



Volume 1, Issue 10, October 2011

Forensic Computing (Dagstuhl Seminar 11401) <i>Felix C. Freiling, Dirk Heckmann, Radim Polcák, and Joachim Posegga</i>	1
Computing with Infinite Data: Topological and Logical Foundations (Dagstuhl Seminar 11411) <i>Ulrich Berger, Vasco Brattka, Victor Selivanov, Dieter Spreen, and Hideki Tsuiki</i>	14
Foundations of distributed data management (Dagstuhl Seminar 11421) <i>Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin</i>	37

ISSN 2192-5283

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany.

Online available at <http://www.dagstuhl.de/dagrep>

Publication date

January, 2012

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 <http://dnb.d-nb.de>.

License

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license: CC-BY-NC-ND.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.
- Noncommercial: The work may not be used for commercial purposes.
- No derivation: It is not allowed to alter or transform this work.

The copyright is retained by the corresponding authors.

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.

Editorial Board

- Susanne Albers
- Bernd Becker
- Karsten Berns
- Stephan Diehl
- Hannes Hartenstein
- Frank Leymann
- Stephan Merz
- Bernhard Nebel
- Han La Poutré
- Bernt Schiele
- Nicole Schweikardt
- Raimund Seidel
- Gerhard Weikum
- Reinhard Wilhelm (*Editor-in-Chief*)

Editorial Office

Marc Herbstritt (*Managing Editor*)

Jutka Gasirowski (*Editorial Assistance*)

Thomas Schillo (*Technical Assistance*)

Contact

Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de

Digital Object Identifier: 10.4230/DagRep.1.10.i

www.dagstuhl.de/dagrep

Forensic Computing

Edited by

Felix C. Freiling¹, Dirk Heckmann², Radim Polcák³, and
Joachim Posegga⁴

1 University of Erlangen-Nuremberg, DE, felix.freiling@cs.fau.de

2 University of Passau, DE, heckmann@uni-passau.de

3 Masaryk University, CZ, radim.polcak@law.muni.cz

4 University of Passau, DE, posegga@uni-passau.de

Abstract

Forensic computing (sometimes also called *digital forensics*, *computer forensics* or *IT forensics*) is a branch of forensic science pertaining to digital evidence, i.e., any legal evidence that is processed by digital computer systems or stored on digital storage media. Forensic computing is a new discipline evolving within the intersection of several established research areas such as computer science, computer engineering and law.

Forensic computing is rapidly gaining importance since the amount of crime involving digital systems is steadily increasing. Furthermore, the area is still underdeveloped and poses many technical and legal challenges.

This Dagstuhl seminar brought together researchers and practitioners from computer science and law covering the diverse areas of forensic computing. The goal of the seminar was to further establish forensic computing as a scientific research discipline, to identify the strengths and weaknesses of the research field, and to discuss the foundations of its methodology.

The seminar was jointly organized by Prof. Dr. Felix Freiling (Friedrich-Alexander University Erlangen-Nuremberg, Germany), Prof. Dr. Dirk Heckmann (University of Passau, Germany), Prof. Dr. Radim Polcák (Masaryk University, Czech Republic), Prof. Dr. Joachim Posegga (University of Passau, Germany), and Dr. Roland Vogl (Stanford University, USA). It was attended by 27 participants.

Seminar 03.–07. October, 2011 – www.dagstuhl.de/11401

1998 ACM Subject Classification K.4.1 [Computers and Society] Public Policy Issues

Keywords and phrases forensic teaching, practical experience in forensics and law, selective imaging, mobile phone forensics, cryptographic hash functions

Digital Object Identifier 10.4230/DagRep.1.10.1

Edited in cooperation with Michael Spreitzenbarth

1 Executive Summary

Felix C. Freiling

Dirk Heckmann

Radim Polcák

Joachim Posegga

License  Creative Commons BY-NC-ND 3.0 Unported license

© Felix C. Freiling, Dirk Heckmann, Radim Polcák and Joachim Posegga

After a brief introduction by the organizers, the seminar started off with a sequence of 3 slide/5 minute talks by all participants stating their research interests, their background and their expectations towards the seminar. In the afternoon, two introductory talks by



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Forensic Computing, *Dagstuhl Reports*, Vol. 1, Issue 10, pp. 1–13

Editors: Felix C. Freiling, Dirk Heckmann, Radim Polcák, and Joachim Posegga



Dagstuhl Reports

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

Dieter Gollmann (“Access control — principles and principals”) and Stig Mjolsnes (“ICT and forensic science”) paved the way for a common understanding of the open questions in the area and the relation of forensic computing to computer security.

Wednesday morning commenced with a first introductory law talk by Focke Höhne (“Introduction to German IT Forensics Law”). It was followed by two insightful technical talks from presenters who had considerable practical experience in the area: Glenn Dardick and Kwok Lam.

The afternoon was spent on a pleasant hike to a nearby village where the Dagstuhl office had organized delicious traditional coffee and cake. On the way back to Schloss Dagstuhl a group of adventurers separated from the main party to explore the woods around Wadern. They only managed to return to Dagstuhl in time because of modern navigation technology (paper maps provided by the Dagstuhl office). Reasons for the failure of more traditional technology (iPhones, etc.) were discussed in the evening in the wine cellar.

Thursday saw a mix of legal and technical talks: Herbert Neumann raised many questions during his presentation of practical (law) case studies while Viola Schmid presented a proposal for a “Casebook on Cyber Forensics”. Harald Baier discussed the deficits of forensic hash functions and Felix Freiling shared some of his experiences from teaching digital forensics. After lunch Michael Spreitzenbarth presented an overview over mobile phone forensics while Radim Polčák gave some background on the issues of data retention relevant in different countries. Joshua James pointed out the necessity to overcome the traditional separation of sciences and encouraged more interaction between computer science and law.

Finally, Johannes Stüttgen introduced the method of “Selective Imaging” to improve the digital evidence collection process.

Friday morning hosted a series of three talks from computer science, law and practice. Stefan Kiltz spoke about techniques to seize transient evidence in networks, Sven Schmitt gave an overview of digital forensics at the German federal police (BKA), and Nicolas von zur Mühlen sparked many discussions during his presentation on transborder searches.

Conclusion

Overall, the seminar was well-received by the participants. They particularly liked the interdisciplinary approach, which is documented by the results of the final Dagstuhl survey: Almost all participants stated that the seminar led to “insights from neighboring fields or communities” and that they made “new professional contacts like an invitation to give a talk or to join an existing project or network”.

The organizers also identified room for improvement: Only about one-third of the participants came from law. This points to a fundamental problem for future seminars since — similar to participants from industry — it is rather untypical for academics in law or for international practitioners to spend an entire week at a seminar or workshop.

In possible future seminars, the set of relevant topics should be broadened to include legal aspects of IT forensics in enterprises. This would substantially enlarge the set of interested international academics and further nourish community building which is currently vital to the field.

2 Table of Contents

Executive Summary

<i>Felix C. Freiling, Dirk Heckmann, Radim Polcák and Joachim Posegga</i>	1
---	---

Overview of Talks


Deficiencies of (Cryptographic) Hash Functions in Digital Forensics <i>Harald Baier</i>	4
Cyber Forensics Assurance Model <i>Glenn S. Dardick</i>	4
Experiences from teaching forensic computing <i>Felix C. Freiling</i>	5
Access Control – Principles & Principals <i>Dieter Gollmann</i>	5
A transparent bridge for forensic sound network traffic data acquisition <i>Stefan Kiltz</i>	6
Forensic Computing: Objectives and Challenges <i>Kwok Lam</i>	6
ICT and Forensic Science <i>Stig Frode Mjølunes</i>	7
A small case of practical experience <i>Herbert Neumann</i>	8
Proportional Cybersecurity <i>Radim Polcák</i>	9
Casebook on Cyber Forensics (CCF) – a proposal for discussion <i>Viola Schmid</i>	10
Mobile Phone Forensics with the help of ADEL <i>Michael Spreitzenbarth</i>	11
Selective Imaging <i>Johannes Stuetgen</i>	11
Legal Challenges of transborder Searches <i>Nicolas von zur Muehlen</i>	12

Participants	13
-------------------------------	----

3 Overview of Talks

3.1 Deficiencies of (Cryptographic) Hash Functions in Digital Forensics

Harald Baier (University of Applied Science Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Harald Baier

Joint work of Baier, Harald; Breiting, Frank

Main reference Harald Baier, Frank Breiting, “Security Aspects of Piecewise Hashing in Computer Forensics,” Proc. 6th Int’l Conf. on IT Security Incident Management & IT Forensics (IMF 2011), IEEE CS, pp. 21–36, 2011.

URL https://www.fbi.h-da.de/fileadmin/gruppen/FG-IT-Sicherheit/Publikationen/2011/2011_05_Baier_IMF2011.pdf

Hash functions are well-known methods in computer science to map arbitrary large input to bit strings of a fixed length that serve as unique input identifier/fingerprints. A key property of cryptographic hash functions is that even if only one bit of the input is changed the output behaves pseudo randomly and therefore similar files cannot be identified.

However, in the area of computer forensics it is also necessary to find similar files (e.g. different versions of a file), wherefore we need a similarity preserving hash function also called fuzzy hash function. In this talk we present use cases of cryptographic hash functions and discuss their drawbacks.

We come up with proposed approaches for fuzzy hashing and discuss the next steps.

3.2 Cyber Forensics Assurance Model

Glenn S. Dardick (Longwood University – Virginia, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Glenn S. Dardick


Main reference Based on Model Previously presented in December 2010 at SECAU Conference

As the usage of Cyber Forensics increases, so does the potential for errors in the practice of applying Cyber Forensic. Errors in opinions derived from faulty practices have resulted in grievous miscarriages of justice. However, utilizing the foundations of Information Systems Assurance and Information Quality, a solid foundation for improving the quality and effectiveness of Cyber Forensics can be derived. The foundations of Information Systems Assurance and information Quality provide a solid foundation for improving the current efforts in Cyber Forensics. With increasing computer and network systems usage as well as the increasing frequency of attacks on information systems, the need for controlling risks in information systems have become more apparent. Meeting that need, Information Systems Assurance has continued to evolve: from the CIA (confidentiality, integrity, and availability) into variations such as the five pillars (confidentiality, integrity, availability, authenticity, and non-repudiation) and the Parkerian Hexad (confidentiality, integrity, availability, authenticity, possession, and utility). Also, with the continuing growth of information systems, the need for improving the quality of such systems has also evolved focusing on various components of information Quality (accuracy, relevance, consistency, timeliness and completeness).

Utilizing the foundations of Information Systems Assurance and information Quality a model has been derived for Cyber Forensics Assurance. However, there is still a need to increase the level of training among digital forensics experts in order to attain the assurance needed as defined by the Cyber Forensics Assurance Model.

3.3 Experiences from teaching forensic computing


Felix C. Freiling (Universität Erlangen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Felix C. Freiling

In the summer of 2004, I gave the first lecture on forensic computing to University students in Germany together with Maximillian Dornseif. Since then, the field of IT forensics has changed dramatically and many more Universities have started to teach the subject. This talk reviews the current European teaching landscape in IT forensics and relates past and future developments in the field to my own teaching experiences. Furthermore, the design of the first German Master degree programme in digital forensics is discussed.

3.4 Access Control – Principles & Principals

Dieter Gollmann (TU Hamburg-Harburg, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Dieter Gollmann


The concepts and terminology for access control were developed in the 1970s and 1980s in the context of closed organizations. In the context it was natural that principals (active entities) security policies referred to were closely related to human users, as is evident from the research literature of that time. By the same measure, there was a close link between access control and accountability.

This paradigm is still highly influential on the perception of access control but it is a poor match for today's situation in Web 2.0 applications. In a world of services, the services become principals; principals have to be named and have to be “authenticated” when issuing access requests. For names, the convention of using host names from the Domain Name System (DNS) has been adopted. However, DNS was not designed as a system supporting access control; in particular, there are no inherent mechanisms that stop authoritative name servers from lying about name/IP address bindings. For authentication, traditional PKI-based solutions do not exist, probably will never exist on a global scale, and are arguably not necessary in the first place. Alternatives are “recognizing the same service as before” (P. Nikander: identification) or distinguishing own requests/scripts from requests/scripts forwarded on behalf of others.

In summary, principals are associated with services, not with persons, authentication (determining origin) may be replaced by different notions, and the close link between access control and accountability no longer exists.

3.5 A transparent bridge for forensic sound network traffic data acquisition

Stefan Kiltz (University of Magdeburg, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stefan Kiltz


Joint work of Kiltz, Stefan; Hildebrandt, Mario; Hoppe, Tobias; Dittmann, Jana
Main reference Stefan Kiltz, Mario Hildebrandt, Jana Dittmann, “A transparent bridge for forensic sound network traffic data acquisition,” In Sicherheit 2010 – Sicherheit, Schutz und Zuverlässigkeit, 5. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI) Berlin, 5–7 Oktober 2010. S. 93.

In this paper we introduce a prototype that is designed to produce forensic sound network data recordings using inexpensive hard- and software, the Linux Forensic Transparent Bridge (LFTB). It supports the investigation of the network communication parameters and the investigation of the payload of network data. The basis for the LFTB is a self-developed model of the forensic process which also addresses forensically relevant data types and considerations for the design of forensic software using software engineering techniques. LFTB gathers forensic evidence to support cases such as malfunctioning hard- and software and for investigating malicious activity. In the latter application the stealthy design of the proposed device is beneficial. Experiments as part of a first evaluation show its usability in a support case and a malicious activity scenario.

Effects to latency and throughput were tested and limitations for packet recording analyzed. A live monitoring scheme warning about potential packet loss endangering evidence has been implemented.

3.6 Forensic Computing: Objectives and Challenges

Kwok Lam (National University of Singapore, SG)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Kwok Lam

In this talk, we discuss the relationship of computing and forensics, and the role of computing in forensics. The objectives and challenges of forensic computing are also discussed. Specifically, the nature of digital evidences, where and how digital evidences may be collected for supporting forensic works in different types of scenario in which digital evidences are of crucial legal implications. We'll identify areas where computing techniques may be applied to support forensic activities and propose approaches for future development of methodologies for forensic computing. We conclude by sketching a proposed collaborative model for legal and computing researchers to contribute to the development of forensic computing methodologies.

3.7 ICT and Forensic Science

Stig Frode Mjølunes (NTNU – Trondheim, NO)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Stig Frode Mjølunes

Main reference Stig F. Mjølunes (ed.), “A Multidisciplinary Introduction to Information Security,” Chapman and Hall/CRC, 2011.

The term forensic is derived from the latin *forum*, denoting the public square of roman cities (foremost Forum Romanum), where all public matters including of judicial nature took place.

The proceedings of disputes and public trials were conducted orally, and the actual evidence supporting the claims were physically and methodically presented to a judge positioned on a tribunal.

The judicial evidence could take the form of testimony of witnesses, physical objects, or documents.

Evidence means what is clearly there for all to see.

The roles of computers and networks at the crime scene can be one or more of the following:

- The direct target of intentional incidents (information security).
- Technical tools and accomplices for crime (cybercrime).
- Instruments assisting the incident investigation process (digital forensics).
- Passive sources of evidence, and witnesses providing technical testimonies (digital evidence).

An after-the-fact investigation of an incident seek answers to the questions of what happened, the true explanation of how it happened, and the attribution to who did what. The judicial verdict must be founded on inculpatory and exculpatory evidence in criminal law. The evidence must be relevant and intelligible in the context of the judicial inquiry. Many disciplines of science are employed in this process of technical evidence.

