DAGSTUHL
REPORTS

Volume 2, Issue 5, May 2012

Dagstuhl Reports, Vol. 2, Issue 5 **ISSN 2192-5283**

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
September, 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`.

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

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

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

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

Digital Object Identifier: 10.4230/DagRep.2.5.i

Report from Dagstuhl Seminar 12181

Quality of Experience: From User Perception to Instrumental Metrics

Edited by

Markus Fiedler[1], Sebastian Möller[2], and Peter Reichl[3]

1 **Blekinge Institute of Technology – Karlskrona, SE,** markus.fiedler@bth.se
2 **Deutsche Telekom Laboratories, DE**
3 **FZ Telekommunikation Wien, AT,** reichl@ftw.at

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 12181 "Quality of Experience: From User Perception to Instrumental Metrics". As follow-up of the Dagstuhl Seminar 09192 "From Quality of Service to Quality of Experience", it focused on the further development of an agreed definition of the term Quality of Experience (QoE) in collaboration with the COST Action IC1003 "Qualinet", as well as inventories of possibilities to measure QoE (beyond the usual user polls) and to exploit feedback between users and systems that reflects QoE issues. The report furthermore describes the mode of work throughout the seminar, with focus on personal statements by the participants, results of the group works, and open challenges.

Seminar 01.–04.May, 2012 – www.dagstuhl.de/12181
1998 ACM Subject Classification H.5.1 Multimedia Information Systems, H.5.2 User Interfaces
Keywords and phrases Quality of Experience, usability, content, network monitoring, quality measurement, service pricing
Digital Object Identifier 10.4230/DagRep.2.5.1

1 Executive Summary

Markus Fiedler,
Sebastian Möller
Peter Reichl

During the recent years, Quality of Experience (QoE) has established itself as a topic of its own for both industrial and academic research. With its focus on the user in terms of acceptability, delight and performance, it is about to take over the role of Quality of Service as key paradigm for provisioning and managing services and networks. As one of the follow-up activities of the Dagstuhl Seminar 09192 "From Quality of Service to Quality of Experience", this Dagstuhl Seminar 12181 focused on the relation between quality perception and QoE quantification, which is among the most challenging tasks for bringing together the three essential corner stones, i.e. user, technology, and business. In particular, qualitative user perception needs to be translated into quantitative input to dimensioning and control of networks and services. Further, different kinds of feedback flows (acceptance, usage, cost, quality) need to be taken into account. Considering the multidisciplinary nature of this problem with complementary and potentially controversial views, the seminar worked towards metrics and measurement techniques aimed at improving QoE prediction and control. The outcomes are expected to become visible in the future QoE research agenda and corresponding standardisation efforts.

2 Table of Contents

3 Introduction

3.1 Introduction

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

Dagstuhl seminars strongly depend on the delegates and their input. In order to give room for both presentations and group discussions during a three-day seminar, the presentations were confined to five minutes and one slide. Each presentation was followed by a short block of questions and answers. In order to truly reflect the delegates' positions with regards to the topic of the seminar, the abstracts have been included in the sequel as-is and in alphabetical order.

The presentation round was followed by the presentation of a QoE White Paper around a QoE definition that has emanated from the Dagstuhl Seminar 09192.

The related discussions of QoE-related definitions and notions were continued and deepened during the first group work entitled "Key aspects of experience perception and their subjective evaluation". The other two groups discussed "Measurable aspects of QoE" and "Identification of QoE-related feedback loops". The outcomes of the group works were presented and discussed in the plenum, and excerpts are presented below.

The seminar was concluded with a plenary discussion of follow-up activities.

4 Overview of Talks

4.1 Combination of multi-source observations at the sub-second scale

Patrik Arlos (Blekinge Institute of Technology – Karlskrona, SE)

Synchronizing measurements from different systems; network, services (app, supporting app, etc...), and human. There is very little usage knowing that at 09:35:01 the user reported/signaled a problem, when there were problems tagged by the network at 09:25:10, 09:35:10 and 09:36:10. At the same range, the services were reporting problems at 09:20:18 and 08:45:10. Then to add to this, how do you add observations from cameras, microphones, EEK etc. . . .

4.2 On-line Estimation of the Quality of Experience

Åke Arvidsson (Ericsson Research – Stockholm, SE)

In cellular networks, the fraction of data traffic surpassed the one of voice traffic in Q4 2009, and by Q1 2012 it has grown to become on the order of three times larger.

The air interface is a limited resource subject to variable demand and variable capacity and most of this often applies to the backhaul as well. This means that cellular networks are likely to encounter temporary bursts of congestion in which case operators have to make intelligent decisions about priorities. In doing so, maintaining QoE should be an important aspect.

The notion of QoE, traditionally measured as MOS, does, however, involve a range of aspects such as, e.g.,

- the terminal (screen, buttons etc.),
- the application (functionality, design etc.),
- the encoding (image resolution, audio fidelity etc.),
- the content (degree of interest, production quality etc.) and
- the presentation (disturbances due to loss, delay etc.).

Since only the last point is directly applicable to managing bursty congestion, it is difficult to rely on MOS in this context. A further complication is that applications and expectations are diverse and subject to constant change whereas MOS measurements are limited and time consuming. Moreover, the special setting around any noticable attempt to estimate MOS may introduce a bias.

For this reason we are interested in the part of QoE that relates to loss and delay only, and to measure it by observing live traffic. To this end we suggest using a binary QoE metric, acceptable or unacceptable.

For example, acceptable web response times are those which do not prompt users to abort their requests by clicking on stop or other links and we note that both response times and user abortions can be measured in the network. After accounting for the fact that not all abortions are related to response times, we can thus assess the "acceptability" of a certain response time as the fraction of users that finds it acceptable. Congestion control may then be tuned to maximise acceptability and operators may set targets for this value. (An interesting remark is that such a target in many aspects is similar to the classical grade-of-service target in circuit switched networks.)

4.3 A User-Centric Service Modeling approach for User Experience Assessment

Sergio Beker (Huawei Technologies – München, DE)

In the context of ever reducing profit margins, operators are turning to Customer Experience Management (CEM) as a strong market differentiator. Current CEM approaches are based on Key Quality Indicators (KQIs) representation of the service performance. Although claiming to be user- oriented, the service modeling frameworks such as eTOM (TMF) and ITIL keep a per-service view, and as such, they are still network-centric. User Experience Indicators (UXIs) would be better suited to represent the user experience, and Customer Experience Indicators (CEIs) to represent the customer outcome of it. The traditional approach to assess UX is to run subjective tests and then to correlate the user answers to the technical measurable aspects of the service. The scope of application for such techniques is narrow around the original context of the test, results in high costs for the operator and proves invasive for the user. In true operational or commercial contexts, a modeling approach would be preferred. Also, the capability of modeling user experience from network observation alone is a key challenge. The User-Centric Service Modeling (USM) approach models the user experience by taking into account the user interaction with the service. This user centered view allows to include the different aspects of the service and the usage context in estimating the per-user- per-service-per-session User Experience, and to derive the corresponding User and Customer Experience Indicators. Also, by following the customer through the daily interaction with the service, the service-related and non- service related aspects, as well as the user experience with time can be modelled. This pioneering framework has been awarded three patents and is setting the industry standards at ETSI. A platform integrating the USM concepts is under development at Huawei's European Research Center in Munich.

4.4 Are engineers from Mars and users from Venus? QoE measurement as an interdisciplinary process

Katrien R. De Moor (Gent University, BE)

From a network engineering and application provision point of view, there is a huge need for unambiguous information: hard numbers, objective data, control. At the other end of the chain are the technology users, who 'as human beings' can be predictable in some aspects, but who can also be highly ambiguous and unpredictable in others. In a user's mind, an experience is not captured or recorded in numerical ratings, it is recorded in thoughts, feelings, stories, images, ... Bridging these differences in 'language' is in my opinion not 'mission impossible', but it requires an interdisciplinary approach, it requires 'translators' (e.g., psychologists, cognitive scientists, social scientists, . . .).

As previous research has already shown, (the quality of) users' experiences are dynamic and variable, influenced by a wide range of both human and technical factors and strongly bounded by multilayered contextual aspects. Moreover, there is no such thing as 'the user'. A person can e.g., be a hardcore online gamer and at the same time an absolute laggard when it comes to mobile services. How to deal with this complexity? There are different levels and stages related to QoE measurement, which in my opinion need a tailored instrumentation. Different steps in this approach could be:

1. Identification of relevant, measurable QoE features (cfr. Qualinet white paper definition: perceivable, recognized and nameable characteristic of the individual's experience of a service which contributes to its quality), as well as possible influencing factors for a specific service or set of services. E.g., seeking to understand the context in which an application is used, understanding user behavior for specific service with specific affordances, . . . ; In terms of instrumentation, different types of input (objective and quantitative as well as subjective and qualitative) and measurements are needed.
2. Isolation of specific influencing factors to investigate if and if yes, how they interplay and may influence users' QoE and the features it is composed of. Which patterns can be detected? This typically requires measurements in controlled environments or settings that are at least to some degree 'controllable'.
3. Extrapolation to more realistic user environments (less control, but higher ecological validity). Understanding QoE from a user perspective requires understanding QoE in the user's natural 'habitat'. This is especially crucial in the case of e.g., mobile applications and services.

4. QoE measurement is not a goal in itself, it should be the basis for action: insights and findings should be translated into actionable input for the different stakeholders in the QoE ecosystem (e.g., operators, network providers, ...)

About myself/how I got to the QoE topic: With a background in social sciences, rolling into the field of QoE some years ago was literally a bit of a 'culture' shock. The first QoE-project I worked on was a very strange experience. A bit comparable to arriving in an completely unknown country, in which people speak a language you don't understand, which uses conventions and signs you are not that familiar with. At first there is a huge barrier to meaningful interaction and you might start thinking maybe I should just go back home, to what I am familiar with. But slowly but surely I tried to learn some of the basics of the language and started to realize that some bits and pieces of knowledge in my social science backpack might be somehow relevant after all. Ultimately, QoE is and should be about people. Although the field is already undergoing a major shift to really start from the user's perception, a major challenge is still related to (1) meaningful interaction: between users, operators and providers, researchers from different backgrounds and with a different expertise; and (2) translating insights arising from that interaction into 'actionable' and tangible input at different levels.

