to have fair application of rules. One person can change a community – you don't see what people don't post because they're scared.
- Felienne has lots of experiences with trolls.

- We foster a culture where being an asshole is how you get attention, how you "lead."
- Teacher culture is not the same. They tend NOT to be loud, to be mean to be a leader.
- "I don't need a professor from Georgia to tell me how to do my job."

- Teachers don't want professors telling them what to do.
- Educators (teachers and leadership) value most heavily local voices.

- CAS was primarily academics, but perceived as being mostly industry-driven.

- In part because Simon Peyton Jones was leading, and he was at Microsoft.
- But was instigated by Eric Schmidt of Google saying "Fix this!"

- In the end, teachers are our main curriculum developers.

5.2.5 How can we foster a culture of plurality of tools and concepts

- Be more tolerant of imperative vs functional, text vs blocks, abstract vs concrete.
- Get past fear that we have to do things one way and that we use up all the other space.

- "Will having more than one thing confuse teachers?"
- In the end, humans are good at filtering out what they don't need/want, and coping with complexity.
- We don't have to find the ONE way.
- It's okay to have competing visions.

- Teachers worry about teaching the wrong thing. They want to know the right thing.
- If you love what you're doing, teach that.

- Dijkstra quote: Basic ruins your brain.

- How do teachers deal with technology rapidly changing? Teachers want to develop depth in a specific tool/PL. But it keeps changing.

- RQ: We need Ethel-like work studying teachers as they transfer knowledge and curriculum and PCK into new versions and new languages. How do they do it?

- Different goals: Breadth vs. depth. Computational practices need depth.
- CS Teachers are hesitant because they don't know the answers.

- Teachers don't want to move to CS because they have to keep updating to keep up.
- In sciences, things change, too. But calculus and other subjects, changes can be about PCK and not underlying content.
- How true is that in CS? Depends on language. Java and JavaScript changes dramatically.
- "Laminating your lesson plans."

- Give teachers a way to stick with what they're teaching, e.g., VMs.

- Python print() vs print.

- Challenge: Making Snap keep looking like Scratch to be comfortable but...

- Changing underneath dramatically.
- Teacher say "There's nothing new here!"

5.3 Program Representation

Participants: **Kathi, Ethel, Michael, Tobias, Amy, Janet**

5.3.1 Synopsis

- There are many questions about why we need representations; are they a means to greater understanding or an essential part of programming and learning?
- There are also many questions about whether automated computational tools for generating representations of program behavior is essential, or just a bias we have as a discipline; is it possible that whiteboarding and sketching skills should get our attention instead?
- There are some ways that representations may benefit from social settings, using communication as a vehicle to generate needed representations as needed.
- Representations may also need to leverage prior knowledge. One example of this is having domain knowledge about the data passing through algorithms, especially personal data. This can promote greater comprehension by leveraging learners' assets.
- There may be some value in trying to name some of the many representations that we've invented and teach people how to use them flexibly to reason about algorithms and program behavior. Maybe they will be compelling to the extent they're situated in relevant domains and personal data.

5.3.2 Notes

Why are we here?

- Amy is overwhelmed by programming language design choices for grade 6-12 creative expression contexts
- Michael is trying to understand challenges with Python Tutor and challenges with a purely functional model that leads multiple executions on one line that isn't well supported by professional stepping tools. Also, Snap doesn't have a lot of tools for observing execution. What should be built into the language?
- Janet is teaching two different classes, one is CS2, one is CS1 for a masters course with students who do not have programming experience. Motivated to restructure CS2 course. Exploring different program representations.

- Tobias is wondering about the relationship between the static program and how it dynamically executes. Students want to modify a program while the debugger is running; how to support them during a debugging process. Just having values alone often does not help with understanding. Comment from Janet: Just the numbers only allows students to implicitly develop a mental model of how program execution works
- Ethel has seen PythonTutor and thought it was helpful but wants to understand reasons for choosing a representation and how representation influences transfer.
- Kathi has been wrestling with the question of why, from a learning perspective, are we creating program representations. Tracing programs isn't necessarily the skill we need students to have; we see tracing as a means to an end of broader programming skills, such as debugging, and thinking about behavior. (Some discussion amongst the group about what student goals we're actually targeting and how much the mechanics of understanding program execution actually helps with higher level understanding of program behavior.)

What is learning programming actually about?