Digital components and systems can be passive sources of information, or even regarded as witnesses that can provide technical testimonies about events. These informational objects are called digital evidence.

The evidence authenticity, both the origin and the integrity, must be assured by proper chain-of-custody/provenance. The current practice of the use of one-way hash function for verifying the integrity cannot become acceptable without some sort of commitment protocol.

The problem of forensic/court presentation of digital evidence is hard.

Can digital evidence be presented directly, or is it only possible to present indirect documentation about the digital evidence? For instance, US Federal Rules of Evidence distinguishes between original and duplicate.

Pragmatically, a paper printout shown to reflect the data accurately is called an “original”. This presupposition of original is not future proof, and new definitions suitable for digital evidence are needed.

The Daubert Test constitutes four criteria for an acceptable forensic theory or method. Currently, there does not exist any theory or method in digital forensics that will satisfy all four soundness criteria. Some of the current analysis methods are image/mirror copy, keyword search, file type search, hash values of known files, timeline analysis using timestamps. The software tools and ad-hoc techniques developed for digital evidence extraction are often very specific to a device or software.

Any digital forensic detection tool will spur, with time, an anti-detection tool.

Are public or secret tools the best in practice? Remember that fingerprint analysis is used although gloves are easily available.

Finally, I list some promising research directions pertaining to digital forensics:

- Time line analysis using temporal logic for (partially) event ordering.
- Reverse engineering techniques of hardware, software, and systems.
- Bayesian causal graphs applied to digital evidence inferences assessing alternative hypotheses.
- Cryptanalysis models and methods from a forensic perspective.
- Shared cross-border investigation and aggregation of technical evidence from internet-based network infrastructures.

This Dagstuhl seminar presentation is based on Chapter 12 in the book *A Multidisciplinary Introduction to Information Security* [1].

References

- 1 Stig F. Mjølsnes (ed.) *A Multidisciplinary Introduction to Information Security*. Chapman and Hall/CRC, 2011. 348 pages. ISBN 978-1420085907

3.8 A small case of practical experience

Herbert Neumann (*Anwaltskanzlei Neumann – Molfsee, DE*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Herbert Neumann

The facts in catchwords: A 56 years old man, civil servant, married, two children (13 and 15) is accused of downloading and possessing child-pornographic photos. The police has searched through his home and confiscated the families entire it-equipment:

- father's PC and 3 laptops
- one external drive for backups
- 37 Original CD's and DVD's
- 9 blank CD's
- 2 photo- and one videocameras
- 29 video tapes
- the NTBA, splitter, DSL-modem and wifi-router

After nine months four child-pornographic photos were found on the PC named: *53896.jpg*, *89463.jpg*, *73346.jpg*, *1397.jpg*. But the defendant says: "I never did such a thing."

The background: The ISP had detected rapidly increasing traffic on 2 domains: *jh-desjn8.khbs23.de*, *jsbggg63.bgsivr5c.de* hosted on his servers.

Content: 500 pornographic photos , 29 thereof clear child-pornographic. Prosecution was informed, the complete communication was monitored and stored.

Result: during one month over 300.000 accesses on the first server, over 92.000 accesses on the second.

Investigation: IP-address – customers name – suspects name. About 12.000 preliminary investigations by public prosecution. Distribution to the local responsible prosecution (i.e. Köln 500).

The prosecutors duty: to investigate not only the inculpatory facts but also the exculpatory ones. (Word-for-word written down in the German Code of Criminal Procedure Art. 160 Ch 2) The attorney has to be the most objective person in the world.

The most important rules of evidence in Criminal Law:

- In doubt in favor of the defendant
- The court has to prove the defendants guilt and not the defendant has to prove his innocence.

Procedure by the public prosecution according to how long the suspects looked at the pictures or downloaded some, the investigations were:

- adjusted instantly (dwell “a few” i.e. 45 seconds only thumbs) or continued (minutes or even downloaded photos)
- search warrants
- prosecution of the sever cases

The Experts duty: The expert is bound to furnish the opinion to the best of his knowledge i.e. to explain about conflicted opinions.

The Courts duty: to exclude all possibility of reasonable doubt otherwise to acquit.


Problem: what are reasonable doubts? at any rate not: the green manikins from Mars

The Court issues the following order: Mr(s) O. is appointed to an expert of Forensic Computing. The expert is assigned to answer the following questions:

- How secure is the investigation of the customer on the basis of the ip-address? resp. is it possible, that an error occurs in the log ? i.e. a wrong ip-address, date or time is stored
- If yes:
 - are there any science-based findings, how often this happens?
 - would an error be noticeable anyway?
- Is it possible to modify the content of a log file?
- If yes, would a manipulation be discoverable anyway?
- Could a site by Firefox be accessed (prefetched) without assistance by the user?
- If yes, is it also automatically downloaded then?
- If yes, would this be discoverable anyway? i.e. stored in ISP’s or users log files or browser-cache
- Could anyhow a photo be stored on the users hdd without his assistance and knowledge?
- If yes, would this be discoverable anyway? i.e. stored in ISP’s or users log files or browser-cache

3.9 Proportional Cybersecurity

Radim Polcák (Masaryk University, CZ)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Radim Polcák

Securing national cyberspace always requires at least marginal infringement of distributive (individual) rights in favor of non-distributive (common) goods. The key issue in this case is to proportionally balance between various constitutionally grounded rights depending on recent state of social and technical development. If a system of national cybersecurity is to be somewhat efficient, it always has to combine gathering information with efficient competences including ultimate ones like blocking. That obviously collides with a set of


individual information rights that the German Constitutional Court originally named as the right to information self-determination as well as with various procedural rights together named as the right for fair trial.

Compared to traditional security issues, there are multiple specific features also in securing of internal information infrastructure of a state and in gathering of respective evidence. Strict centralized security measures always represent an issue as to basic principle of distinction of powers; this applies namely in the case of judicial infrastructure as well as in the case of information space of state offices that are to be treated independently of the rest of state administration.

The note will discuss most recent constitutional issues in developing of efficient national cybersecurity solutions taking into account not just leading constitutional doctrine and recent constitutional case-law (namely those in data retention cases), but also technical features and specifics of various European national laws (like extraordinary strength of principle of legal evidence in the Czech Republic).

3.10 Casebook on Cyber Forensics (CCF) – a proposal for discussion

Viola Schmid (TU Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license

© Viola Schmid

URL http://www.cylaw.tu-darmstadt.de/home_2/forschung_4/onlinepublikationencylawreports_1/online_publikationen_cylaw_reports.de.jsp

Dagstuhl inspired the idea of a “Casebook on Cyber Forensics” (CFF) from a legal perspective. A lot of questions are connected with this endeavour. First the question of terminology: why not name it casebook on “digital forensics”, “forensic informatics”, or “forensic computing”? The title “cyber forensics” was chosen because these forensics are essential for cyberlaw, the law allocating chances and risks, rights and obligations in cyberspace. Moreover, not only digital data but also data written on paper comes into play.

The second question is: How should such a casebook be structured? Two prototypes — one regarding the format of such a casebook, one regarding the content of such a case book — were presented. The link between format and content is the formula: “form follows function”. First, a “Cylaw Report” of the department of Public Law, Technical University of Darmstadt, on the topic of “Subscription Decoys” (Strafbarkeit von “Abo-Fallen”-Betreibern am Beispiel der “kostenpflichtigen” Vermittlung des Zugriffs auf eigentlich kostenlose Software (Freeware)?) was presented as a paradigm for the potential format¹ of such a CCF.

Members of the Computer Science community could contribute to the description of the facts of the case. Then another “Cylaw Report” on the topic of “Online Searches or Remote Acquisition” (Verdeckte Online-Durchsuchungen — zur IT-(Un)Sicherheit in Deutschland)² was offered for discussion and as an example for the potential content of a CCF. In this case, the federal constitutional court of Germany accepted online searches even if they do not guarantee authenticity and integrity of the data in every case. And also the US-American case *Heckenkamp* was cited as an example that online searches are a transatlantic phenomenon.

A potential table of contents for a CCF would divide the casebook in two parts:

- Part one: Scenarios that are distinguished by the information technology that is analyzed.

¹ <http://tuprints.ulb.tu-darmstadt.de/2201/>

² <http://tuprints.ulb.tu-darmstadt.de/1357/>

■ Part two: Legal Principles such as the exclusionary rule.

Summa summa rum: a lot of work has to be done until this proposal becomes reality.

3.11 Mobile Phone Forensics with the help of ADEL

Michael Spreitzenbarth (University of Erlangen-Nuremberg, DE)

License © ⓘ ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Michael Spreitzenbarth

Joint work of Spreitzenbarth, Michael; Schmitt, Sven; Freiling, Felix

URL http://forensics.spreitzenbarth.de/?page_id=258

Due to the ubiquitous use of smartphones, these devices become an increasingly important source of digital evidence in forensic investigations. Thus, the recovery of digital traces from smartphones often plays an essential role for the examination and clarification of the facts in a case. Although some tools already exist regarding the examination of smartphone data, there is still a strong demand to develop further methods and tools for forensic extraction and analysis of data that is stored on smartphones. In this paper we describe specifications of smartphones running Android. We further introduce a newly developed tool –called ADEL– that is able to forensically extract and analyze data from SQLite databases on Android devices. During our evaluation we found that in contrast to data retained by the network operator, location data stored on the mobile device in many cases offers much more precise information than the rather coarse-grained data from the network operator. However, the availability of data shows a much higher variability on the mobile phone than at the network operator. Finally, a detailed report containing the results of the examination is created by the tool. The whole process is fully automated and takes account of main forensic principles.

3.12 Selective Imaging

Johannes Stuetzgen (University of Erlangen-Nuremberg, DE)

License © ⓘ ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Johannes Stuetzgen

Main reference M. Baecker, F.C. Freiling, S. Schmitt, “Selektion vor der Sicherung,” in: Datenschutz und Datensicherheit – DuD, Volume 34, Number 2, 2010.





URL <http://www1.cs.fau.de/filepool/thesis/diplomarbeit-2011-stuetzgen.pdf>

In an increasingly computerized world, the amount of digital evidence in criminal investigations is constantly growing. In parallel, storage capacities of digital devices scale up every year, to a point where current forensic procedures meet inherent limitations. Furthermore, digital evidence acquisition standards are often unable to comply to data protection regulations, forcing investigators to violate the principle of commensurability frequently, to be able to seize any evidence at all.

Our work aims at streamlining the forensic acquisition process, to enable forensic examiners to selectively acquire only those data objects that are of relevance to the investigation. This approach greatly enhances the scalability of data acquisition methods and enables investigators to respect data protection principles without sacrificing important evidence.

3.13 Legal Challenges of transborder Searches

Nicolas von zur Muehlen (MPI für ausländ. u. internat. Strafrecht-Freiburg, DE)

License     Creative Commons BY-NC-ND 3.0 Unported license
© Nicolas von zur Muehlen

Transborder Searches have been an issue since the early days of the Internet and are still one of the biggest challenges for law enforcement agencies when obtaining digital evidence over the internet. This talk aims to explain the basics of the principle of territoriality. It will address the question of whether the violation of this principle –such as the accessing of data stored on a computer outside national territory– is even justified. Furthermore, the basics of mutual assistance are explained. Finally, this talk deals with the problem that traditional legal concepts can reach their functional limits in global cyberspace, especially when the territorial location of data cannot be pinpointed, as for example in cloud systems.

Participants

- Harald Baier
Hochschule Darmstadt, DE
- Glenn S. Dardick
Longwood Univ. – Virginia, US
- Andreas Dewald
Universität Mannheim, DE
- Felix C. Freiling
Universität Erlangen, DE
- Dieter Gollmann
TU Hamburg-Harburg, DE
- Daniel Hammer
Hochschule Offenburg, DE
- Focke Höhne
Universität Passau, DE
- Joshua James
University College – Dublin, IE
- Stefan Kiltz
Universität Magdeburg, DE
- Kwok Lam
National Univ. of Singapore, SG
- Martin Mink
TU Darmstadt, DE
- Stig Frode Mjolsnes
NTNU – Trondheim, NO
- Christian Moch
Universität Erlangen, DE
- Herbert Neumann
Anwaltskanzlei Neumann –
Molfsee, DE
- Radim Polcák
Masaryk University, CZ
- Joachim Posegga
Universität Passau, DE
- Viola Schmid
TU Darmstadt, DE
- Sven Schmitt
Bundeskriminalamt –
Wiesbaden, DE
- Thomas Schreck
Siemens – München, DE
- Andreas Schuster
Deutsche Telekom – Bonn, DE
- Michael Spreitzenbarth
Universität Erlangen, DE
- Johannes Stüttgen
Universität Erlangen, DE
- Stefan Vömel
Universität Erlangen, DE
- Nicolas von zur Mühlen
MPI für ausländ. und internat.
Strafrecht – Freiburg, DE
- Christian Winter
Fraunhofer SIT – Darmstadt, DE



Computing with Infinite Data: Topological and Logical Foundations

Edited by

Ulrich Berger¹, Vasco Brattka², Victor Selivanov³, Dieter Spreen⁴,
and Hideki Tsuiki⁵

1 University of Wales – Swansea, GB, u.berger@swan.ac.uk

2 University of Cape Town, ZA, Vasco.Brattka@cca-net.de

3 Russian Academy of Sciences – Novosibirsk, RU, s_seliv@yahoo.com

4 Universität Siegen, DE, spreen@math.uni-siegen.de

5 Kyoto University, JP, tsuiki@i.h.kyoto-u.ac.jp

Abstract

There is a large gap between mathematical structures and the structures computer implementations are based on. To stimulate research to overcome this—especially for infinitary structures—highly non-trivial problem the Dagstuhl Seminar 11411 “Computing with Infinite Data: Topological and Logical Foundations” was held. This report collects the ideas that were presented and discussed during the course of the seminar.

Seminar 10.–14. October, 2011 – www.dagstuhl.de/11411

1998 ACM Subject Classification F.4.1 Mathematical Logic, F.1.1 Models of Computation, F.4.3 Formal Languages, D.2.4 Software/Program Verification, F.1.3 Complexity Measures and Classes, G.1.0 Numerical Analysis (General)


Keywords and phrases Exact real number computation, Stream computation, Infinite computations, Computability in analysis, Hierarchies, Reducibility, Topological complexity, Dynamical systems, Languages of infinite words

Digital Object Identifier 10.4230/DagRep.1.10.14

Edited in cooperation with Hannes Diener

1 Executive Summary

Dieter Spreen

License  Creative Commons BY-NC-ND 3.0 Unported license
© Dieter Spreen

In safety-critical applications it is not sufficient to produce software that is only *tested* for correctness: its correctness ought to be proven formally. This remark also applies to the area of scientific computation. An important example are autopilot systems for aircrafts. The problem is that the current mainstream approach to numerical computing uses programming languages that do not possess sound mathematical semantics. Hence, there is no way to provide formal correctness proofs.

The reason is that on the theoretical side one deals with well-developed analytical theories based on the non-constructive concept of a real number. Implementations, on the other hand, use floating-point realizations of real numbers which do not have a well-studied mathematical structure. Approaches to tackle these problems are currently promoted under the slogan “Computing with Exact Real Numbers”.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Computing with Infinite Data: Topological and Logical Found., *Dagstuhl Reports*, Vol. 1, Issue 10, pp. 14–36

Editors: Ulrich Berger, Vasco Brattka, Victor Selivanov, Dieter Spreen, and Hideki Tsuiki



Dagstuhl Reports

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

Well-developed practical and theoretical bases for exact real number computation and, more generally, computable analysis are provided by Scott's Domain Theory and Weihrauch's Type Two Theory of Effectivity (TTE). In both theories real numbers and similar ideal objects are represented by infinite streams of finite objects.