4.5 A network based method for Video QoE measurement

Marcus Eckert (TU Chemnitz, DE)

A new method for monitoring the quality of Internet video streaming is presented. The method fully relies on network-centric measurements: it determines the buffered play-out time during progressive download by the evaluation of TCP segment timings in measurement traces. Results from tests carried out in a real mobile network show good agreement with user perceived quality measured directly at the end device. The performance of the method is compared to results of other approaches.

4.6 Waiting for QoE

Sebastian Egger (FZ Telekommunikation Wien, AT)

Recent work on QoE sets out to identify natural psychophysical relationships between the network (stimulus) and user perception (response), with network- induced waiting time being a specific example of such a stimulus which directly affect user satisfaction and thus QoE. Especially in the context of interactive data services, QoE is determined by such waiting times to a large extent, a fact which has led to the catchy notion of WWW as World Wide Wait A large share of services e.g. file downloads, E-Mail browsing, picture viewing or basic web browsing is characterized by an information request from user side and respective waiting times until the request is fulfilled. The past shift from UDP media streaming to TCP media streaming (e.g. youtube.com) has extended the relevance of waiting times also to the domain of online video services. Therefore the following questions are of particular importance to the QoE community: Which waiting times are sufficient to ensure a certain degree of user satisfaction? Are the waiting times translatable between different services?

4.7 Quality of Experience instrumentation: Read the user AND the network

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

My background is teletraffic modelling and analysis in communication networks, which means that I entered the QoE domain from the network measurement, modelling and analysis point-of-view. Triggered by Swedish industry to look for indicators of bad user performance that might cause user churn, our first QoE-related Quality of Service (QoS) study was a bottleneck indicator [1].

During the last decade, QoE has taken over the role that QoS was supposed to have, with focus on multimedia applications and spatial distortions. During classical QoE studies, users have to sit still and rate QoE under very well defined circumstances – this might be repeatable, but also far from real-life situations.

We realise that presence or lack of quality is affecting user's behaviour towards systems, meaning that the way a user interacts with a system mirrors performance and QoE issues. User reactions to application service performance as seen from two-way measurements in user interfaces have a lot to tell and should be exploited to a much larger extent than what is the case today. Furthermore, recent results reveal links between energy consumption patterns and QoE issues (cf. Selim Ickin's contribution). On the other hand, temporal distortions in data flows are visible within the network and can be detected without even bothering the end users [2]. Though having been claimed for many years, efficient feedback channels between users and providers/operators are still missing [3].

Interestingly enough, classical teletraffic modelling and analysis did not follow the above-mentioned shift from QoS to QoE, although they are well prepared to capture (so-far untended) temporal QoE issues. Based on experience with the search for user-perceived bottlenecks through Comparative Output/Input Analysis (COIA) [4] and with recent modelling efforts of mobile connectivity, both based on an underlying fluid model, we postulate the potential of queuing models to quantitatively describe QoE issues [5]

References

1 M. Fiedler, K. Tutschku, P. Carlsson, and A. Nilsson. *Identification of performance degradation in IP networks using throughput statistics.* Proc. 18th International Teletraffic Congress (ITC 18), pp 399–408, Berlin, Germany, Sept. 2003.
2 J. Shaikh, M. Fiedler, D. Collange, and P. Arlos. *Modeling and analysis of web usage and experience based on link-level measurements.* To be presented at the 24th International Teletraffic Congress (ITC 24), Cracow, Poland, Sept. 2012.
3 M. Fiedler. *Quality feedback flows in future networks.* elektrotechnik und informationstechnik (e&i), Springer, Vol. 7-8.2009, pp 269–273.
4 M. Fiedler, K. Wac, R. Bults, and P. Arlos. *Estimating performance of mobile services from comparative output-input analysis of end-to-end throughput.* To appear in IEEE Transactions on Mobile Computing.

5 M. Fiedler. *Teletraffic models for Quality of Experience assessment.* Tutorial at 23rd International Teletraffic Congress (ITC 23), San Francisco, CA, Sept. 2011.

4.8 QoE IPTV

Marie-Neige Garcia (TU Berlin, DE)

Models developed within the standardization groups such as the International Telecom Union (ITU) and the Video Quality Expert Group (VQEG) output video/audio/audiovisual quality (e.g. MOS) scores for short- time (10 s) audiovisual sequences, while providing a score for a whole session (e.g. a whole TV program in case of IPTV) may also be necessary. Moreover, the same score is provided for all types of user and context (e.g. for IPTV: test lab vs. home environment). These short-term quality scores are not sufficient for knowing if a customer is satisfied with the service/product she/he subscribed to. If we want to find the link between the quality scores provided by the quality metrics and user satisfaction, several points have to be addressed:

- How do we identify what the dimensions underlying QoE for a given application are? For IPTV, these dimensions may be for instance the perceived quality, accessibility/stability and usability/joy of use. Subjective tests are required for identifying these QoE dimensions. Multidimensional analysis (e.g. semantic differential scaling followed by Principal Component Analysis) and interview-based methods are examples of mix of qualitative and quantitative test methods which could be applied in that respect.

- How do we make QoE-models ecologically valid, user- and context- dependent? One requirement is bringing subjective testing to the field. In addition, new subjective tests should be designed for addressing long audiovisual sequences (as with the SSCQE of the ITU-R BT-500 Rec.) but with more appropriate task so that the subjects focus on the content and not on quality anymore (in the latter case, subjects are becoming too sensitive to impairments, see Staelens, 2010, IEEE Trans. On Broad.) For making QoE-models user-dependent, the most relevant criteria (personality? degree of expertise? Demographic data?) for characterizing the users should be identified. At last, we need to adapt the models when the 'technical' context is changing (e.g. IPTV is becoming interactive but models are trained in the context of a one-way (non- interactive IPTV services)?

This statement tried to identify requirements for modeling QoE in terms of subjective testing and in the context of a concrete application (IPTV).

4.9 User-centric troubleshooting

Riccardo Guerzoni (Huawei Technologies – München, DE)

The speech introduces a novel approach to current and future ICT networks troubleshooting, designed around user-centric criteria and QoE modeling. The proposed methodology for symptoms localization and root causes identification, denoted as User Centric Troubleshooting

(UTS), goes beyond the classical hierarchical relations between key performance indicators (KPI) and key quality indicators (KQI).

In the proposed process, the triggering point is the detection of performance deteriorations either in quality of experience (QoE) or in related QoS parameters. The input data may be customer complaints or QoE models. (It is well known that Network KPIs are not reliable parameters.) When any of the reference parameters goes below the corresponding threshold, all affected sessions and related pattern of anomalies are identified and aggregated by correlating Transaction Data Records (TDR), derived from network protocols, and Session Data Records (SDR), attained from upper layers protocols. The recurrence of an anomalies pattern (detectable by a clustering algorithm) and related KPI distribution make it possible to identify the root cause of the QoE/QoS issue, which triggered the diagnosis process, exactly like an expert engineer would operate. The information carried by the TDRs can be organized in classes of anomalies, standardizing the diagnosis parameters fed to the clustering algorithm. A pivoting step completes the analysis, identifying the principal components among the KPI dimensions (network context, service context, user context).

The aggregated per user per service (PSPU) SQM approach proposed by the UTS framework links the users segments impacted by the QoE deterioration to the root cause analysis, enabling the network operator to allocate efficiently the budget for the network optimization, working on the issues that are actually affecting the users perception and keeping the churn under control.

Experimental results showed the proposed solution to be an essential component for efficient user centric Customer Service Assurance.

4.10 Long-term QoE: How does the overall quality perception of one user evolve over multiple interactions with IP-based services?

Dennis Guse (TU Berlin, DE)

Shifting the perspective from technical parameter-based to perceptive quality estimation will allow tailoring service quality and network performance perfectly suiting to the user's need, i.e. provide an optimal QoE. In my research, I am focusing on the quality experience over multiple distinct interactions with one or more IP-based services. The user integrates his quality perceptions into an overall perceived quality, which influences his behavior like reusing the service. Understanding these effects would give us the capability to provide not only satisfying single service interactions, but going for real-life long-term QoE.

4.11 QoE Assessment for Web-based systems and IP TV

Richard John Harris (Massey University, NZ)

Subjective assessment for measuring the Quality of Experience may require time-consuming and often expensive methods, and yet, quantitative and accurate user scores are desired. In order to obtain valid correlation between analytical model and user scores, assessment based on networking perspectives and human perception is required. In our study, we have utilized orthogonal arrays using the Taguchi approach to construct an experiment to characterize the application as well as network performance metrics in our QoE assessment model for web-based systems.

For our study of QoE in IPTV and related systems we have divided the model into three basic components, viz: the Content Producer, the Network and the Customer Premises Equipment and selected metrics that influence QoE from each of these categories to study and develop models to be integrated into a single platform.

4.12 You look blocky, is everything alright?

Helmut Hlavacs (Universität Wien, AT)

Non-verbal communication is known to convey a significant part of the message between human beings during their interactions. Among other types of body language, facial expressions represent an important medium to evaluate the mood and feelings in the person one is interacting with. Nowadays, more and more distant communication is made possible by the fast growth of networks and devices such as mobile phones, tablets and traditional computers. For a few years now, we have been facing a rapid development of video communication, which allows distant persons to exchange either live or recorded messages in high fidelity. However, all distant communication has to go through a chain of treatments that can alter the quality of the delivered message. From the capture of the message using a video camera, through compression and transmission, then through decompression, post-treatment and finally to display, each step of the delivery process can introduce degradations in the data and deteriorate the message. The impact of the multimedia processing and delivery channel on the human ability to recognize facial expressions is therefore quite important and currently researched by us.