- Ethel: code comprehension and writing skills doesn't have a strong correspondence. How do we avoid them feeling like things are important when it's the thing that's in the code. [`https://dl.acm.org/doi/10.1145/3341525.3387379` - "If They Build It, Will They Understand It? Exploring the Relationship between Student Code and Performance"]
- Kathi's goal is preventing defects. She doesn't care about correct tracing per se, just that they avoid making defects.
- Amy wondered about whether understanding is necessary for building something that works. (How important is understanding? If students have sufficient ability (in different aspects) to produce a correct program, is that enough?)
- Code writing goal in industrial context: Ship product and make money (Amy); or build something without understanding (Michael)
- For some students, there's a goal of just getting something to work, and understanding becomes important when things aren't working.
- Kathi: Aliasing and mutation is the only reason she teaches notional machines
- Tobias: understanding in industry is becoming important because of security; not just debugging.
- Michael: program representations are also helpful at showing logic errors. Sometimes it's helpful to just have a tool that helps make it clear where something went wrong.
- Amy: everyone needs different representations constantly (not all of them, not the deepest ones); representation as set of possible thinking tools with different utility in different context (lots of hand waving because of missing vocabulary); we create missing representations for thinking
- Kathi: What kind of thinking should students be doing?
- Tobias: Where does change in data happen? as important to understanding
- Representations of rainfall, such that input transforms until final result(avg); each is a different representation; Amy: Why do we need something fancy, when the whiteboard on the fly representation works so well? Maybe we should teach students to get good at generating representations on whiteboards?
- Kathi: We try to make students make debugging plans where they state what they expect memory to look like after to guide their debugging expectations.
- Amy: We see students trying to build and use their own abstract representations.
- Ethel: if students bring some understanding of the world with them, perhaps representations are a way of connecting to that prior knowledge.

- Michael: What we can grade and what we can assign points for is a tension; students have learned for 20 years of their life that are useful skills for following incentives. These skills are also often done with more than one person. Having communication involved may be an important part of representation generation skill.
- Amy: Constraints of circumstances for students learning may make it more difficult to support them
- Kathi: a lot of this depends on what kinds of programs students are creating; content can be a way of helping them build on their prior knowledge.
- Janet: metaphors of variables often do not help with programming skill.
- Amy: start with a feed of social media posts in a list; filtering, counting. . . , just like with Kathi's starting with tables as real data; personal data is even better than real data for learning effects, because the meaning of the data is so important (in ML context, in which students build classifiers for their own data)
- Michael: similar ideas in taking lists of text; using people's names and converting them into initials works similarly well. Personalizing data is powerful. Familiarity of data gives context for reasoning.
- Ethel: Aha moments happened in industry when there was context for the abstract skills they were learning.
- Amy: A tool that helps students build representations that they need; what would be the value in construction for students' learning? -> Janet: We may need a study
- Michael: What is the right time for the aha moment? The programs we teach are not set up to provide the context that leads to the aha. But projects outside of class often are.
- Tobias: The problem with the aha moment is that they are only valuable if they are emergent, not forced. You want to feel like you are a "smart cookie".
- Ethel: This may not happen for everyone; some students may not catch it the first time and may need teachers to help build the aha moment for them.
- Michael: Different projects can build these moments differently. We use a Plants vs zombie-like game, and a data-oriented project which click with different students.
- Tobias: Aha-moments are important and may come later, way after the course
- Kathi: What's the baseline we have to get them to do get them to be able to build their own representations? Have students develop an explanation for other students to understand programs/concepts
- Amy: Design and programming have many parallels, students fear using unfamiliar representations; forcing them using pen-and-paper helped reduce that fear; does that work also for different program representations? Can we force students to get familiar with other representations?
- Michael: Even when using other representations, students are glued to code
- Amy: Start with input – output before submitting code and have students work out possible concrete algorithmic choices about how to process them, getting them use to algorithm design through low-fidelity representations.
- Kathi: Numberless word problems; give scenario and talk about; or take out question entirely and just talk about scenario; only then give out question; related to programming: what if we take out code and only talk about a problem?
- Tobias: How do I teach functions? switch between function and formula until they have understood
- Amy/Diane: representations of fractions for months until they really understood it; now they can apply that
- Tobias: Students do not see that 0.03 and 3% means the same

- Amy: in math, well-defined representations exist, but not in CS
- Michael: in CS representations, there are some variations in exactly what they mean; they aren't identical. That sometimes gets in our way.
- Ethel: Who decides what are equivalent representations?
- Kathi: What are the types of representations we want to see?

- Code
- Amy: Data Struct Transformation Comic Strip -- Show the state changes in a temporal form
- Tobias: Flow charts (Control flow)
- Michael: Data flow charts
- Memory-based representations (lists)
- Janet: user perspective (input/output)
- Recursion evaluation trees
- Kathi: All of these are very geeky -- what about the interaction with other parts of society?