The seminar focused on two problem areas in the realm of computable analysis and computation on infinite streams:

1. Algorithms for stream transforming functions with particular emphasis on (i) logical and category-theoretic methods for the synthesis of provably correct programs, (ii) topological investigations of particular stream representations supporting efficient stream algorithms.
2. Hierarchies and reducibility relations between sets and functions of infinite data as a means of classification. Methods from topology, logic and descriptive set theory were of particular importance in this case.

Infinite streams are infinite words, So there is a close connection to the theory of ω -languages. To study these was a further aim of the seminar.

In the last years much interest in the (logical and topological) structure of the infinite words used to represent continuous data as well as how they code the data space has emerged.

U. Berger and others are developing a constructive theory of digital computation based on co-induction, and are applying it to computable analysis. The aim is to create a mathematical foundation for (lazy) algorithms on analytical data such as real numbers, real functions, compact sets, etc. Since co-induction admits a particularly elegant formalization, this approach is well suited for computer aided modeling and proving in computable analysis. A concrete goal is to use program extraction from proofs as a practical method for obtaining certified programs in computable analysis.

The theory is based on a category of digit spaces. A typical object in this category is a compact metric space with a set of digits, where each digit is a contracting map. A point is then an infinite sequence of function (digit) compositions and hence represented by the corresponding infinite word. Moreover, for any given finite length, the words of that length over the alphabet of digits define a covering of the given space that allows to exactly locate the points represented by infinite words of digits. The set of uniformly continuous functions between such spaces can be characterized by a combined inductive/co-inductive definition that gives rise, via program extraction, to implementations of such functions as non-wellfounded (lazy) trees.

The theory of digit spaces combines known techniques for implementing and verifying stream processing functions (Edalat, Pattinson, Potts, and others) with ideas from co-induction and co-algebra (Jacobs, Rutten, Bertot, Niqui, and others).

H. Tsuiki is pursuing a similar programme, though, from a different perspective. In his case, e.g., every infinite object is represented by exactly one infinite word over $\{0, 1, \perp\}$, where \perp represents "unknown". It turned out that the encoded space has at most topological dimension n , exactly if there are not more than n occurrences of \perp in the corresponding infinite code words.

Besides the problem of how to represent continuous data in an "optimal" way, it is as well an important task to distinguish between computable and non-computable functions and, in the last case, to estimate the degree of non-computability. Most functions are non-computable since they are not even continuous. A somewhat easier and more principal task is in fact to understand the degree of discontinuity of functions. This is mostly achieved by defining appropriate hierarchies and reducibility relations.

In classical descriptive set theory, along with the well-known hierarchies, Wadge introduced and studied an important reducibility relation on Baire space. As shown by van Engelen

et al., von Stein, Weihrauch and Hertling, this reducibility of subsets of topological spaces can be generalized in various ways to a reducibility of functions on a topological space. In this way, the degrees of discontinuity of several important computational problems were classified. It turned out that these classifications refine the so called “topological complexity” introduced in the alternative Blum-Shub-Smale approach to computability on the reals, which is used in complexity considerations in computational geometry.

Recently, reducibilities for functions on topological spaces have been used to identify computational relations between mathematical theorems. This programme, started by V. Brattka et al., can be considered as an alternative to reverse mathematics. Whereas reverse mathematics focuses on set existence axioms required to prove certain theorems, the approach pursued in computable analysis is to identify the computational power required to “compute” certain theorems. The approach, with contributions by M. de Brecht, G. Gherardi, A. Marcone, A. Pauly, M. Ziegler and V. Brattka, has led to deep new insights into computable analysis and has revealed close relations to reverse mathematics, but also some crucial differences.

Motivated initially by decidability problems for monadic second-order logic and Church’s synthesis problem for switching circuits, researchers from automata theory (Büchi, Trakhtenbrot, Rabin, Wagner and many others) developed the theory of ω -languages which provides a foundation of specification, verification and synthesis of computing systems. Topological investigations in this theory have led to several hierarchies and reducibilities of languages of infinite words and trees. Instead of being just continuous, the reduction functions are now required to be computable by automata of suitable type. The resulting hierarchies (like, for example, the Wagner hierarchy) have sometimes the advantage that their levels are decidable.

Note that even the study of automata on finite words now involves topological methods in the classical form of profinite topology. Recently, deep relations of profinite topology to Stone and Priestly spaces were discovered by Pippenger and developed by Gehrke, Grigorieff and Pin. Again, suitable versions of the Wadge reducibility seem to play an important role in development of this field.

The seminar attracted 51 participants representing 15 countries and 5 continents, and working in fields such as computable analysis, descriptive set theory, exact real number computation, formal language theory, logic and topology, among them 10 young researchers working on their PhD or having just finished it. The atmosphere was very friendly, but discussions were most lively. During the breaks and until late into night, participants gathered in small groups for continuing discussions, communicating new results and exchanging ideas.

The seminar led to new research contacts and collaborations. The participants are invited to submit a full paper for a special issue of *Mathematical Structures in Computer Science*. At least one submission deals with a problem posted in the discussions following a talk.

The great success of the seminar is not only due to the participants, but also to the staff, both in Saarbrücken and Dagstuhl, who always do a great job in making everything run efficient and smoothly. Our thanks extend to both groups!

2 Table of Contents

Executive Summary

<i>Dieter Spreen</i>	14
--------------------------------	----

Overview of Talks

An injection from the Baire space to natural numbers	
<i>Andrej Bauer</i>	20
Jumps in the Weihrauch Lattice	
<i>Vasco Brattka</i>	20
Applications of quasi-Polish spaces in computable analysis	
<i>Matthew de Brecht</i>	20
Infinite sets that satisfy the principle of omniscience in all varieties of constructive mathematics	
<i>Martin H. Escardó</i>	21
Kolmogorov complexity and the geometry of Brownian motion	
<i>Willem L. Fouché</i>	21
The effective theory of equivalence relations	
<i>Sy David Friedman</i>	21
Computing invariant measures and pseudorandom points	
<i>Stefano Galatolo</i>	22
Choice operators and the Bolzano-Weierstrass Theorem	
<i>Guido Gherardi</i>	22
Effective theory on arbitrary Polish spaces	
<i>Vassilios Gregoriades</i>	22
Functionals using bounded information	
<i>Serge Grigorieff</i>	23
Types for programming with infinite data	
<i>Peter G. Hancock</i>	23
Countably presentable locales are spatial (proved with a generalization of the Baire Category Theorem)	
<i>Reinhold Heckmann</i>	24
Complexity issues for preorders on finite labeled forests	
<i>Peter Hertling</i>	24
Generalized geometric theories and set-generated classes	
<i>Hajime Ishihara</i>	24
On function spaces and polynomial-time computability	
<i>Akitoshi Kawamura</i>	25
Counterexamples in computable continuum theory	
<i>Takayuki Kihara</i>	25
Boolean algebras of regular languages	
<i>Anton Konovalov</i>	26

Reachability in control polynomial dynamical systems	
<i>Margarita Korovina</i>	26
The Bolzano-Weierstrass principle and the cohesive principle	
<i>Alexander Kreuzer</i>	27
Some remarks related to Katětov’s construction	
<i>Hans-Peter Albert Kuenzi</i>	27
Infinite sequential Nash equilibrium	
<i>Stephane Le Roux</i>	27
Reduction games and reducibilities for sets of reals	
<i>Luca Motto Ros</i>	28
Some steps towards a verified real number arithmetic	
<i>Norbert T. Müller</i>	28
On separation question for tree languages	
<i>Damian Niwinski</i>	29
Conway’s surreal numbers as an inductive-inductive definition	
<i>Fredrik Nordvall Forsberg</i>	29
A finitisation of the infinite Ramsey’s Theorem	
<i>Paulo Oliva</i>	29
The intermediate value theorem is not idempotent	
<i>Arno Pauly</i>	30
Logic, duality and Pervin spaces	
<i>Jean-Eric Pin</i>	30
Computable curves	
<i>Robert Rettinger</i>	30
The extensional versus the intensional hierarchy over the reals	
<i>Matthias Schröder</i>	30
Induction in algebra	
<i>Peter Schuster</i>	31
Simultaneous inductive/coinductive definition of continuous functions	
<i>Helmut Schwichtenberg</i>	31
Computing solution operators of boundary problems for systems of PDE	
<i>Svetlana Selivanova</i>	32
Δ_α^0 -Reductions in quasi-Polish spaces	
<i>Victor Selivanov</i>	32
Joint topologies for finite and infinite words	
<i>Ludwig Staiger</i>	32
Turing machines on represented sets, a model of computation for analysis	
<i>Nazanin Tavana-Roshandel</i>	33
“Good” strategies in infinite games	
<i>Wolfgang Thomas</i>	34
A stream program that takes margin in recursive calls.	
<i>Hideki Tsuiki</i>	34

Computing with infinite data in Lucid
Bill Wadge 34


Uniform polynomial-time maximization of univariate analytic functions
Martin Ziegler. 35

Participants 36

3 Overview of Talks

3.1 An injection from the Baire space to natural numbers

Andrej Bauer (University of Ljubljana, SI)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Andrej Bauer

Infinite-time Turing machines by Joel Hamkins provide a realizability model in which there is an embedding of the Baire space of number-theoretic functions into the natural numbers. The model has other strange features, for example every Π_1^1 statement is decidable, but the model is still intuitionistic.

Nevertheless, the model is interesting both for constructive mathematics as a source of counter-examples, and for computable mathematics as a notion of hyper-computation. In fact, there are Type I and Type II models of infinite-time computability.

(Video of the talk is available at <http://vimeo.com/30368682>.)

3.2 Jumps in the Weihrauch Lattice

Vasco Brattka (University of Cape Town, ZA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Vasco Brattka

We discuss the algebraic structure of the Weihrauch lattice, in particular the operation of a compositional product and a jump. The jump is supposed to play a similar role a Turing jumps play with respect to Turing degrees.

3.3 Applications of quasi-Polish spaces in computable analysis


Matthew de Brecht (NICT – Kyoto, JP)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Matthew de Brecht

We investigate countably based complete quasi-metric spaces, which we call quasi-Polish spaces. We show that this class of spaces is a natural generalization of the class of Polish spaces, but is general enough to include ω -continuous domains and other non-Hausdorff spaces that are important to theoretical computer science. Quasi-Polish spaces can be characterized as the countably based spaces with total admissible representations (with respect to Baire space) and as the spaces that are homeomorphic to the subspace of non-compact elements of an ω -continuous domain. These characterizations suggest that quasi-Polish spaces are important to the study of computable analysis, from both the perspective of Type Two Theory of Effectivity and also for approaches using domain models of spaces. We will also investigate applications of quasi-Polish spaces to the study of degrees of discontinuity of functions between spaces.

3.4 Infinite sets that satisfy the principle of omniscience in all varieties of constructive mathematics


Martin H. Escardó (University of Birmingham, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Martin H. Escardó

In a minimalistic setting for constructive mathematics, without assuming Brouwerian axioms such as continuity, bar induction or fan theorem, we show that there are plenty of infinite sets that satisfy the omniscience principle.

3.5 Kolmogorov complexity and the geometry of Brownian motion


Willem L. Fouché (UNISA – Pretoria, ZA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Willem L. Fouché

We discuss the geometry of Brownian motions which are encoded by Kolmogorov-Chaitin random reals (complex oscillations). We thus interpret Kolmogorov-Chaitin complexity in the context of the geometry of Brownian motion. We outline an effective framework for countable dense random sets of reals.

3.6 The effective theory of equivalence relations


Sy David Friedman (Universität Wien, AT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Sy David Friedman

Effectiveness has played a major role in the development of modern descriptive set theory. Nonetheless, this theme is not emphasized in the current and extremely active area of definable equivalence relations, perhaps due to the fact that there have been so many striking discoveries to be made in the non-effective theory. In this talk I discuss effectively-Borel (= Hyp) reducibility between Hyp equivalence relations on Baire space. The key result that facilitates this study concerns Hyp Wadge-reducibility: There are nonempty Hyp sets of reals A and B such that no Hyp function maps A into B or B into A . From this one can show that there are many pairwise Hyp-incomparable Hyp equivalence relations with only 2 classes and that the Silver and Harrington-Kechris-Louveau dichotomies from the classical theory are not fully effective. Many interesting open questions remain.

3.7 Computing invariant measures and pseudorandom points

Stefano Galatolo (University of Pisa, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stefano Galatolo

Joint work of Galatolo, Stefano; Mathieu Hoyrup; Cristobal Rojas; Isaia Nisoli
Main reference Stefano Galatolo, Mathieu Hoyrup, Cristóbal Rojas, “Dynamical systems, simulation, abstract computation,” arXiv:1101.0833v2 [math.DS]
URL <http://arxiv.org/abs/1101.0833v2>

After recalling some basic notions of ergodic theory, we will consider the problem of (more or less abstract) computation of physical invariant measures of dynamical system.

Those measures contain information on several aspects of the statistical behavior of the dynamics of the system.


We will see that in many interesting situations the physical measure is computable but there are cases of computable systems having no computable invariant measures.

We will apply these results to the problem of the existence of pseudorandom points for the dynamics. These are computable points whose dynamics has a typical statistical behavior. This problem is related to the abstract simulability of the system.

We will also see how to solve some implementation problems in the computation of invariant measures and some experiments in some nontrivial system.

3.8 Choice operators and the Bolzano-Weierstrass Theorem

Guido Gherardi (University of Bologna, IT)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Guido Gherardi


Joint work of Brattka, Vasco; Gherardi, Guido; Marcone, Alberto;
Main reference V. Brattka, G. Gherardi, A. Marcone, “The Bolzano-Weierstrass Theorem is the Jump of Weak König’s Lemma,” arXiv:1101.0792v2 [math.LO]
URL <http://arxiv.org/abs/1101.0792v2>

The computational complexity of the Bolzano-Weierstrass Theorem operator in different computable metric spaces is investigated.

Its relationships with Closed Choice and Compact Choice operators are analyzed. It is in particular shown that in every computable metric space the Bolzano-Weierstrass operator is strongly Weihrauch equivalent to the derivative of Compact Choice operator.

3.9 Effective theory on arbitrary Polish spaces

Vassilios Gregoriades (TU Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Vassilios Gregoriades

Main reference V. Gregoriades, “Effective Theory on arbitrary Polish spaces,” 8th Panhellenic Logic Symposium, Ioannina, Greece, July 4–8, 2011.


We present some results regarding the class of Polish spaces which admit a recursive presentation. In particular we show that -contrary to the classical case- a recursively presented Polish space which is an uncountable set need not be effectively-Borel isomorphic to the Baire space. In order to do this we define a constructive scheme for Polish spaces by assigning to every tree T on ω a Polish space N_T which is recursively presented in T .

The effective structure of N_T depends on the combinatorial properties of T , so that for various choices of a recursive tree T the spaces N_T are not effectively-Borel isomorphic to each other. On the other hand every recursively presented Polish space is up to effective-Borel isomorphism one of the spaces N_T . This leads to the natural question of studying the classes of effective-Borel isomorphism of the spaces N_T . Some preliminary results will be exhibited.