4.13 QoE: Measuring the Immeasurable?

Tobias Hossfeld (Universität Würzburg, DE)

My personal background is performance analysis in communication networks. As focus of this abstract, we consider a user-centric service and network management. From the view-point of a network or service provider, the goal may be to offer a good QoE to its customers while at the same time to optimize the costs for QoE management. Hence, the provider requires some indicators and instruments to quantify the users' (dis-)satisfaction with a service. The question arises whether QoE is the right path for this? Whether QoE can be measured in real-life system or whether QoE is too complex? Whether other (measurable) metrics instead of QoE can be exploited?

The intention of this abstract is to provoke discussions beyond QoE and to ask critically how to utilize QoE, how to implement QoE metrics for a certain purpose in practice, what are the next steps for QoE research in general.

One simple solution to overcome these challenges above is to go into a different direction than QoE. Instead of monitoring and estimating the user perceived subjective quality of a service, we may have a closer look at the user reactions and the user behavior. For example, the actions of an user (where and when he clicks on which elements of a web page) give a clear answer. E.g. when there is a problem with a web page, the user may reload the page or start another web page in parallel in a different tab of the Internet browser application. Another example is to measure how long and whether the user pauses a video to avoid stalling for example. This may show that the user is not happy with the current network situation and overcomes it. For an Internet service provider, the user reaction (especially, when the user churns) may be more important than the user perception! Furthermore, there is a known gap between user perception and user reaction.

In summary, as a next step of QoE research, the consideration and analysis of the user behavior (complementing the QoE perspective) seems to be important. Then, this user feedback should be taken into account in the service delivery accordingly.

4.14 Energy metrics unique enough for smartphone-based video QoE evaluation?

Selim Ickin (Blekinge Institute of Technology – Karlskrona, SE)

Mobile applications and services are used in daily life activities, to support the needs for information, communication or leisure. Mobile video applications, so called apps, are of the most energy/network demanding ones and need to be investigated further with the goal to achieve video quality close to the overall perceived quality of the applications running at fixed networks and terminals. In this talk, we will be presenting one unique metric, energy consumption, as a bellwether to locate both user activity levels and the video application statuses on a smartphone. Most popular video applications work based on transmission-controlled streaming, i.e., video plays without any jumps in the video frames. However, in this type of streaming, a stalling event, so- called freeze, is a common impairment that occurs during a video play-out and it is considered as a key influence factor in user's recent perceived video quality. An instantaneous increase in the network layer metric values such as delay and packet loss rate may not always be a herald for an interruption of a video streaming by a stalling event, due to its dependency on the content/size of the jitter buffer. Therefore, not only focussing on traditional QoS metrics, but investigation of other metrics that represent the state of the video streaming in the application level is necessary.

Our previous study tells that battery lifetime in smartphones is one of the most influencial factors for overall smartphone-based QoE. Monitoring QoE especially on these energy limited devices is a challange and needs smarter QoE monitoring frameworks. More recently, we have investigated that, also by continuous instantaneous per-application/per-service basis energy consumption measurements, it might be possible to detect anomalies such as stalling events in video applications. We have recently investigated that the frequency and the duration of the freezes during the runtime of a mobile phone based video player application are likely revealed by energy consumption values. As our hypothesis, we propose an energy consumption metric that is likely correlated with the stalling events in transmission-controlled video streaming.

Abnormal fluctutations of instantaneous energy consumption metric can indicate the stalling events experienced in the phone display during a user's video streaming session. In this way, instead of using hard to deploy and high energy demanding network level monitoring tools, we can facilitate the built-in energy measurement tools on mobile handheld devices to locate those anomalies. Once a suspicious behaviour such as stalling events are detected through energy measurements, deeper analysis of underlying QoS metrics, sensors, and all other user dependent traces can be enabled, which are expected to provide a more energy friendly future QoE monitoring framework.

4.15 Repeatable Results – The Key to Scientific Accuracy

Lucjan Janowski (AGH Univ. of Science & Technology – Krakow, PL)

As community we produce lots of metrics, indicators and other type results, the problem is that they are not checked by others. The key to have meaningful results is to be able to repeat them. It is not QoE community problem but in general IT. Orange lab took 99 % accurate traffic classification algorithm and it classified 50 % of traffic as unknown. In order to change it AGH would like to propose a join effort of creating a journal focused on QoE problems. This journal will publish not only currently the most important trends in QoE research but also metrics validation, repeating subjective experiments conditions, and other articles focused on checking other researchers' results. In addition the journal could provide a platform to review a paper not only by checking its text but also results.

4.16 Impact of physical layer impairments on higher layer QoS parameters

Maria Kihl (Lund University, SE)

An increasing demand from e.g. real-time multimedia applications (IPTV, OTT) adds strains on especially DSL based access links. Advanced physical layer monitoring tools are deployed, however, the mapping from physical layer impairments to network layer QoS parameters is still rather unknown. In this presentation we argue why it is important to study DSL link impairments and their impact on QoS parameters on higher layers.

4.17 Towards Total Quality of Experience. A Conceptual View on QoE in a communication ecosystem

Khalil Laghari (Télécom SudParis – Evry, FR)

With clear understanding on how human behavior is shaped and what disparate factors could influence his/her needs and expectations in particular context, it is possible to get more reliable and authentic view on QoE. A communication ecosystem brings together various domains such as technical aspects, business models, human behavior and context. For each aspect of a communication ecosystem, various models have been developed. However, few models have been designed to integrate all aspects of a communication ecosystem to understand human behavioral needs in a detailed and structured way. While existing models have produced the basic sketch of QoE modeling, more concepts and inter- domain mapping are to be incorporated in order to have a clear picture of QoE in a communication ecosystem.

4.18 Quality of Experience is not only Quality of User Experience

Patrick Le Callet (Université de Nantes, FR)

Since the late 90's, QoE has reached its great momentum in several communities. Nevertheless, clear understanding of its relative concepts is still missing. QoE to be practicable needs contextualisation. An obvious witness relies on the effort raised by the UX (User Experience) community. Assessment of the user experience is a key aspect and with this respect QoE is certainly related to UX. Nevertheless, QoE encompasses more scenarios .... Experience is not necessarily attached to the user of services or applications and encompasses much more than the relationship between a user and a service in many cases: visiting an expo, watching a movie, wine tasting .... The contents care and the emotion that one be suffered to convey.

4.19 How to come to good QoE instrumentation?

Sebastian Möller (TU Berlin, DE)

In this talk, 10 steps are highlighted which are necessary to define instrumentation concepts for QoE: (1) Define what you understand by Quality of Experience (Qualinet white paper). (2) Analyze relevant perceptual dimensions (features) of QoE. (3) Structurize the space of features (see the example in Möller et al., QoMEx 2011, for speech services). (4) Investigate models to predict desired features (several models are still missing). (5) Analyze how you can obtain the necessary input information (what is possible?). (6) Analyze possible user behavior (carry out data analytics). (7) Investigate whether and how you can predict such behavior. (8) Define what you do instrumentation for (monitoring, adaptation, charging, offers, etc.). (9) Decide which features and which user actions contribute to that aim. (10) Define an instrumentation approach.

4.20 Qo(E) Vadis? Multi-user, multi-service, multi-information, multi-timescales

Alexander Raake (TU Berlin, DE)

The majority of current QoE research is focusing on unidimensional, utilitarian measures of QoE, expressed for an average user of an individual service. Typically, such approaches are taking a system-, that is, object-oriented perspective, where the main goal is the QoE-based performance evaluation of objects such as services, applications or multimedia systems. For practical application, user tests are complemented or fully replaced by instrumental models used to predict QoE as perceived by users. Such predictions are usually based on model input information such as system parameters or multimedia output signals.

These approaches must be extended in different directions:

- For current models, the underlying input information describes the system or partially also the transmitted content, but mostly does not include contextual information or user behavior data. To overcome the limited ecological validity, and exploit all relevant sources of information on user-perceived QoE, multi-level-information approaches must be adopted including both context and behavioral information. Research challenges will lie in identifying the optimal set of information, and developing strategies for handling missing information.
- Most QoE models are agnostic to the kind of user, although QoE is individual to a given user, and largely depends on her personality and current state. For different applications, it is desirable to overcome this limitation by making models user-specific. Here, an appropriate demographic and role-related user classification as well as behavioral information will need to be exploited by future approaches, and included in a respective user-model.
- The still sparse examples of multi-timescale QoE assessment, such as in the case of call quality models for speech QoE of telephony services, must be complemented, to go beyond quality predictions for a single, mostly short-term time scale in the 5-16 sec range. Temporal pooling of QoE features and QoE episodes across different time-scales is not well understood, and a highly relevant topic for further QoE-research.
- If QoE models for an individual service are used for service optimization, the across-service perspective of subscribers is normally not reflected. As a complement, a multi-service perspective must be adopted, catering for how users integrate QoE across different services.

These research topics have to be addressed by finding a good balance between the achieved degree of ecological validity and the practical applicability for deployable QoE assessment solutions.

4.21 On Economics of QoE

Peter Reichl (FZ Telekommunikation Wien, AT)

Probably, QoE can be considered the perfect example for a research topic in the interdisciplinary research field which has recently been termed "Communication Ecosystems" (cf. Kilkki, Laghari et al.), which unites the technological (T), economical (E), user (U) and context (C) perspectives on communication systems as orthogonal and/or competing forces. While the interface U-T has made significant progress by now, both "technoeconomics" (T-E) and "socioeconomics" (U-E) are still lagging behind, while context basically provides an additional orthogonal component. Our main goal is to revisit the economical part of this sketched ecosystem. To this end, we focus on the methodological role of economics as a means of modeling the interaction between users as well as between user and technological environment. Following a hierarchical ecosystem model, competition takes place within a layer, while user utility is maximized by employing lower layer resources.