- Amy: We should engage with design, but we can't teach all of it. We should do justice to it. Which parts do we pull off?
- Alannah Oleson, Amy J. Ko, Brett Wortzman (2020). On the Role of Design in K-12 Computing Education. ACM Transactions on Computing Education, Article 2. https://doi.org/10.1145/3427594
- Maybe they are only geeky in our conventional presentation, but could be brought in more richly through context/domains/etc

- Tobias: Tim Bell's 10 Principles of Computer Science
- Michael: We have useful tools but we need them to be higher level and to connect better.
- Amy: Maybe we need to use more relatable and relevant data.
- Amy: Need to not lose the meaning of data as we represent and work with it
- Tobias: Math edu: We need to formula problems as "word problems". Students need to then extract the information of the word problem before they solve the problem. We risk introducing additional requirements that students need to complete.
- Amy: But sometimes these requirements are skills we want to build.

5.4 Scaffolding Parts of the Process

Understanding the question
- Planning
  - Program Specification / Clarification
  - Program decomposition
  - Choosing overall model / data structures
- Translation – Data Model/Data Structure selection
- Implementation
- Testing
- Debugging
- Reflection

What would a data structure-inspired notational machine look like?
Highlight changes in data, not the control structure.

What variables should you print out when debugging?

Have a data-driven stepper – watch a subset of variables, and have the stepper highlight when each variable value changes. Updated watch variables done in a very friendly way.

Input-Output-Error does not mean the error is in the code, but maybe also on the conceptual level; can students identify relevant parts of source code? So move into a Parson's problem like setting

How to make reliable subgoal labels/plans that actually help students?

Give an example, and then let students work on a similar example to support their understanding

Strategies of students have limits; and if one strategy leads to a dead end, what to do? Allow them to continue that strategy? Highlight a different strategy?

When students have bugs, they often do not know what the intermediate steps are

Have a debugger that works with a student: Ask after each step of execution the state (e.g., fill in numbers in a linked list).

Students do not break down the problem (the debugging problem, not the coding problem), just focusing on the too big a problem, e.g., failing a test case; they do not do the detective work during debugging

Could we generate a sequence of questions to ask them that are the questions they should ask themselves when they debug?

Barb's grant:

Scaffold the TA-student interaction, so that TA can figure out what the misconceptions are;

Nested for loops vs iterating over two lists at the same time; many misconceptions, e.g., schema retrieval, workings of nested loops, incorrect model of data-program state; break down to first steps of loop?

How to detect student's misconceptions in their plans? Break down of iteration step by step, then go one step back

Soloway said that they understand control structures individually but do not understand how to compose them.

There are different kinds of misconceptions at different programming stages (and you would need different supports)

How do we scaffold going from algorithm to code?

Parson's problems of planning steps with vocabulary, e.g., order, clean, filter, aggregate

What kind of language would work for this process? Diana: data-driven approach to let students implement, then analyze and develop vocabulary that works for students

Experiment/study: Audience: starting students or upper level students? Does that make a difference? Also upper-level students have troubles planning

Ask students before and after implementation about their plans

Paper: Scaffolding design problems using Parson's problems: `https://dl.acm.org/doi/10.1145/3279720.3279746`

Should students implement standard data structures? What is the take away for students, what should they learn? Rather using data structures?

5.5 Sustainability notes

- Whatever the source of funding, it's key to align a project's fundraising strategy with the incentives, constraints, and expectations of the funding climate. For some, that might mean particular kinds of scholarship to justify sustain funding, for others that might mean adoption numbers, and for others still that might demonstrate impact on education

systems. Thinking carefully about these incentives is important, as they can often have novelty biases, rigor biases, and technology biases. Accounting for these biases, and ensuring they don't end up warping the scholarship, requires explicit planning.