3.10 Functionals using bounded information

Serge Grigorieff (LIAFA, CNRS & Université Paris 7)

Joint work of Grigorieff, Serge; Valarcher, Pierre

License  Creative Commons BY-NC-ND 3.0 Unported license
© Serge Grigorieff

Let us define the size of a basic clopen set $[u] = \{f \in \mathbb{N}^{\mathbb{N}} \mid f \text{ extends } u\}$ (where u is a partial function $\mathbb{N} \rightarrow \mathbb{N}$ with finite domain) as the cardinality of the domain of u . Continuity of a functional $\Phi : \mathbb{N}^{\mathbb{N}} \rightarrow \mathbb{N}$ requires that there exists some covering of $\mathbb{N}^{\mathbb{N}}$ by basic clopen sets on which Φ is constant. Consider the following conditions on a continuous Φ .

- (1_k) (Φ is locally constant on a covering by clopens of size k)
There is a covering of $\mathbb{N}^{\mathbb{N}}$ by basic open sets with size at most k on which Φ is constant.
- (2_k) (Φ is locally constant on a covering by disjoint clopens of size k)
There is a covering of $\mathbb{N}^{\mathbb{N}}$ by pairwise disjoint basic open sets with size at most k on which Φ is constant.
- (3_k) (Φ uses k -bounded deterministic information)
There exists an algorithm, using some function $\mathbb{N} \rightarrow \mathbb{N}$ as oracle, which, for every $f \in \mathbb{N}^{\mathbb{N}}$, asks for only k values of f to compute $\Phi(f)$.
- (4_k) (Φ can be computed by some Gurevich Abstract State Machine)
There exist an integer a and $k - 1$ functions $(\omega_i : \mathbb{N}^i \rightarrow \mathbb{N})_{0 < i < k}$ such that there is an algorithm, using these functions as oracles, which, for every $f \in \mathbb{N}^{\mathbb{N}}$, computes $\Phi(f)$ using only k values of f at


$$a, \omega_1(f(a)), \omega_2(f(a), f(\omega_1(f(a)))), \\ \omega_3(f(a), f(\omega_1(f(a))), f(\omega_2(f(a), f(\omega_1(f(a)))))), \dots$$

Implications $(4_k) \Rightarrow (3_k) \Rightarrow (2_k) \Rightarrow (1_k)$ are straightforward.

We prove $(1_k) \Rightarrow (4_{k^2})$ (k^2 is optimal already for $(1_k) \Rightarrow (2_{k^2})$) and an effective version $(1_k^{\text{effective}}) \Rightarrow (4_{B(k)}^{\text{effective}})$ for some primitive recursive B .

3.11 Types for programming with infinite data


Peter G. Hancock (The University of Strathclyde – Glasgow, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Peter G. Hancock

It seems to me worthwhile to investigate the type structure needed to program with infinite data. Programs that implement continuous functions on ultrametric spaces are themselves infinite data of “higher order”. Some kind of exponential or arrow type is lurking in the vicinity, perhaps even in broad daylight.

3.12 Countably presentable locales are spatial (proved with a generalization of the Baire Category Theorem)


Reinhold Heckmann (AbsInt – Saarbrücken, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Reinhold Heckmann

We present the category of countably presentable locales, i.e. locales whose frame of opens has a presentation with countably many generators and countably many relations. All such locales are spatial, which can be proved without choice (but with excluded middle) using a generalization of the Baire Category Theorem.

3.13 Complexity issues for preorders on finite labeled forests

Peter Hertling (Universität der Bundeswehr – München, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Peter Hertling

Joint work of Hertling, Peter; Selivanov, Victor;

Main reference P. Hertling and V.L. Selivanov, “Complexity Issues for Preorders on Finite Labeled Forests,” in: B. Löwe, D. Normann, I. N. Soskov, A. A. Soskova (eds) 7th Conference on Computability in Europe, CiE 2011, Proceedings, LNCS, vol. 6735, Springer, pp. 112–121, 2011.

URL http://dx.doi.org/10.1007/978-3-642-21875-0_12

We prove that three preorders on the finite k -labeled forests are polynomial time computable. Together with an earlier result of the first author, this implies polynomial-time computability for an important initial segment of the corresponding degrees of discontinuity of k -partitions on the Baire space.

Furthermore, we show that on omega-labeled forests the first of these three preorders is polynomial time computable as well while the other two preorders are NP-complete.

3.14 Generalized geometric theories and set-generated classes

Hajime Ishihara (JAIST – Nomi, JP)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Hajime Ishihara


We introduce generalized geometric theories over a set S and their models which are subsets of S .

Then we show that the class M of models of a generalized geometric theory is set-generated in CZF, that is, there exists a subset G of M such that each set a in M is the union of subsets of a in G .

We also present some applications to algebra, topology, formal topology and Sambin’s theory of basic pairs.

3.15 On function spaces and polynomial-time computability

Akitoshi Kawamura (*University of Tokyo, JP*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Akitoshi Kawamura


In Computable Analysis, elements of uncountable spaces, such as the real line \mathbb{R} , are represented by functions on strings (or equivalently as infinite strings) and fed to the Turing machine as oracles. To obtain reasonable notions of computability and complexity, it is thus important that we choose the “right” representation (encoding) of the spaces being considered.

Let’s say we have already agreed upon representations γ and δ of spaces X and Y (that are admissible with the topologies of X and Y). How would we represent the space $C[X \rightarrow Y]$ of continuous functions from X to Y ? It is known that there is a natural representation $[\gamma \rightarrow \delta]$ of $C[X \rightarrow Y]$ which is characterized by the property that it is the poorest representation that makes function evaluation computable.

Is there a representation with a similar property also at the level of polynomial-time computability (as introduced by Ko and Friedman and extended by Kawamura and Cook)? In this talk I will point out that there is such a nice representation for the space $C[\mathbb{R} \rightarrow \mathbb{R}]$, but it is not clear how generally a similar construction is possible. I would like to leave this question to open discussion at the seminar.

3.16 Counterexamples in computable continuum theory

Takayuki Kihara (*Tohoku University, JP*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Takayuki Kihara

We provide several pathological examples of tree-like co-c.e. continua fulfilling certain incomputability properties:

(1) there is a computable planar dendrite which is not approximable from the inside by a co-c.e. tree; (2) there is a co-c.e. planar dendrite which is not approximable from the inside by a computable dendrite; (3) there is a computable contractible planar dendroid which is not approximable from the inside by a co-c.e. dendrite; (4) there is a co-c.e. contractible planar dendroid which has no computable point.


As corollaries:

(1) there is a planar computable closed set which is [computable path]-connected, but not [co-c.e. arc]-connected; (2) there is a contractible and locally contractible planar co-c.e. closed set which is not approximable from the inside by a connected computable closed set; (3) there is a contractible planar computable closed set which is not approximable from the inside by a connected and locally connected co-c.e. closed set; (4) there is a contractible planar co-c.e. closed set which has no computable point.

(This is the solution to Question of le Roux-Ziegler, and it is also a new non-basis result on Π_1^0 classes.)

3.17 Boolean algebras of regular languages

Anton Konovalov (A. P. Ershov Institute – Novosibirsk, RU)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Anton Konovalov

Joint work of Konovalov, Anton; Selivanov, Victor

We characterize up to isomorphism the Boolean algebras of regular languages and of regular aperiodic languages, and show decidability of classes of regular languages related to these characterizations.

3.18 Reachability in control polynomial dynamical systems

Margarita Korovina (The University of Manchester, UK and IIS SB RAS Novosibirsk, RU)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Margarita Korovina


Joint work of Korovina, Margarita; Vorobjov, Nicolai

A fundamental problem in the design of biological, chemical or physical processes is to automatically synthesise models from performance specifications. In general, by practical and theoretical reasons, it is highly nontrivial to achieve. However, in some cases, given by partial designs, it could be possible to automatically complete modelling in order to get desired properties. In this talk we discuss synthesis of finite control strategies to meet reachability and time requirements for partial designs given by controlled polynomial dynamical systems.

A controlled polynomial dynamical system is defined by a polynomial depending on control parameters. The choice of a parameter determines a certain motion. In general case, in order to achieve reachability or time requirements it is necessary to switch between motions corresponding to various control parameters at certain points of time. We focus on the following problem. For a partial design, given by CPDS, determine whether there exist finite sequences of time points and control parameters that guide the system from an initial state to a desired state. If the answer is positive then a finite control strategy is automatically synthesised. In the general class of o-minimal dynamical systems this problem is undecidable. Indeed, it has been shown that the reachability problem is already undecidable for three-dimensional piecewise constant derivative systems and two-dimensional o-minimal dynamical systems with non-determinism. Moreover, for one-dimensional o-minimal dynamical systems with non-determinism, the problem remains open. In this paper we show that it is possible to find finite control strategies for the certain broad class of one-dimensional controlled polynomial dynamical systems. The key results are summarised next. Firstly, we prove that for this class of one-dimensional controlled polynomial dynamical systems the existence of a finite control strategy is decidable and construct an polynomial complexity algorithm which synthesises finite control strategies. Secondary, for the algorithm we show an upper bound on the numbers of switches in finite control strategies. Finally, we prove that finite control strategies generated by the algorithm are time-optimal.

3.19 The Bolzano-Weierstrass principle and the cohesive principle

Alexander Kreuzer (TU Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Alexander Kreuzer

Let BW be the usual Bolzano-Weierstrass principle, i.e. the statement that each bounded sequence of real numbers contains a subsequence converging at the rate 2^{-n} . Let BW_{weak} be the statement that each bounded sequence of reals contains a -possibly slowly- converging subsequence, i.e. a subsequence converging but possibly without any computable rate of convergence.

We show that BW is instancewise equivalent to WKL for Σ_1^0 -trees and that BW_{weak} is instancewise equivalent to the cohesive principle.


Thus the degrees $d \gg 0'$ (i.e. the degrees d that contain an infinite branch of each $0'$ -computable 0/1-tree) are exactly those degrees that contain for each computable sequence (x_n) a subsequence converging at the rate 2^{-n} . In particular, there is a degree d that is low over $0'$ that contains a solution of each computable instance of BW.

Using the classification of the cohesive principle of Jockusch and Stephan one obtains that a slowly converging subsequence of (x_n) is computable in a degree d that is low_2 , i.e. $d'' = 0''$, and thus that BW_{weak} does not compute $0'$.

We also comment on the strength of Bolzano-Weierstrass principle for weak compactness on the Hilbert space ℓ_2 .

3.20 Some remarks related to Katětov's construction

Hans-Peter Albert Kuenzi (University of Cape Town, ZA)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Hans-Peter Albert Kuenzi

If one deletes in the usual definition of a metric space the symmetry condition, one obtains the concept of a quasi-metric space. (A more precise definition will be given in the talk.) We discuss an asymmetric approach to M. Katětov's functions (compare with his well-known article "On universal metric spaces").

Our approach turned out to be very useful in our recent work on hyperconvexity (joint work with E. Kemajou and O.O. Otafudu) and on universality (joint work with M. Sanchis) in quasi-metric spaces.

3.21 Infinite sequential Nash equilibrium

Stephane Le Roux (TU Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Stephane Le Roux

Generalisation of Martin's Theorem for games with many agents (instead of two) and many outcomes (instead of two) and Nash equilibrium (instead of winning strategy).

3.22 Reduction games and reducibilities for sets of reals

Luca Motto Ros (Universität Freiburg, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Luca Motto Ros

The Wadge hierarchy has been generalized in many directions. One of these directions is to enlarge the class of reducing functions from the set of continuous functions to some natural class \mathcal{F} , usually called *reducibility*. Examples of these reducibilities are the classes D_α of those f such that $f^{-1}(D) \in \Delta_\alpha^0$ for every $D \in \Delta_\alpha^0$.


Given such an \mathcal{F} and two sets $A, B \subseteq {}^\omega\omega$, we say that A is \mathcal{F} -reducible to B ($A \leq_{\mathcal{F}} B$ in symbols) just in case there is $f \in \mathcal{F}$ such that $f^{-1}(B) = A$. The preorder induced by $\leq_{\mathcal{F}}$ on the quotient of $\mathcal{P}({}^\omega\omega)$ with respect to the equivalence relation induced by $\leq_{\mathcal{F}}$ is called \mathcal{F} -*hierarchy*. The structure of each of the D_α -hierarchies, $\alpha < \omega_1$, has been completely determined under a certain weakening of the Axiom of Determinacy, called Semi-Liner Ordering principle for continuous functions, and it turns out to be isomorphic to the Wadge one. However, while in the cases $\alpha = 1, 2$ the proof of this fact follows from a characterization of the functions in D_α in terms of games, in the other cases the original proof is of a topological nature and heavily relies on the structure of the Wadge degrees.

In this talk I will introduce the concept of a reduction game, and show that many reducibilities can be characterized in terms of these games. This approach has two advantages:

- one can use reduction games to have a more “combinatorial” proof of the fact that the D_α -hierarchies are isomorphic to the Wadge one;
- unlike in the original proof, the axioms needed for this alternative proof are potentially weaker than the one used to determine the Wadge hierarchy.

3.23 Some steps towards a verified real number arithmetic

Norbert T. Müller (Universität Trier, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Norbert T. Müller


Computing with real numbers requires special care about the validity of the found results.

Quite often, billions of operations can be performed without big problems concerning the precision. But there are also important examples where already less than 100 dependent arithmetic operations lead to grossly wrong results when performed with the usual 64bit floating point numbers.

In the talk we present an approach where exact real arithmetic is used, so rounding errors are completely avoided. Currently, we are working on tools that aim at the use of interactive proofs systems like COQ for the formal verification of the (partial) correctness of the software package.

3.24 On separation question for tree languages

Damian Niwinski (University of Warsaw, PL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Damian Niwinski

Joint work of Niwinski, Damian; Arnold, Andre; Michalewski, Henryk

Once we know that a class C is different from $\text{co-}C$, we may ask a more subtle question: can any two disjoint sets in C be “approximated” by some disjoint super-sets in $\text{co-}C$?


In the classical hierarchies (Borel, projective, Wadge...), typically one of the two dual classes on each level enjoys this property. We pursue the question for the Rabin-Mostowski index hierarchy of alternating automata on infinite trees.

The property turns out to fail for all Sigma classes, however it is open if it holds for the Pi classes for levels > 2 .

The result for trees comes through an analogous result for words, more precisely, for the index hierarchy of deterministic automata on infinite words. In this case we solve the problem completely: the approximation (separation) property holds for Pi classes and fails for Sigma classes. As a by-product we discover a simplification of the Arnold’s proof of the strictness of the Rabin-Mostowski index hierarchy. More specifically: the use of the Banach Fixed-Point Theorem can be replaced by an explicit construction of a fixed point.

3.25 Conway’s surreal numbers as an inductive-inductive definition


Fredrik Nordvall Forsberg (University of Wales – Swansea, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Fredrik Nordvall Forsberg

The class of surreal numbers contains both the real numbers and the class of ordinals. We show how they can be naturally represented in Martin-Löf type theory by an inductive-inductive definition, and take the opportunity to introduce the principle of such definitions.

3.26 A finitisation of the infinite Ramsey’s Theorem


Paulo Oliva (Queen Mary University of London, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Paulo Oliva

Alexander Kreuzer and Ulrich Kohlenbach have recently shown how the Erdős/Rado proof of the infinite Ramsey theorem can be formalised using “weak” König’s lemma for Σ_1^0 definable trees. One can view their proof as a combination of three major principles: (1) Π_1 countable choice, (2) weak König’s lemma, and (3) the infinite pigeon-hole principle. In this talk we see how each of these three principles has a neat computational interpretation via the product of selection functions. A combination of these three applications of the products of selection functions gives a construction that witnesses the no- counterexample (meta-stability) version of the infinite Ramsey’s theorem.