4.22 Automatic QoE assessment

Gerardo Rubino (INRIA – Rennes, FR)

1/ The PSQA (Pseudo-Subjective Quality Assessment) technology.

In the past, we developed a no-reference parametric assessment methodology for the evaluation of the perceptual quality of a voice or video + audio communication when the channel includes a packet network. PSQA works by measuring specific QoS metrics and specific application-based parameters, and then by invoking a particular function that maps all these variables into a MOS-like number. These metrics and parameters must then be accessible, and with almost no cost (that is, they must be measured efficiently, or available "for free"). The function is built using statistical learning tools, working on data coming from subjective testing sessions, and a mix of random and quasi-random sampling techniques to prepare the sequences to evaluate in those sessions. Once built, PSQA works in real time, if useful or necessary.

We claim that PSQA is accurate enough for any network-oriented usage (typically, for network monitoring or for network control). The procedure is network- and application-dependent: if the network and/or the application evolve, or change, the measuring module must be developed again from the beginning. However, this effort is done only once, before putting PSQA at work. Once built, the module just measures the instantaneous perceptual quality at time t.

We are currently in the process of putting PSQA in industry, in a "2.0" version, where we are integrating the experience cumulated over some years of application in several different domains, together with improvements on the original techniques.

2/ From perceptual quality to QoE

Compared to the case of voice or video communications, QoE covers an immensely larger range. This makes the idea of extending the parametric approach we followed for the simpler perceptual quality assessment of voice or video communications difficult to apply as such. Said in other terms, this huge universe makes difficult to achieve the same accuracy than PSQA in case of, say, a given Web service, or when considering the generic class of Web services as a whole. Instead, we are exploring two ideas.

(i) First, we are looking at a procedure based on the assumption that the target, the system to be assessed from the QoE point of view, exists in a large number of instances. The idea is to request specific users, at specific points in time, to provide an opinion about the QoE of the system, and to integrate the collected information in a way such that we can elaborate an automatic and accurate QoE measuring tool.
(ii) Second, we are considering the idea of building a QoE metric with values in a multi-dimensional space equipped with some mathematical structure.

We are currently working on the first point above, point (i), focusing on the mathematical foundation of the approach. The point is that the subjective views given by the users of the measured system will not have the same value (the same "quality") as the scores obtained from a controlled subjective testing experiment. The other side of the coin is the fact that we expect a large number of opinions, must larger than the number that can be obtained using panels of users in the lab. The situation is similar (at least, formally) as what one encounters in Monte Carlo, where for estimating the integral of $f()$ in some interval, we only need to answer the question "is $f(x) > u$?" given x and some random value u, for many pairs (x, u). The question is quite simple, but the result can be an extremely accurate answer to a non-trivial question. Technically, we are looking at what happens with accelerated Monte Carlo where the system's dynamics is completely changed in order to reach efficiency criteria. We are also looking at Quasi-Monte Carlo methods, where instead of random u values with use quasi-random ones, that is, weak discrepancy sequences. Here the goal is some regularity in the sample (instead of something mimic randomness), a tool we already use in PSQA.

Our first objective is to develop a first set of mathematical tools able to be combined into a measuring technique. Then, we must test the idea on well-chosen specific cases, including some of the scenarios we are identifying in the QuEEN project.

4.23 Non-intrusive network-based estimation of QoE

Junaid Shaikh (Blekinge Institute of Technology – Karlskrona, SE)

The number and types of applications on the Internet are increasing. Each application has its own performance requirements to work smoothly on the network. On a network with the same available resources, the same user can have different experiences based on the type of application used. Therefore, dimensioning a network for all types of applications according to the same criterion may either lead to the congestion (and user churn) or waste of resources (under-utilization of available capacity). Both situations are undesirable for the network operators. It triggers the need for QoE-aware management of networks, in order to organize networks dynamically based on the real-time estimation of QoE.

Several assessment models have been proposed to estimate QoE. Most of them are intrusive and require knowledge of the content reference. In contrast, the network operators require non-intrusive methods, which allow models to be implementable on the network-level without having much knowledge about that reference. The methods should be able to monitor QoE passively in real-time, based on the information readily available on network level. Considering the high-speed networks, methods should also be fast and practically implementable.

Our work is based on the non-intrusive methods to infer QoE based on the objective indicators obtained from the network traffic. The methods take into account ON and OFF phases of user traffic on the network. They do not require deep packet header information. Amongst others, these methods capture the temporal aspects of QoE and locate those time instances where the problems occur. By using these methods, the frequency and duration of user-perceived problems could be visualized at varying time scales.

Once developed, the above-mentioned methods are intended to be used in the live networks to estimate QoE in the quasi-real-time.

4.24 From modeling QoE to QoE management: challenges for domain-wide QoE-driven resource allocation

Lea Skorin-Kapov (University of Zagreb, HR)

While a large number of approaches to date employing utility-based resource allocation concepts have linked technical performance metrics (associated with QoS) with perceived service quality, the advent of QoE-related research has led to advancement in the understanding of QoE metrics, focusing on subjective quality perception and subsequent human psychological and behavioral models. It has been widely accepted that QoE is a multidimensional construct, comprised of multiple user perceived quality dimensions. In the domain of QoE-driven service optimization, the concept of utility functions can hence be applied to include such multiple dimensions (e.g., ease-of-use, efficiency, comfort, satisfaction, visual quality, willingness to pay, perceived value for money, etc.) and their impact on an overall (integral) QoE. Furthermore, it should be noted that different dimensions are relevant for different types of services (e.g., conversational voice, streaming video, cloud applications, multiplayer games, etc.).

Based on QoE models and estimation methods, we can attempt to correlate QoE dimensions at a given point in time with QoE influence factors (related to application/service parameters, allocated resources, user parameters, context parameters). By understanding this correlation, and the user, service, and network constraints which are pertinent, we can formulate QoE optimization problems related to tuning (where possible) influence factors (e.g., resource-related, application configuration-related) to maximize QoE.

A question we raise is how to go from our estimation of QoE in a multidimensional QoE space (having so far considered a single user and a single service), to performing domain-wide QoE optimization, whereby we are looking for a global solution in a multi-user, multi-service QoE space, constrained by available network resources, user subscriptions, different user preferences and capabilities, operator policy, etc.? We discuss some challenges related to QoE-driven resource allocation in a domain-wide scenario. Not only do we have multiple dimensions of QoE that need to be considered for different services (and media flows in the case of multimodal services), but we also have multiple simultaneous user sessions. In practice, formulation of the objective function for optimizing resource allocation may differ depending on whose interests are being considered (e.g., user, network operator, or both in the case of multiobjective optimization). Different examples include: (1) maximizing the (weighted) sum of QoE values across end users, expressed generally as functions of QoS parameters, (2) maximizing number of "satisfied" users, i.e., with QoE above a certain threshold, (3) maximizing operator profit, by minimizing operator costs, or (4) a combination of the previous objectives.

4.25 A Generic Approach for Understanding QoE

Martin Varela (VTT Technical Research Centre of Finland – Oulu, FI)

QoE research has so far been mainly focused on media services and applications, such as (Internet) telephony, video, etc. The concept of QoE, however, is applicable (and important!) to a much broader domain of services and applications, which are becoming ever more important in everyday life, as more and more activities take place online. Online collaboration, social networking, e-banking, and a myriad other applications, often web-based and residing in the Cloud, are now an integral part of everyday life. Oftentimes, applications which used to be local to the user's devices, now reside somewhere else, with a network, hosting platform and application platform in between them and the users, all of which introduce potential degradations in the way the user experience the service.

Given the ubiquity of these applications and the increasing importance of their role in our lives, it makes sense to try and understand how we experience our interactions with them (i.e. what is 'their QoE'). This poses several challenges, starting with basic ones such as understanding what quality means for users in the context of each of these services.

In the context of the Celtic+ QuEEN project, and also together with Qualinet colleagues, we are currently working on a conceptual framework for understanding QoE for any service. The goal is to be able to reason about QoE and exploit it in different ways (e.g. SLA negotiation/enforcement, QoE-driven network management, etc.)

To this end, we consider QoE as a multi-dimensional construct which depends on several (with 'several' » 'a few') factors, and we strive to understand the relations between these factors and the many aspects (or dimensions) of QoE. For this we use different tools, such as Lea Skorin-Kapov's ARCU model for classifying QoE-influencing factors, and a layered, compositional view of quality for understanding the relations between different parts of the service-user system.

We further consider that services often comprise multiple modalities, which play differently on the overall (integral) QoE of the service (i.e. they can relate differently to different dimensions of QoE), and which may change over time within an usage session (e.g. video-conferencing with screen-sharing facilities).

As of this writing, we have a first approach to formalizing the concepts above, and we are currently working on incorporating the temporal/multi-modality aspects to the model. The main challenge, however, lies in finding efficient ways of constructing a mapping from the quality-influencing factors, and QoE itself, taking into account all of the above. We have, in the past, constructed such mappings in mono-modal applications (VoIP, video) for single dimensions of QoE (e.g. with PSQA). More recently, we have successfully created such a mapping for another mono-modal application (video), considering several QoE dimensions in the output, as well as an integral QoE estimation. Making this work in a more generic setting, however, with larger number of influencing factors and QoE dimensions, presents a non-trivial challenge in terms of the amounts of data needed to be able to use statistical approaches to create the mappings. We are currently working towards such a more complex use case, for a web-based service.

4.26 Factors Influencing Quality of Experience of Commonly-Used Mobile Applications

Katarzyna Wac (University of Geneva, CH)

Increasingly, we use mobile applications and services in our daily life activities, to support our needs for information, communication or leisure. However, user acceptance of a mobile application depends on at least two conditions: the application's perceived experience and the appropriateness of the application to the user's context and needs. Yet, we have a weak understanding of a mobile user's Quality of Experience (QoE) and the factors influencing it. We present a week long, 29 Android phone users study, where we collected both QoE and underlying network's Quality of Service (QoS) measures through a combination of user, application and network data on the user's phones. We aimed to derive and improve the understanding of users' QoE for a set of widely used mobile applications in users' natural environments and different daily context. We present data acquired in the study and discuss implications for mobile applications design.