- Another consideration is how much funding is restricted toward particular expenses; obviously, having unrestricted funding is the most valuable, as it allows for a project to meet whatever needs come, rather than being constrained. It's also the least common and hardest to obtain.
- There are almost always politics that influence how long money can be held and what it is spent on, whether it's a corporate politics game or an academic or foundation politics game. It's key to have someone who can manage those politics and relationships and ensure that they do not interfere with project goals in problematic ways. That could be a project lead, or a principal investigator, or even a funder or corporate partner who provides cover.
- There was a tangible sense of a continuum from deception to omission when it came to reporting and persuading funders. Everyone agreed that deception is unacceptable, but everyone also agreed that sometimes omission was necessary to amplify the outcomes that a funder might care most about, while hiding other details that might be in alignment with academic or innovation goals, but in tension with funder goals.
- It was quite common for successful projects to have one or two people with semi-permanent positions as the backbone of a project, from a maintenance perspective and from a resource perspective. This might be a corporate job or academic position. But it also means that projects have a single point of failure.
- Backend infrastructure can be a key risk to project sustainability. It requires regular maintenance, cloud costs, and staffing. Committing to back end infrastructure is a significant decision with long term consequences. Using university bandwidth and hosting is one way to avoid these costs, but is often restricted to static hosting, or has costs for university IT. But the quality of university IT service can be highly variable, and can also require some political negotiations to navigate policy restrictions.
- It's important to consider other sources of revenue. Communities can be a source of revenue. User events can have registration fees and that can generate unrestricted funding. But where to store that money can be complicated and impose accounting challenges. Donations can also be a source of revenue. Sometimes people will just give and this can also be a source of funding, especially when requests are targeted towards those with philanthropic capacity. Some talked about offering nearly meaningless premium services that offer almost nothing extra, but allow corporations that are often reluctant to donate to pay for a service. This might not sustain the project, but it can help sustain it. All of this requires a place to keep money, which may or may not be the backbone organization.
- We also talked about other sources of staffing, such as ways of trying to onboard, supervise, and engage students to contribute, for credit or for modest pay. There are all kinds of challenges of doing this, including ensuring they have sufficient expertise, that there is onboarding for them, and that they get feedback through code reviews or other mechanisms. We also talked about open source contributions and the limited value of "drive by" contributors that don't have enough context for the project to make meaningful contributions. Some talked about ways of engaging contributors socially first, to get context, and then have them contribute later after they have it.

5.6 How to Snap

Participants: **Jens, Elena, Michael Lee, Kenichi, Jadga, Mark**

5.6.1 Meta-Discussion

Snap is challenged to fit within two words:

- Some users come in from Scratch and know "sprite" and "costume"
- Some users come in from programming and think "objects" and "bitmaps"

It's hard to make all features visible when the programming is all visible.

- No API documentation
- Have to use drag feedback, like colors.
- Has to be an ecosystem – curriculum, communities, documentation.

- Current Snap manual is SICP-lite.

Remixing in Scratch – big part of the community culture.
New features in Snap support this sharing/remixing in more granular ways

- No real way to link code to people, because it's coming from SAP. There's a danger about a large corporate entity being able to track code and people.

Supporting teachers in sharing, remixing, and ownership in the community is even harder but more important than students.

- Hedy's support for teachers is fairly impressive.

Object model in Snap:

- Sprites and clones use Henry Lieberman style prototypes and delegation of slots. Can inherit something like y-position, so clones are linked vertically.

- Sprites are objects with methods and instance variables, and can send messages and data between each other.
- Can connect sprites so that they are sub-parts of one another.

- Scenes are totally different worlds.

6 Open problems

6.1 What studies should we do together? (and which not) / what collaborations could come out of this week?

Mark Proposed Study: Test transitions between our languages and tools. Pairs of us work on transitions from (for example) from Hedy to Pyret, or Snap to Hedy. (Requesting Ethel to be part of this!)-I love this Mark-Ethel.

This is Felienne: I love this too! Many of our Hedy classrooms do Scratch first! We even already have data of users of whether or not they have used Scratch before! We can totally do data analysis on behavioral differences between Hedy users with and without Scratch experience!

Of course, Mark is interested in ANYBODY transitioning from Teaspoon to ANY PL! I'm so interested to see how we support that transition.

Ethel-i think this is the right time to explore transitions, i would encourage people to participate in this and understand transfer at a deeper level. Also mapping constructs across different programming languages is something I'm looking forward to. The mappings can be put in a public domain later on if possible ->Love this idea!!

Teacher Transitions: Examine the transition/process of teachers' learning and choice of second language. Can I suggest you read Ethel's paper about this :) ->`https://dial.uclouvain.be/pr/boreal/object/boreal:251251`

- Ethel's paper is great <3 In our breakout, we also talked about the challenge of moving curriculum, learning content. Do teachers focus on making their content better, or transitioning to new languages and versions?
- Thanks Mark! I was trying to remember how we phrased it in our breakout. – Mike

Kathi wants to explore program representations/notional machines that center around data transformations, partly to help students create and execute debugging plans. What would such a representation look like? What differences might arise based on the semantics of the programming language? Would a data-centered approach help connect data to plans to programs? Johan: Don't you want more than debugging plans: also the kind of planning you discussed in your talk. Kathi: yes, plans as well. I'm wondering whether data-centered plans would also give debugging guidance, which would again tie to notional machines.