3.27 The intermediate value theorem is not idempotent

Arno Pauly (University of Cambridge, GB)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Arno Pauly

As an example for both results and proof-techniques in computable reverse mathematics, a proof is given that the intermediate value theorem is not idempotent. This means that we it is impossible to solve two instances of the intermediate value theorem using a solution of a single instance together with computable means.

The proof techniques we use apply to choice principles in general, these are multi-valued functions mapping negative information about closed sets to members of the sets. As choice principles seem to be ubiquitous in computable reverse mathematics, the techniques are promising to be useful in many more cases.

3.28 Logic, duality and Pervin spaces

Jean-Eric Pin (University Paris-Diderot, FR)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Jean-Eric Pin

In a recent paper, we proved that any lattice of [regular] languages can be defined by a set of [profinite] “equations”. This result applies to any set of languages defined by a reasonable fragment of logic.

This result involves the description of the Stone dual of a lattice of languages of A^* . For the Boolean algebra of all regular languages, the dual space is the completion of A^* for a certain metric. What about the general case? Completions of quasi-uniformities do the job, but actually only a very special case is needed: the Pervin spaces. Turning to Pervin spaces simplifies a number of results and leads to an alternative point of view on Stone’s duality.

3.29 Computable curves


Robert Rettinger (FernUniversität in Hagen, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Robert Rettinger

Several computability notions of curves in \mathbb{R}^2 will be discussed. We put the emphasis on open problems on this topic.

3.30 The extensional versus the intensional hierarchy over the reals

Matthias Schröder (Unibw – München, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Matthias Schröder

In functional programming, there are essentially two approaches to computability on the real numbers. The *extensional* approach assumes an idealistic functional language containing


the real numbers as an own datatype. The *intensional* approach uses data structures of ordinary functional languages and encodes real numbers as streams using the signed-digit representation.

It is known that both approaches yield the same classes of Type-1 and Type-2 functionals over the reals. This has been shown by Bauer, Escardó and Simpson. Whether this is also the case for functionals of type $n \geq 3$ was an open problem for a long time.

In this talk I will prove the non-coincidence of the hierarchies from level 3 on. I will do this by using a result by Normann. He came up with a purely topological condition on the Kleene-Kreisel functionals over the natural numbers which is equivalent to the Coincidence Problem. So I will show that the Kleene-Kreisel space $\mathbb{N}^{\mathbb{N}}$ does not satisfy this topological property.

3.31 Induction in algebra

Peter Schuster (University of Leeds, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Peter Schuster

Many a concrete theorem of abstract algebra admits a short and elegant proof by contradiction but with Zorn's Lemma (ZL). A few of these theorems have recently turned out to follow in a direct and elementary way from the Principle of Open Induction distinguished by Raoult. A proof of the latter kind may be obtained systematically from a proof of the former sort, and the tree one can grow alongside the induction encodes the computation corresponding to the theorem. If the theorem has finite input data, then a finite partial order carries the required instance of induction, which thus is constructively provable. The ideal objects characteristic of any invocation of ZL are eliminated, and it is made possible to pass from classical to intuitionistic logic. This approach is intended as a contribution to a partial realisation of Hilbert's Programme, and was motivated by related work of Berger, Coquand and by the rise of dynamical and logical approaches to algebra.

3.32 Simultaneous inductive/coinductive definition of continuous functions


Helmut Schwichtenberg (Universität München, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Helmut Schwichtenberg

When extracting computational content from proofs in constructive analysis it can be helpful to use simultaneous inductive/coinductive definitions of (uniformly) continuous real functions. The talk reports on an attempt to design the underlying theory, based on recent work of Ulrich Berger.

3.33 Computing solution operators of boundary problems for systems of PDE


Svetlana Selivanova (Sobolev Institute of Mathematics – Novosibirsk, RU)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Svetlana Selivanova

We discuss possibilities of applications of numerical analysis methods to proving computability (in the sense of the TTE approach) of the solution operators for boundary problems for systems of PDE.

3.34 Δ_α^0 -Reductions in quasi-Polish spaces

Victor Selivanov (A. P. Ershov Institute – Novosibirsk, RU)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Victor Selivanov

There are several directions to generalize the classical Wadge reducibility on the Baire space, in particular to a wider class of reducing functions or to more complicated topological spaces. For a space X and a pointclass $C \subseteq P(X)$, C -reducibility is the preorder on $P(X)$ corresponding to many-one reductions by functions on X such that the preimage of any set in C is again in C . In a series of papers, A. Andretta, D. Martin and L.M. Ros have shown that, under suitable set-theoretic assumptions, the structure of C -degrees in the Baire space is isomorphic to the structure of Wadge degrees, where C is the class of Borel sets or is a level of the Borel hierarchy. P. Hertling has shown that the structure of Wadge degrees in the space of reals is much more complicated than the structure of Wadge degrees in the Baire space.

We show that for many C -reducibilities (this applies e.g. to the case when C is any infinite level of the Borel hierarchy) the structure of C -degrees in any uncountable quasi-Polish space X is isomorphic to the structure of Wadge degrees in the Baire space. This immediately follows from the following extension and refinement of a classical fact: any two uncountable quasi-Polish spaces X, Y are C -isomorphic, where $C(X)$ is the class of sets of finite Borel rank in X , and C -isomorphism is a bijection between X and Y which preserves the classes $C(X)$ and $C(Y)$ in both directions. Quasi-Polish spaces is a natural class of spaces (countably-based completely quasi-metrizable spaces) containing all Polish spaces and all omega-continuous domains.

3.35 Joint topologies for finite and infinite words

Ludwig Staiger (Martin-Luther-Universität Halle-Wittenberg, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Ludwig Staiger

Infinite words are often considered as limits of finite words. As topological methods have been proved to be useful in the theory of omega-languages it seems to be providing to include finite and infinite words into one (topological) space. The attempts so far have their drawbacks.


Therefore, in the present paper we investigate the possibility to join separate topologies on the space of finite words with a topology in the space of infinite words via a natural mapping. A requirement in this linking of topologies consists in the compatibility of topological properties (openness, closedness etc) of images with pre-images and vice versa.

Here we choose the natural CANTOR topology for infinite words and the delta-limit as linking mapping, and we show that several natural topologies on the space of finite words prove to be compatible with the topology of the CANTOR space. It is interesting to observe that besides the well-known prefix topology there are at least two more whose origin is from language theory, from the construction of centers and super-centers of languages.

These center- and supercenter-topologies on the space of finite words, fit into the class of L-topologies investigated by Proding. Moreover they exhibit special properties within the classes of topologies compatible with the CANTOR topology.

3.36 Turing machines on represented sets, a model of computation for analysis

Nazanin Tavana-Roshandel (Amir Kabir University of Technology – Teheran, IR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Nazanin Tavana-Roshandel

Joint work of Tavana-Roshandel, Nazanin; Weihrauch, Klaus


We introduce a new type of generalized Turing machines (GTMs), which is intended as a tool for the mathematician who studies computability in Analysis.

In a single tape cell a GTM can store a symbol, a real number, a continuous real function or a probability measure, for example. The model is based on TTE, the representation approach for computable analysis. As a main result we prove that the functions that are computable via given representations are closed under GTM programming. This generalizes the well known fact that these functions are closed under composition. The theorem allows to speak about objects themselves instead of names in algorithms and proofs. By using GTMs for specifying algorithms, many proofs become more rigorous and also simpler and more transparent since the GTM model is very simple and allows to apply well-known techniques from Turing machine theory. We also show how finite or infinite sequences as names can be replaced by sets (generalized representations) on which computability is already defined via representations.

This allows further simplification of proofs. All of this is done for multi-functions, which are essential in Computable Analysis, and multi- representations, which often allow more elegant formulations. As a byproduct we show that the computable functions on finite and infinite sequences of symbols are closed under programming with GTMs. We conclude with examples of application.

3.37 “Good” strategies in infinite games

Wolfgang Thomas (RWTH Aachen, DE)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Wolfgang Thomas

We discuss the algorithmic synthesis of winning strategies in regular infinite games, focusing on the following aspects:

- (1) a new approach to connect the logical format of winning conditions (requirements) and of winning strategies,
- (2) quantitative refinements concerning properties of infinite plays (e.g. the amount of nondeterminism that can be realized in a strategy, the amount of lookahead that can be granted to the opponent without affecting the possibility to win),
- (3) the finite number of moves that suffices to decide the winner of a game.

3.38 A stream program that takes margin in recursive calls.

Hideki Tsuiki (Kyoto University, JP)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Hideki Tsuiki

Joint work of Tsuiki, Hideki; Yamada, Shuji

Main reference Hideki Tsuiki, Shuji Yamada, “On Finite-time Computability Preserving Conversions,” J. UCS 15(6): 1365–1380 (2009).

In this talk, we present a new kind of recursively-defined function on infinite sequences. We introduce the notion of a finite-time computable function and then a finite-time computability preserving conversion (ftcp in short) as a function which preserve finite-time computability. Then, we show that a ftcp function f can be written as


$$f(x) = (gx) ++ (\text{drop}(\mu x)(f(\text{tail } x)))$$

for a computable function g which takes an infinite list and returns a finite list, and a computable function μ which takes an infinite list and returns a number. That is, the computation of f proceeds as producing by itself some part of the output (gx) and then makes a recursive call with the tail, and takes (μx) number of output as a kickback and outputs only the rest.

We show that a ftcp function can be represented as an extended sliding block function, and that it is suffix-identity preserving.

3.39 Computing with infinite data in Lucid

Bill Wadge (University of Victoria, CA)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Bill Wadge

The dataflow language Lucid, invented by the author and E. A. Ashcroft around 1976, was one of the first to incorporate infinite data.

Originally, the infinite data took the form of streams of finite objects indexed by the natural numbers. Programmers could think of these streams as being generated incrementally


in a dataflow network. This modest feature proved surprisingly powerful when combined with recursively defined stream functions (“filters”) because the corresponding dataflow network itself grows (incrementally). The next simple step was to add so-called “space” (and other) dimensions. A variable could, for example, depend on one time and two space dimensions and be thought of as a stream of infinite matrices. We give examples of useful programs using these possibilities.

Two problems arose, however, which still await complete solutions. One is the problem of caching values of variables in the presences of large numbers of dimensions – when searching for a cached value, how do we know which coordinate values are relevant to our search? The second is the semantics of local (in terms of scope) dimensions. In particular, when local dimensions are combined with filter recursion, it appears to entail the existence of infinitely many simultaneously active dimensions.

We will discuss some of the current approaches (due to the author, A. Faustini, J. Plaiice and others) to solving these serious problems.

3.40 Uniform polynomial-time maximization of univariate analytic functions

Martin Ziegler (TU Darmstadt, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Martin Ziegler

Joint work of Kawamura, Akitoshi; Müller, Norbert; Rettinger, Robert; Rösnick, Carsten

Ko and Friedman have shown in the 1980ies that the maximum and the integral of a smooth (i.e. infinitely differentiable) polynomial-time computable real function is in general again polytime-computable iff P equals NP and $\#P$, respectively. For polytime-computable analytic functions f , on the other hand, both maximum and integral are polytime-computable [Ko’91, Section 6.2], [Müller’87] — nonuniformly, i.e. for fixed f :

One reason being that a satisfactory uniform complexity theory of real operators has only been devised recently by Kawamura and Cook (2010).

But secondly, the known algorithms actually implicitly exploit ‘knowing’ many unspecified parameters of the fixed function to be maximized.

We present an (almost) uniform algorithm for calculating the maximum of a given analytic function on a given Jordan domain in polynomial time.

Participants

- Andrej Bauer
University of Ljubljana, SI
- Veronica Becher
University of Buenos Aires, AR
- Ulrich Berger
Univ. of Wales – Swansea, GB
- Jens Blanck
Univ. of Wales – Swansea, GB
- Vasco Brattka
University of Cape Town, ZA
- Matthew de Brecht
NICT – Kyoto, JP
- Hannes Diener
Universität Siegen, DE
- Martin H. Escardo
University of Birmingham, GB
- Willem L. Fouché
UNISA – Pretoria, ZA
- Sy David Friedman
Universität Wien, AT
- Stefano Galatolo
University of Pisa, IT
- Guido Gherardi
University of Bologna, IT
- Vassilios Gregoriades
TU Darmstadt, DE
- Serge Grigorieff
University Paris-Diderot, FR
- Peter G. Hancock
The University of Strathclyde –
Glasgow, GB
- Reinhold Heckmann
AbsInt – Saarbrücken, DE
- Peter Hertling
Universität der Bundeswehr –
München, DE
- Tie (Caroline) Hou
Univ. of Wales – Swansea, GB
- Hajime Ishihara
JAIST – Nomi, JP
- Akitoshi Kawamura
University of Tokyo, JP
- Takayuki Kihara
Tohoku University, JP
- Anton Konovalov
A. P. Ershov Institute –
Novosibirsk, RU
- Margarita Korovina
Manchester University, GB
- Alexander Kreuzer
TU Darmstadt, DE
- Hans-Peter Albert Künzi
University of Cape Town, ZA
- Stéphane Le Roux
TU Darmstadt, DE
- Luca Motto Ros
Universität Freiburg, DE
- Norbert T. Müller
Universität Trier, DE
- Damian Niwinski
University of Warsaw, PL
- Fredrik Nordvall Forsberg
Univ. of Wales – Swansea, GB
- Dag Normann
University of Oslo, NO
- Paulo Oliva
Queen Mary University of
London, GB
- Arno Pauly
University of Cambridge, GB
- Jean-Eric Pin
University Paris-Diderot, FR
- Robert Rettinger
FernUniversität in Hagen, DE
- Carsten Roesnick
TU Darmstadt, DE
- Jan Rutten
CWI – Amsterdam, NL
- Matthias Schröder
Unibw – München, DE
- Peter Schuster
University of Leeds, GB
- Helmut Schwichtenberg
Universität München, DE
- Victor Selivanov
A. P. Ershov Institute –
Novosibirsk, RU
- Svetlana Selivanova
Sobolev Institute of Mathematics
– Novosibirsk, RU
- Anton Setzer
Univ. of Wales – Swansea, GB
- Dieter Spreen
Universität Siegen, DE
- Ludwig Staiger
Martin-Luther-Universität
Halle-Wittenberg, DE
- Nazanin Tavana-Roshandel
Amir Kabir University of
Technology – Teheran, IR
- Wolfgang Thomas
RWTH Aachen, DE
- Hideki Tsuiki
Kyoto University, JP
- Bill Wadge
University of Victoria, CA
- Klaus Weihrauch
FernUniversität in Hagen, DE
- Martin Ziegler
TU Darmstadt, DE



Foundations of distributed data management

Edited by

Serge Abiteboul¹, Alin Deutsch², Thomas Schwentick³, and
Luc Segoufin⁴

1 INRIA - Orsay Cedex, FR, serge.abiteboul@inria.fr

2 University of California – San Diego, US, deutsch@cs.ucsd.edu

3 TU Dortmund, DE, thomas.schwentick@udo.edu

4 ENS – Cachan, FR, luc.segoufin@inria.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11421 “Foundations of distributed data management”.