4.27 The influence of contextual factors on quality ratings

Ina Wechsung (TU Berlin, DE)

User eXperience (UX) is widely understood as a highly context-dependent construct based on individual perceptions. This position implies that to achieve meaningful measurements of UX the context needs to be taken into account.

Research indicates that judgment and decision making involves two systems, the cognitive-rational and the intuitive-emotive system (e.g. Kahneman 2003). Compared to the emotive system, the cognitive-rational system is more analytic, logical, abstract, active, controlled, rule-based and slower; it is the deliberate mode of judgment [1]. The intuitive-emotive system on the other hand is characterized by automatic, associative, effortless, and often emotionally charged operations; it is the automatic mode of judgment , and is also shown to be more context- specific than the rational system [2, 3]. The heuristic, context-specific, emotive system determine preferences and judgments unless the cognitive system intervenes [1].

To gain a better understanding of the role of context in the judgmental process, we investigated whether findings from cognitive psychology could be transferred to the judgment of interactive systems. For example we found the mood congruency effect [4] to be also applicable to HCI: the better the mood of the participants, the better the ratings for perceived hedonic quality. Furthermore, we showed that increasing mental workload by introducing a parallel task decreased the perceived pragmatic quality. In our most recent study, we compared ratings collected "online" during a field trial with ratings assessed after the usage period. While quantitative ratings of overall quality were similar, the qualitative data differed: comments collected during usage were more specific with respect to certain negative or positive aspects of the apps performance. Participants often only reported problems, not judgments. Comments collected after usage were often rather general, however they also contained an affective appraisal of the experience. Thus the remembered user experience does not necessarily represent a one-to-one reflection of the actual user experience.

The results reported above show that although laboratory studies that aim to strictly eliminate contextual factors might be appropriate for performance evaluation, such settings are certainly not the best approach for meaningful assessments of the multi-faceted concept UX.

References

1 Kahneman, D. (2003) A Psychological Perspective on Economics, American Economic Review, 93(2), pp. 162-168.
2 Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996) Individual differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. Journal of Personality and Social Psychology, 71(2), pp. 390-405.
3 Denes-Raj, V. and Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgement. Journal of Personality and Social Psychology, 66(5), pp. 819-829.
4 Schwarz, N. and Clore, G. L. (1983) Mood, misattribution and judgement of well-being. Informative and directive functions of affective states. Journal of Personality and Social Psychology, 45(3), pp. 513-523.

5 Working Groups

5.1 QoE White Paper and Group Work 1: Key Aspects of Experience Perception and Their Subjective evaluation

Sebastian Möller, Sebastian Egger, and Markus Fiedler

Main reference "Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)", Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Lausanne, Switzerland, Output version of the Dagstuhl seminar 12181, 2012. (last seen 2012-07-21).
URL http://www.qualinet.eu

On the end of the first day, after the participants' presentations, the QoE White Paper prepared by the Qualinet group, led by Sebastian Möller and Patrick Le Callet, was presented. The QoE definition in this document ("degree of delight") originates from the Dagstuhl Seminar 09192 "From QoS to QoE". The subsequent discussion dealt with the relation between QoE and User eXperience (UX); the concept of experience; the relationship of QoE and the Communication Ecosystem; influencing factors on user level; quality features on service level; and the relationship between QoS and QoE.

These discussion points led to group work, which resulted in an updated version of the Qualinet White Paper, in particular with respect to the concepts of experience and quality of experience, as well as with respect to multimedia learning. This updated version was presented in the Friday morning wrap-up session.

During this session, it was agreed to compile an Output Version of the White Paper until 18.05.2012, work led by Sebastian Möller, with the following tasks:

- Start a discussion in a larger group around the term "experience" and clarify whether this also include perception and judgment processes and user states; potentially provide an update of the definition in Section 2 of the White Paper;
- Check whether the definition of QoE needs to be modified given a new definition of experience; check whether the sentence related to telecommunication services can be excluded from the definition to reflect that QoE exists also without telecommunication services;
- Provide a tentative text to be added to Section 4 which clarifies the different roles users might take in a communication ecosystem, and how this relates to QoE (Kalevi Kilkki could provide support);
- Provide an update of the paragraph of Section 4 which related to multimedia learning, so that it better reflects QoE; provide a short explanation of what the "level of service" means for the QoE features in Section 6.

In the meantime, these targets were all reached, and the Output Version of the Qualinet White Paper is now available through the Qualinet Web Site [1].

References

1 "Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)", Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Lausanne, Switzerland, Output version of the Dagstuhl seminar 12181, 2012. http://www.qualinet.eu (last seen 2012-07-21).

5.2 Group Work 2: Measurable Aspects of QoE

Peter Reichl, Martin Varela, and Markus Fiedler

In parallel with group work 1, this group discussed issues and challenges related to measurable QoE-affecting factors. The results were summarised in a table, listing features and related properties such as measurability, relevance and challenges. A summary of the results is provided hereafter:

- Personality, health aspects and mood are measurable in general (personality rather indirectly). Challenges relate to how to measure them, in particular in field measurements.
- Experience and demographic data are knowable and can be measured both in- and directly. No particular challenges were identified.
- Expectations are indirect measures. They are deemed the more relevant and the more difficult to estimate, the longer the underlying time perspective becomes. Challenges are amongst others found in dependencies on applications and in the roughness of the estimates.
- Technical factors can be measured for the most part in a quite straightforward way. Environmental factors, as far as they are relevant, might be difficult to measure in the field. Socio-economic factors can be estimated, while it is difficult to capture their influence in models.
- Group and role factors can be estimated and are knowable. Role factors are deemed more influential and can be assessed through social graph analysis.

- Mobility factors are directs measure. Task factors can also be measured indirectly and might be difficult to assess.
- Cost factors are knowable, while emergency factors cannot be measured.

A use case exemplifying the above factors was prepared by Tobias Hossfeld.

5.3 Group Work 3: Identification of QoE-Related Feedback Loops

Markus Fiedler and Riccardo Guerzoni

This group work was also performed in parallel to the other two group works, and was structured as follows:

1. *History and motivation.* One of the strongest drivers of QoE is the risk of user churn in face of perceived problems. It is important to find out about users' real opinions and their dynamics in order to pave proactive ways to avoid churn. As questionnaires add noise to customer experience, objective measurements for the entire population are needed.
2. *Classification.* A set of explicit feedbacks (e.g. push an "anger" button; call the support; body response from face expression via exclamations and device abuse to violence), implicit feedbacks (e.g. more and longer activity; increased usage frequency and degree of completed transactions; higher spendings; word of mouth etc.) and hybrids (e.g. timings in the user interface; stop/reload buttons) were identified.
3. *Inventory.* It was discussed which QoE-related feedback a user expects, which feedback facilities a user might miss, and which feedback should be provided from the operator's point of view. Obviously, there is a risk of over-polling the user. Feedback should correlate to the possibilities of the user to control the situation. It is definitely dependent of the type of service. A faithful user would consider no news as good news and value notifications of "turbulence ahead", which is particularly important for transactions that involve several steps.
4. *Construction of new feedback loops.* This part, which seems to be best done based on case studies, still needs to be addressed, e.g. for some interactive service. From the manufacturer's point of view, questions on how to use QoE evaluation for network management and how to assure that the investments into quality improvements feed back positively into the accounts of the network operators are of high importance.

The sub-group planned to follow up on this topic with a survey paper (during Summer/Autumn 2012).

6 Conclusions and Open Challenges

6.1 Conclusions and Open Challenges

Markus Fiedler (Blekinge Institute of Technology – Karlskrona, SE)

The first part of the last – half – day of the seminar (i.e. Friday morning) was devoted to summaries and discussion of the outcomes of the group works in the plenary. For group 2, a case study was presented. Group 3's feedback scenario triggered a lively exchange of ideas. The omnipresent challenge of how to provide feedback to one's users was stressed in particular by a representative of a vendor. Further issues addressed were how to make users aware of quality, how to avoid cheating by users, and how to tackle privacy issues when it comes to monitoring user traffic. The discussion regarding group 1's work circled around the notions of experience (as opposed to events, and in relationship to anticipations) and quality (in relationship to qualitas; absence of the temporal aspect). In principal, everybody using the term QoE should answer the question "How does my usage of the QoE limit a general QoE definition", as for instance provided in the forthcoming QoE White Paper [1], cf. also Section 5.1.

The final session on Friday morning was devoted to the discussion of generic items of joint interest, such as

- a specific journal on QoE, following the proposal of Lucjan Janowksi presented in Section 4.15;
- standardisation (realising the orthogonality of the communities and particularities of standardisation work,such as obsession about details); in this context, an upcoming ETSI workshop on QoE [2] and work on a QoE framework and an communication ecosystem was announced by Sergio Beker;
- follow-up work on the QoE White Paper and the documentation of this Dagstuhl Seminar (with corresponding deadlines);
- ideas for the organisation of a follow-up seminar, such as mobilisation of the intercontinental communities already in the application stage, and the use of 30-second elevator pitches and pre-prepared contributions to the group discussions;
- brainstorming about intermediate activities in order to keep the momentum.

It remains to point out that the reduction of the seminar from five to three days in combination with an extensive discussion of the notion of QoE left (much) less time for the treatment of the instrumentation topic as compared to the initial planning. On the other hand, new ideas for future directions of a deepening of the QoE topic in the Dagstuhl context emanated from those discussions. Two-and-a-half months after the end of the seminar, a set of outcomes can be reported (the Qualinet White Paper; joint work on papers; discussions between academia and industry). Thus, there are good reasons to assume that the "Dagstuhl QoE community" will remain active and visible.

References

1 Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Lausanne, Switzerland, Output version of the Dagstuhl seminar 12181, 2012. http://www.qualinet.eu (last seen 2012-07-21).