Johan proposed study: (Should probably be combined with others.) Different approaches to how we can support planning. Using Mio, Shriram's, and Kathi's approach to planning, add a testing based approach to it (as discussed in the break-out session this morning), and study how it works. Ben likes this idea. See related idea below. Youyou also likes this idea. Kathi is interested. Diana

Diana Proposed study: Cog Sci & strategies people pair up to better understand the processes involved in different steps of programming (program clarification, planning / decomposition, algorithm development, coding, debugging) and propose new strategies. Eva is keen to join :) Kathi is interested.

Ben: Program annotation, planning, and debugging – If students are writing interleaved (i.e. not very decomposed) programs, can they annotate pieces of their code to indicate which plans/goals pieces of their code are meant to accomplish? A visual representation of this could also be contrasted with a visualization of data flow within their programs (e.g. to indicate that some information (e.g. input from a test case) is not making it into a vital part of computing a correct result). This could be used to build more focused steppers or to help students to figure out where to focus when dealing with incorrect behavior. Johan: can we for example connect annotations/subgoal labels to test cases? Or ask for test cases at these annotations? Can we grey out parts of the program that are not part of a subgoal in a stepper/debugger? +Barb

Bastiaan: nice idea! Kathi is interested.

Diana: How can we provide real-time hints / scaffolding to help students design "good" test cases (black-box tests) for the purposes of both design and testing? +Barb

Kathi is also interested in this, noting that tests characterize a space of inputs/scenarios not just individual scenarios. I see this as a form of sensemaking.

Diana: Can we analyze problem statements and develop a process for identifying keywords or aspects of the problem statement that point towards specific aspects of the design and/or implementation? +Barb: This might also help students write good purpose statements.

Barb: Can we all study the same problem – like Kathi's shopping cart problem or the rainfall problem in a variety of contexts with a variety of scaffolding? +Johan: I think taking a particular example to discuss our ideas will be very fruitful +Tobias +Diana

- MIMN? Multi-institutional, multi-national?
- ITiCSE working group: "Everybody teach X in any way you want. We all use the same measure of learning/understanding at the end."

Kenichi: Programming language teachers in CS departments do not necessarily know how their ways of teaching are effective or not. It would be great if education experts can see the class (although it takes quite a long time) and make advice on how to improve the class and even how to conduct experiments. Conversely, CS people could hear what teachers want to teach and advise what is feasible technically.

Diana: Not quite a study, but it seems like a Computing Research Infrastructure NSF (USA) grant could fund the development of an open-source framework for testing many innovations (in one particular language) so that research could be pushed forward. This could help many people.

Amy proposal: I'd love to see us try 5-10 radically different approaches to educational programming languages than exist now, both to explore the design space more broadly through the lens of cultural responsiveness, but also to understand more deeply the role of syntax, semantics, and representation on learning, teaching, curriculum. (What would the world look like if we had 10 different mature Scratch/Snap etc platforms, but doing things very differently for very different audiences). Anyone want to write a $25 million 5 year grant? Kathi would be interested in brainstorming around or discussing this.

Ben would as well . . . depending upon how radical we can be :-)

References

1 Garcia, Dan and Ball, Michael and Garcia, Yuan. 2022. *Snap! 7 – Microworlds, Scenes, and Extensions!. Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2. ACM: New York, NY USA.*

Participants

- Kenichi Asai
Ochanomizu University – Tokyo, JP
- Michael Ball
University of California – Berkeley, US
- Neil Brown
King's College London, GB
- Youyou Cong
Tokyo Institute of Technology, JP
- Barbara Ericson
University of Michigan – Ann Arbor, US
- Kathi Fisler
Brown University – Providence, US
- Diana Franklin
University of Chicago, US
- Elena Leah Glassman
Harvard University – Allston, US
- Mark J. Guzdial
University of Michigan – Ann Arbor, US
- Bastiaan Heeren
Open University – Heerlen, NL
- Felienne Hermans
Leiden University, NL
- Jadga Hügle
SAP SE – Walldorf, DE
- Johan Jeuring
Utrecht University, NL
- Amy Ko
University of Washington – Seattle, US
- Tobias Kohn
Utrecht University, NL
- Shriram Krishnamurthi
Brown University – Providence, US
- Michael J. Lee
NJIT – Newark, US
- Eva Marinus
Pädagogische Hochschule Schwyz, CH
- Jens Mönig
SAP SE – Walldorf, DE
- R. Benjamin Shapiro
University of Colorado – Boulder, US
- Janet Siegmund
TU Chemnitz, DE
- Ethel Tshukudu
University of Botswana – Gaborone, BW