Seminar 16.–21. October, 2011 – www.dagstuhl.de/11421

1998 ACM Subject Classification C.2.4 Distributed Systems, H.2 Database Management, H.3.5 Online Information Services

Keywords and phrases XML Query language, Distribution, Incompleteness

Digital Object Identifier 10.4230/DagRep.1.10.37

Edited in cooperation with Tom Ameloot

1 Executive Summary

Serge Abiteboul

Alin Deutsch

Thomas Schwentick

Luc Segoufin

License  Creative Commons BY-NC-ND 3.0 Unported license
© Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin

Description of the Seminar’s Topic

The Web has brought fundamentally new challenges to data management. Web data management differs from traditional database management in a number of ways. First, Web data differ in their structure: trees with links (usually described by mark-up languages such as XML) instead of tables. Also, Web data are by nature distributed, often on a large number of autonomous servers. Finally, Web data are typically very dynamic and imprecise.

Unlike for the classical relational database model, there is still no commonly accepted model for data management over the Web. The lack of a clean, simple, mathematical model further prevents us from designing general solutions to typical data management problems, such as building indexes, optimizing queries, and guaranteeing certain properties of applications.

As witnessed by the two seminars that previously occurred in Dagstuhl on this topic (Seminar 01361 in 2001 and Seminar 05061 in 2005, both entitled “Foundations of Semistructured Data”), most of the recent research efforts have concentrated on adapting traditional database techniques to the XML setting. In particular, foundational research on XML focused on the tree structure of XML documents, applying well-developed techniques based on logic and



Except where otherwise noted, content of this report is licensed
under a Creative Commons BY-NC-ND 3.0 Unported license
Foundations of distributed data management, *Dagstuhl Reports*, Vol. 1, Issue 10, pp. 37–57
Editors: Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin



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

automata for trees. These lines of research have been very successful. However, they do not address all the facets of Web data. In particular distribution, dynamicity, incompleteness and reliability had received limited attention in past work, but play a central role in a Web setting. The aim of Seminar 11421 was to bring together researchers covering this spectrum of relevant areas, to report on recent progress in terms of both results as well as new, relevant research questions. It was organized at the initiative of members of the EU funded research projects FoX (fox7.eu) and Webdam (webdam.inria.fr) that are acknowledged for their support.

The seminar focused on the following key aspects of Web data management.

Semistructured data/query languages, with particular emphasis on XML/XPath and RDF/SPARQL. Semistructured data is the preferred way to organize information on the Web, and de facto and de jure standards are emerging. The study of XML data management and XML query languages remains a constant in the entire line of seminars culminating with 11421. Four notable additions over the predecessor seminars deserve mentioning. First is the emphasis on XML with Data values (and in its simplified version, on Data Words), where the data labeling of an XML tree (a word) is drawn from an infinite domain. Query evaluation, and static analysis tasks in this setting are considerably harder, often undecidable, and finding the right limitations of the logics used for querying and of the data models is still an open research challenge. Second is the RDF data model, with its associated query language SPARQL. They are used for modeling/querying semantic Web data (ontologies), but also classical semistructured data. Since the standardization process is still ongoing, the work performed by researchers in our extended community has significant potential for impact. Indeed, one of the seminar participants, Marcelo Arenas, sits on the standard working group and is a leader in studying the semantics, query evaluation complexity, as well as optimization potential of the SPARQL language. Third is the emphasis on static typing. This is applied to XML data, schema inference, and the experimental evaluation of large collections of XML schemas found in real life (the work conducted in the framework of the above mentioned project FoX is relevant). In addition, the new SPARQL language is in need of foundational contributions towards type inference. Fourth is the work on related languages which, while not being under consideration as official standards, have established themselves as quasi-standards for querying graph databases in the database theory community. These include variations on the language of regular path queries, in which reachability queries in the graph are expressed using various classes of regular expressions over the alphabet of edge labels.

Incomplete and Probabilistic Databases. Information found on the Web is often incomplete, or uncertain due to contradictory facts across distinct data sources. Blindly applying classical query evaluation techniques to such databases leads to inconsistent answers. In the past, the database community has proposed a revolutionary way to view such information, namely as a set of possible databases, sometimes with an associated probability distribution. Query evaluation becomes a more refined task, in which query results are classified as *possible*, i.e. they belong to the answer over some possible database, or *certain*, i.e. they belong to the query answer over all possible databases. When the set of possible databases is accompanied by a probability distribution, the likelihood of possible answers can be derived. Not surprisingly, query evaluation in this setting is harder than in the standard relational setting, and work on finding the trade-offs between evaluation complexity and query language expressivity is always challenging. For Web data management, with its in-flux design for the data models and query languages, answering these questions is particularly timely.

Data Exchange is concerned with the (materialized or virtual) migration of data between data sources. Since in a Web setting such data sources are likely autonomous and have

distinct schemas even when modeling similar real-life concepts, it is proposed to specify declaratively how data from the source database relates to the data published into the target database. These specifications are known as *schema mappings*, and they are exploited for various tasks, ranging from actually migrating the data from source to target, to leaving the data at the source but migrating queries from the target schema to the source schema. Seminar 11421 gave particular attention to the case of data exchange for XML data (prior work confines itself mostly to relational data sources), and for incomplete data (prior work focuses solely on complete data). It also addressed the problem of inferring schema mappings from examples given by less sophisticated users, who simply associate source/target data pairs and expect a tool to automatically generate the mappings. The seminar included a tutorial by Phokion Kolaitis, co-founder of the data exchange field.

Distribution of data across sources (typically within peer-to-peer networks), as well as of the computation performed by queries, and more generally, processes on top of such data, is another prominent topic of Web data management. The seminar explored recent answers to the long-standing challenge of coming up with models of computation that enable expressive languages that are semantically clean, efficiently executable and nevertheless admit automatic optimization. The above mentioned, highly visible, European research project Webdam proposes a vision inspired by the quintessentially declarative Datalog language from classical relational database research. A notable related approach is motivated by the area of declarative networking, which has gained the attention of the systems community in past years, and more recently of the theory community, which is now carrying out foundational research to complement and enhance the existing systems contributions. Such models as the relational transducer networks are being proposed to formalize famous (but so far informally stated) conjectures about expressivity and evaluation complexity of declarative networking programs. The seminar was also interested in general questions on Peer-to-Peer networks.

Static verification of temporal properties is key to increasing the reliability and facilitating the design of various classes of processes powered by an underlying (collection of) databases. Notable examples include electronic commerce Web sites, declarative networking programs, and general business processes. In all these cases, the underlying data is dynamic, its evolution in time governed by large collections of declarative rules, whose interference with each other and global effect are impossible to predict without automatic verification tools. Of particular interest is the verification of properties pertaining to the temporal evolution of the system, which are naturally expressed in various temporal logic flavors.

Crowdsourcing is another highly relevant recent development in the Web data management arena, one in which practice has pressed ahead of foundational work, which is now attracting the interest of the theory community. The seminar dedicated particular attention to this topic, reserving a long talk slot for a survey.

Organization of the Seminar and Activities

The workshop brought together 51 researchers from complementary areas of database theory, logic, and theoretical computer science in general, all with an established record of excellence in Web data management. The participant pool comprised both senior and junior researchers, including several advanced PhD students.

Participants were invited to present their own work, and/or survey state-of-the-art advances and challenges in the field. Thirty-four talks were given, which included four (60-90 minute) tutorials and thirty regular (30 minute) talks. All presentations were scheduled

prior to the workshop, and due to the flood of volunteered talks, the organizers had to cap the number of slots. Talks were chosen so as to represent well the aspects of Web data management described above. The talks are listed below, classified by the covered topics. The classification is necessarily rough, as many talks crossed the boundaries between areas, in keeping with the seminar's intent. To the organizers' pleasant surprise, some of the results established surprising bridges between fields previously seen as unrelated (such as Machine Learning and Data Exchange), and brought in techniques from novel areas (such as Nominal Sets).

Crowdsourcing

- Tova Milo, Research Challenges in Crowdsourcing [tutorial]

Webdam Project Overview

- Emilien Antoine, Social Networking with WebdamExchange and WebdamLog
- Meghyn Bienvenu, A rule-based Language for Web Data Management

Web Data Management: Static Analysis

- Tom Ameloot, Relational Transducers for Declarative Networking
- Marie-Christine Rousset, Alignment-based Trust for Resource Finding in Semantic P2P Networks
- Alin Deutsch, Feasible Verification of Expressive Business Processes
- Anca Muscholl, Some Decidability Results on Distributed Games
- Evgeny Kharlamov, Evolution of DL-Lite Knowledge Bases
- Sophie Tison, Views and Updates

XPath with Data and Data Words

- Diego Figueira, XPath Decidability [tutorial]
- Mikołaj Bojańczyk, Temporal Logic on Changing XML Documents
- Szymon Toruńczyk, Automata-based Verification Over Linearly Ordered Data Domains
- Thomas Zeume, Two-Variable Logic, Orders and Successors

Probabilistic Data

- Pierre Senellart, PARIS: Probabilistic Alignment of Relations, Instances and Schemas
- Robert Fink, Aggregation in Probabilistic Databases via Knowledge Computation
- Serge Abiteboul, Finding Optimal Probabilistic Generators for XML Collections

FoXLib Project Overview

- Frank Neven, An Overview of FOXLIB
- Maarten Marx, Collections of XML, Schemas and Queries in FoxLib

SPARQL and Regular Expressions

- Wim Martens, The Complexity of Evaluating SPARQL Property Paths
- Giorgio Ghelli, Type-checking for SPARQL
- Juan Reutter, Parameterized Regular Expressions and their Languages

Schema Mappings and Data Exchange

- Phokion G. Kolaitis, Characterizing Schema Mappings with Examples [tutorial]
- Filip Murlak, XML Data Exchange
- Balder ten Cate, Learning Schema Mappings
- Marcelo Arenas, Data Exchange Beyond Complete Data

Incompleteness

- Leonid Libkin, A New Look at Incompleteness in Relations, XML, and Beyond [tutorial]

Nominal Sets and Access Paths

- Mikołaj Bojańczyk, Nominal Sets: An Introduction
- Sławomir Lasota, Nominal Sets: Automata
- Pierre Bourhis, Querying Access Paths

Queries

- Nicole Schweikardt, Expressiveness and Static Analysis of Extended Conjunctive Regular Path Queries
- Wojtek Kazana, Query Enumeration with Constant Delay
- Frank Neven, Deciding Twig-Definability of Node-Selecting Tree Automata
- Stijn Vansummeren, A New Characterization of the Acyclic Conjunctive Queries and Its Application to Structural Indexing
- Dan Olteanu, Factorized Representation of Query Results

Concluding Remarks and Future Plans

Due to the rich coverage of the area of foundations of Web data management, as achieved by both the presentations and the informal interactions, the organizers regard the seminar as a great success.

The weeklong format was well-suited to such an ambitious topic. The topic was well-received, as witnessed by the high rate of accepted invitations, and the exemplary degree of involvement by the participants. These volunteered such a high number of exceptional-quality talks that the organizers were faced with not being able to accommodate demand.

Bringing together researchers from different areas of data management, programming languages, theoretical computer science and logic fostered valuable interactions and led to fruitful collaborations, as reflected also by the very positive feedback from the audience.

The organizers wish to express their gratitude toward the Scientific Directorate of the Center for its support of this seminar, and hope to continue this seminar series on Web data management.

2 Table of Contents

Executive Summary

<i>Serge Abiteboul, Alin Deutsch, Thomas Schwentick, and Luc Segoufin</i>	37
---	----

Overview of Talks


Deduction in the Presence of Distribution, Contradictions and Uncertainty <i>Serge Abiteboul</i>	44
Deciding Eventual Consistency for a Simple Class of Relational Transducer Networks <i>Tom Ameloot</i>	44
Distributed Knowledge Base: Webdam System <i>Emilien Antoine</i>	45
A Rule-based Language for Web Data Management <i>Meghyn Bienvenu</i>	45
Efficient evaluation for a temporal logic on changing XML documents <i>Mikołaj Bojańczyk</i>	46
Nominal Sets <i>Mikołaj Bojańczyk</i>	46
Feasible Verification of Expressive Business Processes <i>Alin Deutsch</i>	46
Satisfiability for XPath <i>Diego Figueira</i>	47
Aggregation in Probabilistic Databases via Knowledge Compilation <i>Robert Fink</i>	47
Query Enumeration with Constant Delay <i>Wojtek Kazana</i>	48
Evolution of Knowledge Bases: DL-Lite case <i>Evgeny Kharlamov</i>	48
Schema Mappings and Data Examples <i>Phokion G. Kolaitis</i>	49
A new look at incompleteness in relations, XML, and beyond <i>Leonid Libkin</i>	49
The Complexity of Evaluating Path Expressions in SPARQL <i>Wim Martens</i>	50
Collections of XML, Schemas and Queries in FoXLib <i>Maarten Marx</i>	50
Asking the Right Questions in Crowd Data Sourcing <i>Tova Milo</i>	51
Solutions in XML data exchange <i>Filip Murlak</i>	51
Controlling distributed systems <i>Anca Muscholl</i>	52

Deciding Twig-definability of Node Selecting Tree Automata	
<i>Frank Neven</i>	52
Factorised Representations of Query Results	
<i>Dan Olteanu</i>	53
Alignment-based Trust for Resource Finding in Semantic P2P Networks	
<i>Marie-Christine Rousset</i>	53
Expressiveness and static analysis of extended conjunctive regular path queries	
<i>Nicole Schweikardt</i>	54
Finding Optimal Probabilistic Generators for XML Collections	
<i>Pierre Senellart</i>	54
Learning Schema Mappings	
<i>Balder Ten Cate</i>	55
Verification over linearly ordered data domains	
<i>Szymon Toruńczyk</i>	55
A new characterization of the acyclic conjunctive queries, and its application to structural indexing.	
<i>Stijn Vansummeren</i>	56
Two-Variable Logic, Orders and Successors	
<i>Thomas Zeume</i>	56
Participants	57

3 Overview of Talks

3.1 Deduction in the Presence of Distribution, Contradictions and Uncertainty


Serge Abiteboul (ENS – Cachan, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Serge Abiteboul

We study deduction, captured by datalog-style rules, in the presence of contradictions, captured by functional dependencies (FDs). We start with a simple semantics for datalog in the presence of functional dependencies that is based on inferring facts one at a time, never violating the FDs, until no further facts can be added. This is a non-deterministic semantics, that may lead to several possible worlds. We present a proof theory for this semantics and compare it to previous work on datalog with negation. We also discuss a set-at-a-time semantics, where in each iteration, all facts that can be inferred are added to the database, and then choices are made between contradicting facts. We then proceed to our main goal of defining a semantics for the distributed setting. Note that contradictions naturally arise in a distributed setting since different peers may have conflicting information, opinions or recommendations. In the distributed case, we propose and study a concrete semantics for (an important fragment of) a previously proposed distributed datalog idiom, namely Webdamlog, that we enrich to account for FDs. Here again, we compare the semantics with previously studied semantics and in particular Webdamlog with negation in the absence of FDs. Finally, we note that in a distributed environment, it is natural to settle contradictions by introducing probabilities. We consider a simple adaptation of the distributed semantics to a probabilistic setting and show that it captures an intuitive way of resolving contradictions. We propose a sampling algorithm for evaluating queries under this semantics.

3.2 Deciding Eventual Consistency for a Simple Class of Relational Transducer Networks

Tom Ameloot (Hasselt University, BE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Tom Ameloot

Joint work of Ameloot, Tom; Van den Bussche, Jan
Main reference T.J. Ameloot, J. Van den Bussche, “Deciding Eventual Consistency for a Simple Class of Relational Transducer Networks,” in Proceedings of the 15th International Conference on Database Theory, 2012, to appear.