2 ETSI workshop on QoE. http://www.etsi.org/WebSite/NewsandEvents/2012_STQWORKSHOP.aspx (last seen 2012-07-21).

Participants

- Patrik Arlos
Blekinge Institute of Technology – Karlskrona, SE
- Ake Arvidsson
Ericsson Res. – Stockholm, SE
- Sergio Beker
Huawei Techn. – München, DE
- Katrien R. De Moor
Gent University, BE
- Marcus Eckert
TU Chemnitz, DE
- Sebastian Egger
FZ Telekommunikation Wien, AT
- Markus Fiedler
Blekinge Institute of Technology – Karlskrona, SE
- Marie-Neige Garcia
TU Berlin, DE
- Riccardo Guerzoni
Huawei Techn. – München, DE
- Dennis Guse
TU Berlin, DE
- Richard John Harris
Massey University, NZ
- Helmut Hlavacs
Universität Wien, AT
- Tobias Hoßfeld
Universität Würzburg, DE
- Selim Ickin
Blekinge Institute of Technology – Karlskrona, SE
- Lucjan Janowski
AGH Univ. of Science & Technology – Krakow, PL
- Maria Kihl
Lund University, SE
- Khalil Laghari
Télécom SudParis – Evry, FR
- Patrick Le Callet
Université de Nantes, FR
- Sebastian Möller
TU Berlin, DE
- Alexander Raake
TU Berlin, DE
- Peter Reichl
FZ Telekommunikation Wien, AT
- Gerardo Rubino
INRIA – Rennes, FR
- Junaid Shaikh
Blekinge Institute of Technology – Karlskrona, SE
- Lea Skorin-Kapov
University of Zagreb, HR
- Martin Varela
VTT Technical Research Centre of Finland – Oulu, FI
- Katarzyna Wac
University of Geneva, CH
- Ina Wechsung
TU Berlin, DE

Report from Dagstuhl Perspectives Workshop 12182

Social, Supply-Chain, Administrative, Business, Commerce, Political Networks: a Multi-Discipline Perspective

Edited by

Matthias Häsel[1], Thorsten Quandt[2], and Gottfried Vossen[3]

1 **Otto Group – Hamburg, DE,** Matthias.Haesel@otto.de
2 **Universität Hohenheim, DE,** thorsten.quandt@uni-hohenheim.de
3 **Universität Münster, DE,** vossen@uni-muenster.de

Abstract

The information society is shaped by an increasing presence of networks in various manifestations, most notably computer networks, supply-chain networks, and social networks. Online networks nowadays connect people all around the world at day and night, and allow to communicate and to work collaboratively and efficiently. What has been a commodity in the private as well as in the enterprise sectors independently for quite some time now is currently growing together at an increasing pace. As a consequence, the time has come for the relevant sciences, including computer science, information systems, social sciences, economics, communication sciences, and others, to give up their traditional "silo-style" thinking and enter into borderless dialogue and interaction. The purpose of this Perspectives Workshop has been to play an enabling role in this future dialogue, to review where we stand today, and to outline directions in which we urgently need to move, in terms of both research and teaching, but also in terms of funding. This report summarizes the discussions of the workshop and their findings.

Perspectives Workshop 02.–04. May, 2012 – www.dagstuhl.de/12182

1998 ACM Subject Classification A.0 General, A.2 Reference; H. Information Systems, J.4 Social and Behavioral Sciences, K.4 Computers and Society.

Keywords and phrases Networks, network infrastructure, network types, network effects, data in networks, social networks, social media, crowdsourcing

Digital Object Identifier 10.4230/DagRep.2.5.26

Social, Supply-Chain, Administrative, Business, Commerce, Political networks: a multi-discipline perspective, *Dagstuhl Reports*, Vol. 2, Issue 5, pp. 26–42

Editors: Matthias Häsel, Thorsten Quandt, and Gottfried Vossen

Dagstuhl Reports

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

1 Executive Summary

Matthias Häsel
Thorsten Quandt
Gottfried Vossen

The information society is shaped by an increasing presence of networks in various manifestations. Efficient computer networks are regarded as a significant enabler for the process of change towards networks of any size and complexity. They serve as an administrative and technological basis for social network structures, with the result that online networks connect people all around the world at day and night, and allow to communicate and to work collaboratively, efficiently, and without recognizable time delay. Companies reduce their in-house production depth, join forces in supply chain networks and establish cooperation with their suppliers, with their customers, and even with their competitors. By now, social networks like Facebook, Google+, LinkedIn or XING are seen as the de facto standard of "social networking" in the information society. Companies are mimicking their effects internally, allow overlays of networking applications with regular business ones, and a use of social networks for enterprise purposes including and beyond advertising has become common. Public administrations create and improve shared services and establish "Private Public Partnerships (PPP)" to benefit from synergetic effects of cooperation with private and public organizations.

As the interactions between people in these networks increase at various levels, new approaches are needed to analyze and study networks and their effects in such a way that individuals as well as organizations and enterprises can benefit from them. This Perspectives Workshops has convincingly shown that more interaction and collaboration between fields such as information systems, computer science, social sciences, economics, communication sciences and others is needed. The fields need to identify a common level of language, tools and set of methodologies so that the various aspects of networking can be addressed and jointly developed further. The most important point is the need for a renewed multi-disciplinarity. To a great extent, networks are driven and further developed by practitioners; which also means that they are evolving in a very fast manner and not emanating from a single scientific discipline. To be able to both understand them and contribute to the state of art, true inter- or multi-disciplinary research is needed that involves the fields mentioned. As these distinct disciplines grow together and embark on collaborative research, it is also important to convince funding agencies that multi-disciplinary research should arrive on their agendas. Finally, Web sciences need to be developed as a field, and also need to be integrated into teaching. This will most likely lead to novel curricula which receive their content from multiple disciplines in a balanced way.

2 Table of Contents

3 Introduction

The information society is shaped by an increasing presence of networks in various manifestations. Efficient computer networks are regarded as a significant enabler for the process of change towards networks of any size and complexity. They serve as an administrative and technological basis for social network structures, with the result that online networks connect people all around the world at day and night, and allow to communicate and to work collaboratively, efficiently, and without recognizable time delay. Companies reduce their in-house production depth, join forces in supply chain networks and establish cooperation with their suppliers, with their customers, and even with their competitors. By now, social networks like Facebook, Google+, LinkedIn or XING are seen as the de facto standard of "social networking" in the information society. Companies are mimicking their effects internally, allow overlays of networking applications with regular business ones, and a use of social networks for enterprise purposes including and beyond advertising has become common. Public administrations create and improve shared services and establish "Private Public Partnerships (PPP)" to benefit from synergetic effects of cooperation with private and public organizations.

All types of networks have in common that they consist of actors (such as persons, computers, companies, administrations, etc.) which are related amongst and to each other; in a graph-based representation, actors become nodes, and their relationship become edges. At an abstract level, actors and their relationships can be subject to network theory, which is related to mathematical graph theory. From this perspective networks are systems that can be represented as (directed or undirected) graphs consisting of vertices and edges, representing actors and their relationships. As a consequence of ubiquitous computing and application integration that transcends organizations, networks are increasingly complex, from rather small ones (e.g., business contacts, friendships, federal organizations) towards highly complex networks of networks. The emergence of the latter suggests that the involved participants expect benefits from joining and participating in a network, such as by accessing information more easily or at lower cost (than without the network). Additionally, participants expect that uncertainty in business and private relations can be reduced by establishing networks. For instance, companies become capable of reacting more quickly to demand and supply fluctuations if routines such as load balancing have appropriately been established in their supply chain networks. Moreover, companies strive to reduce uncertainty related to hiring new employees by consulting social networks, while individuals keep themselves informed about the latest news in their social network by receiving information from their social networks — often aided by mobile devices that provide ubiquitous access to the Web. Networks thus emerge, persist, and dissolve not only in commercial, public, or private settings, but transcend traditional boundaries.

4 Purpose of the Workshop

With this Perspectives Workshop, it has been our intention to focus on three fundamental aspects of networks in order to analyze and study the design, interplay, and behavior of networks in the Information Society:

1. *Drivers*: Networks can be regarded as systems that are continuously shaped by their environment. In fact, the emergent structure and properties of networks are subject to

self-organizing processes — not unlike evolutionary processes — that create structure in the form of temporarily stable patterns of interaction between actors.
2. *Cohesion*: In a general context cohesion describes the phenomenon of (economic and/or social) solidarity, or, in other words, the intention of actors to act in the middle of their neighbors. Structural cohesion is the sociological and graph-theoretical conception for evaluating the behavior of social groups and networks.
3. *Dynamics*: A dynamic system is a system that changes its state over time. Concerning different network application areas, we regard the dynamics of a system as the change of states a system takes. On the one hand, we consider a change of state in a network as the exchange of entities (information, goods, etc.) between its actors. On the other, the change of state in a network is regarded as its evalution, which may involve, among other aspects, a change of the underlying system's structure over time.

The resulting research question of the Perspectives Workshop has been:

> *How do online networks evolve, how can they be conceptualized, and how can they be consciously designed and influenced, given the fact that multiple disciplines have a increasingly common interest in networks?*

This question, which was adapted from the regulation framework originally described in Figure 0.1 of [2], has been addressed from a thorough multi-disciplinary point of view which allowed participants to look at the evolution of networks, the drivers for evolution that promote the emergence of online networks, and the lifecycle of networks, their emergence, behavior, and maturity. Since networks are influenced by numerous aspects, the workshop was structured along two dimensions:

- *Views on networks*: From a perspective of communication and trust, IT-enabled social interactions between individual human actors must be considered. Issues of trust are of particular importance in this context [14]. Taking a view on governance, formal rules organizing the relations of actors in the network have to be discussed. Processes being performed are another aspect deserving closer attention. Finally, the underlying information infrastructure, shaped by developments of computer networks (e.g. mobile applications, concepts like Web 2.0, etc.), must be taken into consideration as it clearly is a major factor influencing network evolution. An integrative treatment of these issues takes account of their interdependencies.
- *Domains of networks*: Networks can emerge in different areas. These can be structured into economy, politics and administration as well as the society at large, each having unique an impact on a network. Thus, social, supply-chain, administrative, business, commerce and political networks will be considered.