Networks of relational transducers can serve as a formal model for declarative networking, focusing on distributed database querying applications. In declarative networking, a crucial property is eventual consistency, meaning that the final output does not depend on the message delays and reorderings caused by the network. Here, we show that eventual consistency is decidable when the transducers satisfy some syntactic restrictions, some of which have also been considered in earlier work on automated verification of relational transducers. This simple class of transducer networks computes exactly all distributed queries expressible by unions of conjunctive queries with negation.

3.3 Distributed Knowledge Base: Webdam System

Emilien Antoine (INRIA Saclay – Orsay, FR)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license
© Emilien Antoine

Joint work of Abiteboul, Serge; Antoine, Emilien; Bienvenu, Meghyn; Galland, Alban

Main reference S. Abiteboul, M. Bienvenu, A. Galland, E. Antoine, “A rule-based language for web data management,” Proceedings of the 30th ACM Symposium on Principles of Database Systems (PODS’11), pp. 293–304, Athens, Greece, 2011.

URL <http://hal.inria.fr/docs/00/58/28/91/PDF/pods17a-abiteboul.pdf>

URL <http://dx.doi.org/10.1145/1989284.1989320>

As an extension to the talk about WebdamLog which is a model for a distributed declarative database in a peer to peer environment, I present (1) the WebdamExchange system for access control management on top of WebdamLog and (2) some clues about the implementation of a WebdamLog engine. WebdamExchange provides also an architecture for communication, to deal with heterogeneity of peers on the Internet, introduced as Webdam for the poor.

References

- 1 Serge Abiteboul and Meghyn Bienvenu and Alban Galland and Emilien Antoine. *A rule-based language for web data management*. Proceedings of the Symposium on Principles of Database Systems, pp. 293–304, Athens, Greece, 2011.
- 2 Serge Abiteboul and Alban Galland and Neoklis Polyzotis. *A Model for Web Information Management with Access Control*. Proceedings of the International Workshop on the Web and Databases, Athens, Greece, 2011.

3.4 A Rule-based Language for Web Data Management

Meghyn Bienvenu (INRIA Saclay – Orsay, FR)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license
© Meghyn Bienvenu

Joint work of Abiteboul, Serge; Antoine, Emilien; Bienvenu, Meghyn; Galland, Alban

Main reference S. Abiteboul, M. Bienvenu, A. Galland, E. Antoine, “A rule-based language for web data management,” Proceedings of the 30th ACM Symposium on Principles of Database Systems (PODS’11), pp. 293–304, Athens, Greece, 2011.


URL <http://hal.inria.fr/docs/00/58/28/91/PDF/pods17a-abiteboul.pdf>

URL <http://dx.doi.org/10.1145/1989284.1989320>

There is a new trend to use Datalog-style rule-based languages to specify modern distributed applications, notably on the Web. In this talk, I will introduce such a language (called Webdamlog) for a distributed data model where peers exchange messages (i.e. logical facts) as well as rules. After illustrating the language, I will mention some results concerning the connection with centralized Datalog semantics and the impact on expressiveness of “delegations” (the installation of rules by a peer in some other peer) and explicit timestamps.

3.5 Efficient evaluation for a temporal logic on changing XML documents

Mikołaj Bojańczyk (University of Warsaw, PL)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Mikołaj Bojańczyk

Joint work of Bojańczyk, Mikołaj; Figueira, Diego

This talk is about a logic that describes a changing XML document. The changing XML document is modeled as a sequence of trees, over a finite alphabet. The logic can define properties such as: “every node changes its label at most twice”, or “whenever a node gets label a , then one of its descendants eventually gets label c ”. The contribution is an evaluation algorithm, which tests if a formula is true in a sequence of trees, assuming that the edit distance between consecutive trees is at most 1. The algorithm runs in time $n \log(k)$, where n is the number of trees and k is their maximal size.

3.6 Nominal Sets

Mikołaj Bojańczyk (University of Warsaw, PL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Mikołaj Bojańczyk

Joint work of Klin, Bartek; Bojańczyk, Mikołaj; Lasota, Sławomir

Nominal sets are a different kind of set theory. Nominal sets were invented by Abraham Fraenkel in the 1920's. They were rediscovered for computer science by Gabbay and Pitts in 1999, as a way of talking about name binding in lambda terms and logical formulas. We rediscover them yet again, this time as a way of talking about data values, including data words and automata for data words.

The point in nominal sets is that there is a different notion of finite set, e.g. the set of all data words of length at most 10 is a finite set.

3.7 Feasible Verification of Expressive Business Processes

Alin Deutsch (University of California – San Diego, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Alin Deutsch

Joint work of Deutsch, Alin; Damaggio, Elio; Vianu, Victor

We revisit the static verification problem for data centric business processes, specified in a variant of IBM's “business artifact” model. Artifacts are records of variables that correspond to business-relevant objects and are updated by a set of services equipped with pre-and postconditions, that implement business process tasks. The verification problem consists in statically checking whether all runs of an artifact system satisfy desirable properties expressed in a first order extension of linear-time temporal logic.


In previous work we identified the class of guarded artifact systems and properties, for which verification is decidable. However, the results suffer from an important limitation: they fail in the presence of even very simple data dependencies or arithmetic, both crucial to real-life business processes. In an ICDT 2011 paper, we extend the artifact model and

verification results to alleviate this limitation. We identify a practically significant class of business artifacts with data dependencies and arithmetic, for which verification is decidable, but the upper bound is non-elementary.

This talk reports on a new technique developed since, leading to more palatable upper bound (EXPSPACE). The technique makes practical implementation feasible, and a preliminary experimental evaluation of our prototype verifier yields encouraging results.

3.8 Satisfiability for XPath


Diego Figueira (University of Edinburgh, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Diego Figueira

XPath is a node selecting language for XML documents. Although its satisfiability problem is in general undecidable, there are several syntactic restrictions that make the problem decidable. I will present a survey on the existing results on the satisfiability problem for fragments of XPath in the presence of data values. I will also mention some open problems and conjectures.

3.9 Aggregation in Probabilistic Databases via Knowledge Compilation

Robert Fink (University of Oxford, GB)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Robert Fink
Joint work of Fink, Larisa; Han, Larisa; Olteanu, Dan
Main reference R. Fink, L. Han, D. Olteanu, “Aggregation in Probabilistic Databases via Knowledge Compilation,” VLDB 2012, pp. 490–501.

This talk presents a query evaluation technique for positive relational algebra queries with aggregates on a representation system for probabilistic data based on the algebraic structures of semirings and semimodule. The core of our evaluation technique is a procedure that compiles semimodule and semiring expressions into so-called decomposition trees, for which the computation of the probability distribution can be done in polynomial time in the size of the tree and of the distributions represented by its nodes. We give syntactic characterisations of tractable queries with aggregates by exploiting the connection between query tractability and polynomial-time compilation into decomposition trees.

The technique is incorporated into the probabilistic database engine SPROUT, which is built on top of PostgreSQL. We report on extensive performance experiments with synthetic datasets and TPC-H data.

3.10 Query Enumeration with Constant Delay

Wojtek Kazana (*ENS – Cachan, FR*)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Wojtek Kazana

In many applications the output of a query may have a huge size and enumerating all the answers may already consume too many of the allowed resources. In this case it may be appropriate to first output a small subset of the answers and then, on demand, output a subsequent small numbers of answers and so on until all possible answers have been exhausted. To make this even more attractive it is preferable to be able to minimize the time necessary to output the first answers and, from a given set of answers, also minimize the time necessary to output the next set of answers – this second time interval is known as the delay. For this it might be interesting to compute adequate index structures. The ultimate goal being to obtain index structures easily computable (say in linear time in the size of the database), that allow constant delay in the enumeration process. In this case we speak of constant delay enumeration of the query that was introduced by Durand and Grandjean.

In this talk I will outline the differences between query evaluation and enumeration. Using the example of MSO queries over trees I will try to illustrate some techniques useful in obtaining constant delay enumeration algorithms.

3.11 Evolution of Knowledge Bases: DL-Lite case

Evgeny Kharlamov (*Free University Bozen-Bolzano, IT*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Evgeny Kharlamov

Joint work of Calvanese, Diego; Kharlamov, Evgeny; Nutt, Werner; Zheleznyakov, Dmitriy

Main reference E. Kharlamov, D. Zheleznyakov, “Capturing Instance Level Ontology Evolution for DL-Lite,” in Proc. of ISWC, 2011.

URL <http://www.inf.unibz.it/~kharlamov/publications-date.html>

We study the problem of evolution for Knowledge Bases (KBs) expressed in Description Logics (DLs) of the DL-Lite family. DL-Lite is at the basis of OWL 2 QL, one of the tractable fragments of OWL 2, the recently proposed revision of the Web Ontology Language. We review known model (MBAs) and formula-based approaches (FBAs) for evolution of propositional theories. We exhibit limitations of MBAs: they intrinsically ignore the structural properties of KBs, which leads to undesired properties of KBs resulting from such an evolution, i.e., DL-Lite is not closed under all considered MBAs. We show what causes inexpressibility and exhibit a fragment of DL-Lite that is closed under a number of MBAs. We show that standard FBAs are also not appropriate for DL-Lite evolution, either due to high complexity of computation, or because the result of such an action of evolution is not expressible in DL-Lite. We propose two formula-based approaches for which evolution is expressible in DL-Lite and can be computed in polynomial time.

The talk is based on the following papers [1, 2].

References

- 1 Diego Calvanese, Evgeny Kharlamov, Werner Nutt, and Dmitriy Zheleznyakov. *Evolution of DL-Lite Knowledge Bases*. In Proc. of ISWC, 2010
- 2 Evgeny Kharlamov and Dmitriy Zheleznyakov. *Capturing Instance Level Ontology Evolution for DL-Lite*. In Proc. of ISWC, 2011

3.12 Schema Mappings and Data Examples

Phokion G. Kolaitis (University of California – Santa Cruz, US)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license
© Phokion G. Kolaitis

Joint work of Alexe, Bogdan; ten Cate, Balder; Kolaitis, Phokion G.; Tan, Wang-Chiew

Schema mappings are high-level specifications that describe the relationship between two database schemas. Schema mappings are considered to be the essential building blocks in such critical data interoperability tasks as data exchange and data integration. For this reason, they have been the focus of extensive research investigations over the past several years. Since in real-life applications schema mappings can be quite complex, it is important to develop methods and tools for illustrating, explaining, and deriving schema mappings. A promising approach to this effect is to use “good” data examples that illustrate the schema mapping at hand.

In this talk, we present an overview of recent work on characterizing and deriving schema mappings via a finite set of data examples. We show that every LAV schema mapping (i.e., a schema mapping specified by a finite set of local-as-view tuple-generating dependencies) is uniquely characterized by a finite set of universal data examples with respect to the class of all LAV schema mappings. We also show that this type of result does not hold for arbitrary GAV schema mappings (i.e., schema mappings specified by a finite set of global-as-view tuple-generating dependencies). After this, we give a necessary and sufficient algorithmic condition for a GAV schema mapping to be uniquely characterizable by a finite set of universal examples with respect to the class of all GAV schema mappings. Along the way, we establish tight connections between unique characterizability of schema mappings and homomorphism dualities.

This is joint work with Bogdan Alexe (IBM Research – Almaden), Balder ten Cate (UC Santa Cruz), and Wang-Chiew Tan (UC Santa Cruz and IBM Research – Almaden).

3.13 A new look at incompleteness in relations, XML, and beyond

Leonid Libkin (University of Edinburgh, GB)

License © ⓘ ⊕ Creative Commons BY-NC-ND 3.0 Unported license
© Leonid Libkin


While incomplete information is ubiquitous in all data models - especially in applications involving data translation or integration – our understanding of it is still not completely satisfactory. For example, even such a basic notion as certain answers for XML queries was only introduced recently, and in a way seemingly rather different from relational certain answers.

Here we propose a general approach to handling incompleteness, and test its applicability in known data models such as relations and documents. The approach is based on representing degrees of incompleteness via semantics-based orderings on database objects. We use it to both obtain new results on incompleteness and to explain some previously observed phenomena. Specifically we show that certain answers for relational and XML queries are two instances of the same general concept; we describe structural properties behind the naive evaluation of queries; answer open questions on the existence of certain answers in the XML setting; and show that previously studied ordering-based approaches were only adequate

for SQL’s primitive view of nulls. We define a general setting that subsumes relations and documents to help us explain in a uniform way how to compute certain answers, and when good solutions can be found in data exchange. We also look at the complexity of common problems related to incompleteness, and generalize several results from relational and XML contexts.

3.14 The Complexity of Evaluating Path Expressions in SPARQL

Wim Martens (*Universität Bayreuth, DE*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Wim Martens

The World Wide Web Consortium (W3C) recently included property paths in the working draft for SPARQL 1.1, a query language for RDF data. Property paths give SPARQL queries the power of evaluating regular expressions over graph data. However, they differ from regular expressions in several notable aspects.

For example, they include a limited form of negation and they can use counters as syntactic sugar. Furthermore, their semantics on graphs is defined in a non-standard manner.

We formalize the W3C semantics of property paths and investigate the impact on the complexity of various query evaluation problems on graphs. More specifically, let x and y be two nodes in an edge-labeled graph and r be a regular expression. We investigate the complexities of (1) deciding whether there exists a path from x to y that matches r and (2) counting how many paths from x to y match r .

Our main results show that, compared to an alternative semantics of regular expressions on graphs, the W3C semantics causes a significant increase in the complexity of problems (1) and (2). Whereas the alternative semantics remains in polynomial time for fairly large fragments of expressions, the W3C semantics makes problems (1) and (2) intractable almost immediately.

As a side-result, we also prove that the membership problem for regular expressions with counters and negation is in polynomial time.

3.15 Collections of XML, Schemas and Queries in FoXLib

Maarten Marx (*University of Amsterdam, NL*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Maarten Marx

Joint work of Grijzenhout, Steven; Marx, Maarten

Main reference S. Grijzenhout, M. Marx, “The Quality of the XML web,” in Proc. of 20th ACM Int’l Conf. on Information and Knowledge Management (CIKM’11), pp. 1719–1724, 2011.

URL <http://dx.doi.org/10.1145/2063576.2063824>

We collect evidence to answer the following question: Is the quality of the XML documents found on the web sufficient to apply XML technology like XQuery, XPath and XSLT? XML collections from the web have been previously studied statistically, but no detailed information about the quality of the XML documents on the web is available to date. We address this shortcoming in this study. We gathered 180K XML documents from the web. Their quality is surprisingly good; 85.4% is well-formed and 99.5% of all specified encodings is correct. Validity needs serious attention. Only 25% of all files contain a reference to a DTD

or XSD, of which just one third is actually valid. Errors are studied in detail. Automatic error repair seems promising. Our study is well documented and easily repeatable. This paves the way for a periodic quality assessment of the XML web.

All data is publicly available at the url <http://data.politicalmashup.nl/xmlweb>.

3.16 Asking the Right Questions in Crowd Data Sourcing

Tova Milo (Tel Aviv University, IL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Tova Milo

Crowd-based data sourcing is a new and powerful data procurement paradigm that engages Web users to collectively contribute data, analyze information and share opinions. This brings to light, out of the huge, inconsistent Web ocean, an important body of knowledge that would otherwise not be attainable. Crowd-based data sourcing democratizes data-collection, cutting companies' and researchers' reliance on stagnant, overused datasets and bears great potential for revolutionizing our information world. Yet, triumph has so far been limited to only a handful of successful projects such as Wikipedia or IMDb. This comes notably from the difficulty of managing huge volumes of data and users of questionable quality and reliability. Every single initiative had to battle, almost from scratch, the same non-trivial challenges. The ad hoc solutions, even when successful, are application specific and rarely sharable.

In this talk we consider the development of solid scientific foundations for Web-scale data sourcing. We believe that such a principled approach is essential to obtain knowledge of superior quality, to realize the task more effectively and automatically, be able to reuse solutions, and thereby to accelerate the pace of practical adoption of this new technology that is revolutionizing our life.

We discuss the desired logical, algorithmic, and methodological foundations for the management of large scale crowd-sourced data and for the development of applications over such information. This encompasses formal models capturing all the diverse facets of crowd-sourced data. This also means developing the necessary reasoning capabilities for managing and controlling data sourcing, cleaning, verification, integration, sharing, querying and updating, in a dynamic Web environment.

3.17 Solutions in XML data exchange

Filip Murlak (University of Warsaw, PL)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Filip Murlak


Joint work of Bojańczyk, Mikołaj; Kołodziejczyk, Leszek A.; Murlak, Filip

The task of XML data exchange is to restructure a document conforming to a source schema under a target schema according to certain mapping rules. The rules are typically expressed as source-to-target dependencies using various kinds of patterns, involving horizontal and vertical navigation, as well as data comparisons. The target schema imposes complex conditions on the structure of solutions, possibly inconsistent with the mapping rules. In consequence, for some source documents there may be no solutions. I will discuss three

computational problems: deciding if all documents of the source schema can be mapped to a document of the target schema (absolute consistency), deciding if a given document of the source schema can be mapped (solution existence), and constructing a solution for a given source document (solution building). It turns out that the complexity of absolute consistency is rather high in general, but within the polynomial hierarchy for bounded depth schemas. The combined complexity of solution existence and solution building behaves similarly, but the data complexity is very low. In fact, even for very expressive mapping rules, based on MSO definable queries, absolute consistency is decidable and data complexity of solution existence is polynomial.

3.18 Controlling distributed systems

Anca Muscholl (*Université Bordeaux, FR*)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Anca Muscholl

Joint work of Genest, B.; Gimbert, H.; Muscholl, Anca; Walukiewicz, I.

We consider the problem of controlling distributed automata that cooperate via shared variables (*rendez-vous*). The setting corresponds to the framework of Ramadge and Wonham, where certain actions (controllable ones) can be forbidden by the local controller. Although the general question is still open, we can show that the problem is decidable on acyclic architectures, albeit of non-elementary complexity.

3.19 Deciding Twig-definability of Node Selecting Tree Automata

Frank Neven (*Hasselt University, BE*)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Frank Neven

Node selecting tree automata (NSTAs) constitute a general formalism defining unary queries over trees.

Basically, a node is selected by an NSTA when it is visited in a selecting state during an accepting run.

We consider twig patterns as an abstraction of XPath. Since the queries definable by NSTAs form a strict superset of twig-definable queries, we study the complexity of the problem to decide whether the query by a given NSTA is twig-definable. In particular, we obtain that the latter problem is EXPTIME-complete.

In addition, we show that it is also EXPTIME-complete to decide whether the query by a given NSTA is definable by a node selecting string automaton.

3.20 Factorised Representations of Query Results

Dan Olteanu (University of Oxford, GB)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
 © Dan Olteanu
Joint work of Olteanu, Dan; Zavodny, Jakub
Main reference D. Olteanu, J. Zavodny, “Factorised Representations of Query Results,” arXiv:1104.0867v1 [cs.DB],
 to appear in ICDT 2012.
URL <http://arxiv.org/abs/1104.0867v1>

We introduce a representation system for relational data based on algebraic factorisation using distributivity of product over union and commutativity of product and union.

We give two characterisations of conjunctive queries based on factorisations of their results defined by a certain class of hyperpath decompositions of the query hypergraph.

The first characterisation concerns sizes of factorised representations. For any query, we derive a size bound that is asymptotically tight within our class of factorisations.

For relations where tuples are annotated with identifiers we also characterise the queries by the readability of their results, which is the minimum over all equivalent factorisations of the maximum number of occurrences of any identifier in that factorisation. We give a dichotomy of queries based on the readability of their results for any database and define syntactically the class of queries with bounded readability.

3.21 Alignment-based Trust for Resource Finding in Semantic P2P Networks

Marie-Christine Rousset (Université de Grenoble, FR)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
 © Marie-Christine Rousset
Joint work of Atencia, Manuel; Euzenat, Jérôme; Pirrò, Giuseppe; Rousset, Marie-Christine
Main reference M. Atencia, J. Euzenat, G. Pirrò, M.-C. Rousset, “Alignment-Based Trust for Resource Finding in Semantic P2P Networks,” in Proc. of Int’l Semantic Web Conference (ISWC’11), pp. 51–66, 2011.
URL http://dx.doi.org/10.1007/978-3-642-25073-6_4


In a semantic P2P network, peers use separate ontologies and rely on alignments between their ontologies for translating queries.

Nonetheless, alignments may be limited (unsound or incomplete) and generate flawed translations, leading to unsatisfactory answers. In this paper we present a trust mechanism that can assist peers to select those in the network that are better suited to answer their queries. The trust that a peer has towards another peer depends on a specific query and represents the probability that the latter peer will provide a satisfactory answer. In order to compute trust, we exploit both alignments and peers’ direct experience, and perform Bayesian inference. We have implemented our technique and conducted an evaluation. Experimental results showed that trust values converge as more queries are sent and answers received.

Furthermore, the use of trust improves both precision and recall.

3.22 Expressiveness and static analysis of extended conjunctive regular path queries

Nicole Schweikardt (Goethe-Universität Frankfurt am Main, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Nicole Schweikardt

Joint work of Freydenberger, Dominik; Schweikardt, Nicole

Main reference D. Freydenberger and N. Schweikardt, “Expressiveness and static analysis of extended conjunctive regular path queries,” in Proc. of the 5th Alberto Mendelzon Int’l Workshop on Foundations of Data Management (AMW’11), vol. 749 of CEUR Workshop Proceedings, CEUR-WS.org, 2011.

URL <http://ceur-ws.org/Vol-749/paper9.pdf>

We study the expressiveness and the complexity of static analysis of extended conjunctive regular path queries (ECRPQs), introduced by Barcelo et al. (PODS’10). ECRPQs are an extension of conjunctive regular path queries (CRPQs), a well-studied language for querying graph structured databases. Our first main result shows that query containment and equivalence of a CRPQ in an ECRPQ is undecidable. This settles one of the main open problems posed by Barcelo et al.


As a second main result, we prove a non-recursive succinctness gap between CRPQs and the CRPQ-expressible fragment of ECRPQs. Apart from this, we develop a tool for proving inexpressibility results for CRPQs and ECRPQs. In particular, this enables us to show that there exist queries definable by regular expressions with backreferencing, but not expressible by ECRPQs.

This is joint work with Dominik D. Freydenberger.

The material presented in this talk was published in the proceedings of the 5th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2011), vol. 749 of CEUR Workshop Proceedings, CEUR-WS.org, 2011.

3.23 Finding Optimal Probabilistic Generators for XML Collections

Pierre Senellart (Telecom Paris Tech, FR)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Pierre Senellart

Joint work of Abiteboul, Serge; Amsterdamer, Yael; Deutch, Daniel; Milo, Tova; Senellart, Pierre

Main reference S. Abiteboul, Y. Amsterdamer, D. Deutch, T. Milo, P. Senellart, “Finding Optimal Probabilistic Generators for XML Collections,” in Proc. ICDT, Berlin, Germany, March 2012.

URL <http://pierre.senellart.com/publications/abiteboul2012finding.pdf>

We study the problem of, given a corpus of XML documents and its schema, finding an optimal (generative) probabilistic model, where optimality here means maximizing the likelihood of the particular corpus to be generated. Focusing first on the structure of documents, we present an efficient algorithm for finding the best generative probabilistic model, in the absence of constraints. We further study the problem in the presence of integrity constraints, namely key, inclusion, and domain constraints. We study in this case two different kinds of generators. First, we consider a continuation-test generator that performs, while generating documents, tests of schema satisfiability; these tests prevent from generating a document violating the constraints but, as we will see, they are computationally expensive. We also study a restart generator that may generate an invalid document and, when this is the case, restarts and tries again. Finally, we consider the injection of data values into the structure, to obtain a full XML document. We study different approaches for generating these values.

3.24 Learning Schema Mappings

Balder Ten Cate (University of California – Santa Cruz, US)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Balder Ten Cate

Joint work of Kolaitis, Phokion G.; Dalmau, Victor; Cate, Ten Cate, Balder

Main reference B. Ten Cate, V. Dalmau, Ph. Kolaitis, “Learning Schema Mappings,” in Proc. of ICDT 2012, to appear.

A schema mapping is a high-level specification of the relationship between a source schema and a target schema. Recently, a line of research has emerged that aims at deriving schema mappings automatically or semi-automatically with the help of data examples, i.e., pairs consisting of a source instance and a target instance that depict, in some precise sense, the intended behavior of the schema mapping. Several different uses of data examples for deriving, refining, or illustrating a schema mapping have already been proposed and studied.

In this paper, we use the lens of computational learning theory to systematically investigate the problem of obtaining algorithmically a schema mapping from data examples. Our aim is to leverage the rich body of work on learning theory in order to develop a framework for exploring the power and the limitations of the various algorithmic methods for obtaining schema mappings from data examples. We focus on GAV schema mappings, that is, schema mappings specified by GAV (Global-As-View) constraints. GAV constraints are the most basic and the most widely supported language for specifying schema mappings.

We present an efficient algorithm for learning GAV schema mappings using Angluin’s model of exact learning with membership and equivalence queries. This is optimal, since we show that neither membership queries nor equivalence queries suffice, unless the source schema consists of unary relations only. We also obtain results concerning the learnability of schema mappings in the context of Valiant’s well known PAC (Probably-Approximately-Correct) learning model.

Finally, as a byproduct of our work, we show that there is no efficient algorithm for approximating the shortest GAV schema mapping fitting a given set of examples, unless the source schema consists of unary relations only.

3.25 Verification over linearly ordered data domains

Szymon Toruńczyk (ENS – Cachan, FR)

License © © © Creative Commons BY-NC-ND 3.0 Unported license
© Szymon Toruńczyk

Joint work of Segoufin, Luc; Toruńczyk, Szymon

We work over linearly ordered data domains equipped with finitely many unary predicates and constants. We consider nondeterministic automata processing words and storing finitely many variables ranging over the domain. During a transition, these automata can compare the data values of the current configuration with those of the previous configuration using the linear order, the unary predicates and the constants.

We show that emptiness for such automata is decidable, both over finite and infinite words, under reasonable computability assumptions on the linear order. Finally, we show how our automata model can be used for verifying properties of work-flow specifications in the presence of an underlying database.

3.26 A new characterization of the acyclic conjunctive queries, and its application to structural indexing.

Stijn Vansummeren (Université Libre de Bruxelles, BE)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Stijn Vansummeren

We present a new structural characterization of the expressive power of the acyclic conjunctive queries in terms of guarded simulations. The study of this fragment of first order logic is motivated by the central role it plays in query languages across a wide range of data models.

We discuss the relevance of this result as a formal basis for constructing so-called structural indexes. Structural indexes were first proposed in the context of semi-structured query languages and later successfully applied as an XML indexation mechanism for XPath-like queries. We discuss how our main result can be instantiated to the construction of structural indexes for RDF on the Semantic Web.

3.27 Two-Variable Logic, Orders and Successors

Thomas Zeume (TU Dortmund, DE)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Thomas Zeume

Joint work of Manuel, Amaldev; Schwentick, Thomas; Zeume, Thomas

Recent results for the finite satisfiability problem for two-variable logic over structures with linear order, preorder, successor and unary relations will be discussed in this talk.

Two-variable logic with one total preorder relation, its induced successor relation, one linear order relation and some further unary relations is EXPSPACE-complete. Actually, EXPSPACE-completeness already holds for structures that do not include the induced successor relation. As a special case, the EXPSPACE upper bound applies to two-variable logic over structures with two linear orders. A further consequence is that satisfiability of two-variable logic over data words with a linear order on positions and a linear order and successor relation on the data is decidable in EXPSPACE. Furthermore, two-variable logic is decidable on structures with two linear order successors and an order corresponding to one of the successors.

Those results are complemented by the undecidability of the finite satisfiability problem for two-variable logic over structures with two total preorder relations as well as over structures with one total preorder and two linear order relations.

Participants

- Serge Abiteboul
ENS – Cachan, FR
- Tom Ameloot
Hasselt University, BE
- Emilien Antoine
INRIA Saclay – Orsay, FR
- Timos Antonopoulos
Hasselt University, BE
- Marcelo Arenas
Univ. Católica de Chile, CL
- Pablo Barcelo
Univ. of Chile – Santiago, CL
- Meghyn Bienvenu
INRIA Saclay – Orsay, FR
- Mikołaj Bojańczyk
University of Warsaw, PL
- Pierre Bourhis
University of Oxford, UK
- Claire David
Université Paris-Est –
Marne-la-Vallée, FR
- Alin Deutsch
University of California – San
Diego, US
- Diego Figueira
University of Edinburgh, UK
- Robert Fink
University of Oxford, UK
- Amélie Gheerbrant
University of Edinburgh, UK
- Giorgio Ghelli
University of Pisa, IT
- Florent Jacquemard
ENS – Cachan, FR
- Ahmet Kara
TU Dortmund, DE
- Wojtek Kazana
ENS – Cachan, FR
- Evgeny Kharlamov
Free Univ. Bozen-Bolzano, IT
- Pekka Kilpeläinen
University of Kuopio, FI
- Christoph Koch
EPFL – Lausanne, CH
- Phokion G. Kolaitis
University of California – Santa
Cruz, US
- Sławomir Lasota
University of Warsaw, PL
- Leonid Libkin
University of Edinburgh, UK
- Sebastian Maneth
Univ. of New South Wales, AU
- Wim Martens
Universität Bayreuth, DE
- Maarten Marx
University of Amsterdam, NL
- Tova Milo
Tel Aviv University, IL
- Filip Murlak
University of Warsaw, PL
- Anca Muscholl
Université Bordeaux, FR
- Frank Neven
Hasselt University, BE
- Matthias Niewerth
TU Dortmund, GE
- Dan Olteanu
University of Oxford, UK
- Pawel Parys
University of Warsaw, PL
- Juan L. Reutter
University of Edinburgh, UK
- Marie-Christine Rousset
Université de Grenoble, FR
- Anne Schuth
University of Amsterdam, NL
- Nicole Schweikardt
Goethe-Universität Frankfurt am
Main, DE
- Thomas Schwentick
TU Dortmund, DE
- Luc Segoufin
ENS – Cachan, FR
- Helmut Seidl
TU München, DE
- Pierre Senellart
Telecom Paris Tech, FR
- Cristina Sirangelo
ENS – Cachan, FR
- Tony Tan
University of Edinburgh, UK
- Balder Ten Cate
University of California – Santa
Cruz, US
- Sophie Tison
Université de Lille I, FR
- Szymon Toruńczyk
ENS – Cachan, FR
- Jan Van den Bussche
Hasselt University, BE
- Stijn Vansummen
Université Libre de Bruxelles, BE
- Domagoj Vrgoc
University of Edinburgh, UK
- Thomas Zeume
TU Dortmund, DE