The diversity of these research areas as well as the fact that various issues arise in more than one field necessitates a multi-disciplinary approach. Consequently, researchers from various disciplines, including computer science, information systems, economics, supply-chain management, and communication studies, participated in the workshop and helped to shed light on online networks from various angles. Their objective was to develop mutual understanding and compile a comprehensive state-of-the-art of networks across the involved research disciplines, thereby unifying the perspectives of different researchers on networks.

The discussions started out in four different areas pertaining to networks, which will be discussed in Section 5. They led to a comprehensive set of recommendations for future research and teaching as well as for future funding as reported in Section 6, thereby paving the way for a network agenda in the information society for the coming decade.

5 Discussion Areas

The workshop started out with group work in four distinct areas that pertain to networks and networking:
1. Data in Networks,
2. network infrastructures,
3. the specifics of social networks and social media, and
4. the (observable) network effects of crowdsourcing.

For each group, whose work is reported in the four subsections of this section, the guiding goals were as follows:
- Identify the state-of-the-art and most pressing research issues for your group topic. What is already there, and what is needed or missing (in terms of expertise, research, products, mutual influence from other areas)?
- Is your topic (or its current treatment in research) purely academic or has it reached industrial exploitation already?
- Where do you expect networks to be in five years from now?
- If you have a new PhD student coming in, what do you want him or her to work on?

5.1 Data in Networks

Networks produce massive amounts of data, either automatically through machines (e.g., Web server logs, supply-chain control) or through user input. Indeed, user-generated content has been one of the distinctive features of "Web 2.0" or the evolution of the Web from a "read-only Web" to a "read-write Web" [19]. Moreover, accessible data on the Web, whether created by computers, by Web users, or generated within professional organizations, are growing at a tremendous pace. Social networks like Facebook, search engines like Google, or e-commerce sites like Amazon store new data in the terabyte range on a daily basis. Due to the emerging usage of cloud computing, this trend will not only continue, but accelerate over the coming years, as not only more and more data is generated, but also more and more data is permanently stored online, is linked to other data, and is aggregated in order to form new data.

Regarding the various kinds of data on the Web and in networks, including linked open data, socio-economic data, big data, and user-supplied data, relevant topics are technical aspects of data, usage patterns of data, types of data in networks (e.g., process data). Questions to be asked include, but are not limited to the following: Is storing all this data necessary? What can be done with all this data? How can data flow between networks? How can data produced in one network be beneficial for another?

Regarding data arising in the context of computer networks, a first observation was that the term "big data" should rather be "*broad data*," as various developments, including linked data, the Web of data, and others are currently coming together. In particular linked data [4] has gained recent popularity in the context of the Semantic Web [3], as Semantic Web people think in terms of *links*, as opposed the previous thinking in terms of *pages*. This perception nowadays also applies to data creation, updating, and analysis. Surprisingly, data scraping is still in wide use, since linking is not yet fully understood and reasonable alternatives are not available (e.g., based on metadata standard formats). On the other hand, data is most useful when it can be combined with other data, which is what we currently see on social networks like Facebook with their underlying graph databases, where there is a rising usage of inherent semantics as well as implicit context.

Clearly, data is heavily spreading across networks, but we still do not understand how to create networks appropriate for a specific purpose, where the spreading of data can be directed and controlled in some way, or how to bring together (structured, semi-structured etc.) data and information. Besides data, it is important to distinguish *networks of machines* from *networks of people*: There are attempts which try to join the two, while others want to keep them apart. Human actors are obviously important in the network picture, since they are prime suppliers of data and its connections, especially in social networks.

Whether data is linked or not, what is of increasing importance is to be able to identify *data provenance* (or data lineage, i.e., the ability to trace given data to its roots, points of creation, and along its history of changes). Provenance has a different meaning for data ("where does it come from?") and for networks; in the latter case, it is more process- or document-oriented. Data provenance [6] [7] has originated from scientific applications, e.g., in physics or in molecular biology, where reproducibility has always been an important aspect; however, provenance has meanwhile reached even business areas. From a technical perspective, provenance is often seen in connection to data curation, as exemplified, for instance, in the DBWiki project [5].

The traditional database approach to large data collections has been the *data warehouse* [15], where data is collected from various sources, put through an ETL process (short for *extraction*, *transformation*, and *loading*) and finally integrated into a single data collection, the warehouse. The latter then forms the basis for data analysis, online analytical processing (OLAP), as well as data mining. If that is to happen on the fly, a data warehouse is not good enough, since an ETL process takes too much time. A data warehouse stands for slow integration with high quality, whereas fast integration with lower quality is often more desirable, in particular since data changes occur on the network, "by themselves." For that, a linking of data sets seems again more appropriate; networks of data are needed, which at best amounts to a warehouse in a cloud-like world. This is in a way similar to developments in software engineering (from traditional to agile approaches).

Another important aspect is that data is increasingly considered as *goods* which have a *value* or come at a price: If the goods are rare, you collect them; if there is abundance, collecting is no longer necessary (an example is music, in particular records vs. music obtained from the Web). Indeed, marketplaces for data are on the rise (see, for example, the MIA project at TU Berlin[1]) which aim at the development of reliable and trusted platforms for the production, provision, and use of data. In this area, where sophisticated search and analysis tools are needed, there is a link to *crowdsourcing*, i.e., the idea to outsource a task to a possibly anonymous group of people. Numerous examples from recent years prove that having the user in the loop can improve data quality (e.g., maps of Haiti before and after the quake; the UK map of bus stops before and after it had been opened to the public). However, the effects achievable with crowdsourcing depend on the specifics of the crowd. Here it is important to distinguish between a "pre-defined" crowd in a professional environment (a "club," e.g., for building a plane) and a "randomly gathered" crowd (as in the case of the bus stops). The techniques used in either category may or may not be the same; the size of the crowd may be a determining factor: As the crowd gets larger, the need for individual experts potentially decreases, but control remains an issue and beyond a certain size experts are needed again for helping to separate useful information from nonsense. So a question is how

[1] http://www.dima.tu-berlin.de/menue/research/current_projects/mia/

a crowd can be triggered to do what it is expected to do or what a system (e.g., Wikipedia) requires them to do. Ultimately, such a crowd will decide about what is right and what is wrong.

A final aspect relating to data is that of a *process view*, which can be considered similar to views in relational databases; one process could be about gathering data, but integration and usage are done by other processes. Both processes and the data they operate upon might change over time. Processes may have common patterns; if they do, these (usage) patterns could have an impact the design of data gathering processes.

5.2 Network Infrastructures

Network infrastructures increasingly shape modern societies. In comparison to traditional infrastructures such as traffic, energy or health care, network infrastructures based on the Internet and its services are developed much faster, at a considerably wider scale, and they facilitate widespread participation. Computerized network infrastructures are easily scalable due to the availability of massive computing and storage power as well as network bandwidth and due to the availability of standardized protocols. They are versatile and represent a generative regime (they facilitate the growth of new infrastructures). Several network infrastructures can thus be conceptualized as commodities and are seen as a societal resource for innovation, economic development and welfare.

Topics to be discussed in this area include decentralized network architectures, cloud computing, emergence and design of network infrastructures, simulation of network behavior, informated logistics infrastructures. Questions include the following: Which infrastructures are particularly suited for which area (e.g., SCM and logistics, service industry)? Do we still need to care about infrastructure, or will it soon be all invisible like electrical current?

The *definition* of a network infrastructure should cover aspects such as non-rivalry access, one infrastructure for one purpose, and visibility only in the case of failure. According to Nicholas Carr,[2] infrastructure does not really make a difference, at least as far as IT infrastructure is concerned: If each enterprise has it, it can no longer help to sustain a competitive advantage. Important are standards; there is a technology stack with infrastructure at the lowest level. Infrastructures require administration, (legal) regulation, and accessibility. The proliferation of the network society may have an effect on infrastructures. Research topics to be studied include governance, comparison of infrastructure types, infrastructure lifecycles, vulnerability of infrastructures, as well as privacy. Twitter and Facebook have the potential to become infrastructures.

5.3 The Specifics of Social Networks and Social Media

Social networks are at the heart of modern network usage, demonstrated by the wide user coverage (if Facebook was a country, it would currently be the third biggest in the world). They have different foci, be it on personal or professional issues (or a mixture of both), they serve as extremely efficient and sometimes highly specialized news and communication platforms (think, for example, of the role of Twitter in the Arab spring of early 2011), and they are to an increasing degree discovered by enterprises as an instrument for reaching out

[2] http://www.nicholasgcarr.com/doesitmatter.html

internally to employees and externally to customers. The result is an increasing professional investment in social media technology and advertising, although the ultimate effects, in particular the external ones, still remain to be seen.

Topics for discussion in this area include (social) network analysis, social networks for the public domain, social media (networks), and social commerce. Questions to be asked are: Which distinctions can currently be made between various social networks (e.g., Facebook, Google+, Path et al.)? How could the future of social networks look like? Will Facebook be the new "operating system" of the Internet? What value do online social networks have for an economy from a macro-economic perspective? What influence does my online social network have on me, and what influence do I have as a node in that network? Can I influence my personality by forming specific (online) relationships? Is the Internet a special case for all existing research results on social networks? What are the specifics of online social networks that the social sciences provide? Does the Internet enhance existing or enable new social behavior? Does the mere size of a network or community make possible new effects that have not been possible before due to quantitative thresholds? What are parallels (and metaphors to describe them) between the real world and the online world?

It is obvious that the social sciences know a lot about social networks, but miss the technical expertise, a fact that needs to change (see the findings in Section 6). Yet the question is how social scientists (who have questions) can be brought together with information systems researchers (who have tools to answer these questions). What social sciences *can* contribute and study are questions like "what influence does my network have on me?" or "Can I form my personality through an architecture of social contacts?" Some people claim that the Internet as well as mobile devices fundamentally change the behavior of people and the way they communicate, and that the Internet is hence not just "yet another medium."[3]

For online networks as they exist today and the transparency they provide (which is much larger than it used to be prior to online networks), control and regulation are needed, yet how to do this (if it can be done at all) is still vastly unclear. Governance approaches are also needed for commercial networks. On the other hand, the added value of online social networks is undeniable.

The added value of online social networks can be discussed in multiple dimensions: The personal value of users, the commercialization value generated by platform providers (such as Facebook), and the value generated by businesses (such as brands that advertise on Facebook). However, from a provider perspective, commercialization does currently focus on simply-targeted advertisements. New business models in terms of bringing together supply and demand will appear and need to be researched in the future. For example, being able to develop social software using APIs such as the Facebook Platform or OpenSocial opens up a wealth of different business opportunities because businesses do not have to build a new social graph from scratch. Internet companies can exploit this, for instance, to boost their outreach and profile immensely — by positioning their existing product on other networking sites as a social application [10].

For now, a big question for businesses is how to attribute revenues from social media to the different touchpoints a customer has had with their brand or products. People tend to switch between distinct contexts all the time (reading email on their smartphones, browsing the Web, shopping online, participating in a chat, checking in on Foursquare etc.), but the interplay of digital touchpoints still needs to get on the research agenda. Also, it is largely unclear how user behaviour in social networks depends on the device used (e.g., PC,

[3] http://www.theshallowsbook.com/nicholascarr/Nicholas_Carrs_The_Shallows.html

smartphone, or tablet) and context of usage.

Physical presence has a distinct meaning in a social context. The same applies to shopping: Online shopping is different from going to a store. For example, it is usually more focused, yet there may be more impulses. *Servicescape* is a concept that was developed by Booms and Bitner to emphasize the impact of the physical environment in which a service process takes place.[4] As they state in their paper, "the ability of the physical environment to influence behaviors and to create an image is particularly apparent for service businesses such as hotels, restaurants, professional offices, banks, retail stores, and hospitals." But what holds for the physical world may as well apply to the virtual world; in other words, how can I find out that the world that Google is showing me is the real world?

Along similar lines of thought goes the observation made, for example, by N.A. Christakis (Harvard Medical School) that we influence our networks, but at the same time they send information and emotion back to us and are a source of infection to ideas and even diseases.[5]

5.4 The (Observable) Network Effects of Crowdsourcing

One of the most striking "network effects," besides the creation of large friends networks, is the already mentioned area of *crowdsourcing*. According to [1], "Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken." Crowdsourcing has been successfully applied to tasks that are easier to solve for a human than for a computer (e.g., image analysis), but also to many other areas, and it has meanwhile developed "subareas" such as *crowdfunding* or *crowdvoting*.

Topics in this field include large-scale cooperation, collaborative editing, constructivism via digital means, knowledge management, IT supported collaboration in logistics networks, and agent-based coordination. Questions here are the following: Which effects can be observed by employing crowdsourcing? In which areas has crowdsourcing failed up to now and why? Which new areas could benefit from crowdsourcing (technical ones such as query optimization, social ones such as crowdfunding)?

Collaboration between people is often seen as a form of art (e.g., in music[6] or social writing) and differs from the kind of collaboration we practice today. Most examples of crowdsourcing we see today are those which are working well. Different forms of crowdsourcing have had different successes (e.g., Galaxy Zoo[7] and Zooniverse[8]), and it turns out that creative work

[4] http://www.jstor.org/stable/1252042, http://en.wikipedia.org/wiki/Servicescape
[5] http://www.wjh.harvard.edu/soc/faculty/christakis
[6] See http://www.npr.org/2012/05/13/151712146/first-listen-hilary-hahn-and-hauschka-silfra for an example representing the taste of one the authors and http://www.inc.com/articles/201103/ted-collaborative-communication-social-media-age.html for a more general coverage.
[7] http://www.galaxyzoo.org/
[8] https://www.zooniverse.org/

requires a selection process for the crowd. This might lead to a revised notion of "crowd," e.g., "active," "participatory," or "unconsciously voluntary:" A member may be invited, participate actively, register by herself, or be used without knowing about it.[9] Data querying and analysis are increasingly seen as an application for crowdsourcing (a mixture of human computation and automation) [9]. Crowdsourcing is constantly producing process data and content data. Another emerging specialization that might arise in the future is the *private crowd* (e.g., inside an enterprise) vs. the *public crowd* (e.g., AWS Mechanical Turk).

Social sciences should take a leading role in the design of social networks, not just analyze them. The goal should be to make it easier for people to meet, get together, and let them do the rest themselves. For example, flashmobs minimize risk and maximize outcome; they are a low-level form of crowdsourcing.

There is a link between economics, social media, and human computation. It is about give and take, giving and receiving: I am willing to contribute to a social network because it, say, saves me an address book. People contribute because they expect a return. Another principle is self-sufficiency: In gaming, you may need data for a game you do not have; if your game is not self-sufficient, you constantly need to crowdsource to get more data. So data is an input to crowdsourcing.

More targeted social media are coming (especially to the workplace, in the sense of a given activity or given goals). For example, car navigation systems show the effects to be expected: They may help, but can overdo it, i.e., create a traffic jam they actually want to avoid.

There are also cases where the crowd is less efficient than a hierarchy. Examples include warfare (but there is also network-centric warfare) or emergencies. On the other hand, even social media might be designed in such a way that they incorporate hierarchy (such as Wikipedia). In some areas or applications crowdsourcing is not just inapplicable, but has failed or is (or has become) inappropriate, e.g., teaching material in a university program, military applications based on classified information, or generally applications requiring fast decisions. Crowdsourcing is not even useful in an arbitrary application, since it may destroy creativity or a vision in a given setting.

6 Findings

In this section we report on the findings of the workshop. We first review general findings which arose from plenary discussions following the group discussions. We then state a number of suggestions for future research, and finally summarize other findings that were brought about.

6.1 General Findings

The workshops has convincingly shown that more interaction and collaboration between the various fields represented here (and others who were not here) is needed. The fields need to identify a common level of language, tools and set of methodologies so that the various aspects of networking we discussed can be addressed and jointly developed further. Indeed, the most

[9] Matt Ridley describes in "When ideas have sex" that it's all about creating and sharing (he calls it trading), see http://www.ted.com/talks/matt_ridley_when_ideas_have_sex.html.

important point in our findings was the need for a renewed multi-disciplinarity. To a great extent, the networks we have discussed are driven and further developed by practitioners; which also means that they are evolving in a very fast manner and not emanating from a single scientific discipline. To be able to both understand them and contribute to the state of art, we need true inter- or multi-disciplinary research that involves computer science, social sciences, economics, and more. Much can be learned by viewing a network-based situation from an alternative disciplinary perspective.

A crucial issue in this context is grasping the dynamics of networks at a conceptual as well as a methodological level:

- Levers of change include technology, as it has proliferated across societies;
- spill-over effects across domains, e.g., the public, political, and commercial domain;
- counter-forces, dark networks show a similar dynamics;
- innovation and defective behavior: innovation is often driven by defective behavior, e.g. young people challenging the power of global media companies;
- methodologically, e.g., living labs.

Science is currently driven by the fast development and changing character of social networks. Taking into account the high relevance of understanding the dynamics of networks, only an inter-disciplinary view on the different aspects of networks could develop the chance to grasp the nature of networks dynamics. A methodological mesh of different approaches used in the various disciplines could be a promising way to tackle the numerous research questions. Furthermore the comparison of the different network characteristics (social, business, logistics, etc.) and investigating the possibilities of transferring principles between the different network types could bring up new ways of understanding and managing these networks.

Business networks will grow (we will see more and different shaped business networks), personal networks will change (Facebook in the future will not be as Facebook is right now). We will see an integration of business networks and personal networks (social networking platforms like Facebook, XING, LinkedIn etc. will be integrated parts of businesses and business networks). Law will not be able to cover all the implications of computer-supported networks and will loose its controlling function. Interdisciplinary research is necessary to recognize, to describe, to explain and, even more important, to design and to innovate social, supply-chain, administrative, business, commerce, and political networks.

We are stuck analyzing the present; we should start developing the future. We are also stuck in the silos of our disciplines. One crucial aspect that will help to change this situation is student training. Three key takeaways are:

1. We need to get ahead of the curve, i.e., instead of dealing with old networks, we need to understand how to monitor dynamically and predict the development of current and future networks at some appropriate level of abstraction.
2. The notion that methods working for offline networks (viz. process mining in supply chains and/or enterprises) can be used for exploring online mechanisms needs to be explored more.
3. The traditional economic models applied for networks may be from a wrong perspective; for example, the increasing interest (by commercial providers such as Factual, Socrata, or Kasabi, to name just a few) to view data as “goods” that have a price tag and that can be traded on a market may help to explain changes and predict needs for a Web free of data issues.

In particular, we should strive to make methods of IS research, such as business process management, decision support systems or data mining, better accessible for investigating phenomena of social networking.

A question immediately following from the discussions summarized above is whether we can use what is happening on the Web today in order to help to solve real world problems. Network research is a crucial piece of this, but it remains open — at least for the time being — how the different research areas and findings can be brought together.

Deriving solutions from the network effects that we observe today would ultimately result in the development of *social machines* [12]. These and other approaches such as *Games with a Purpose* (GWAP)[10] bring many disciplines together (analysis and theory, engineering and technology, social sciences and communications, economics and law etc.) [18]. We need methodologies for studying these approaches, their effects, their consequences, and their interplay (e.g., cyber-bullying).

The important observation is that networks cross all disciplinary boundaries. Research communities are discovering this at the moment, while funding agencies are not. In order to change this situation, it would be advisable to start by creating a common vocabulary and/or methodology to bring together the various disciplines required. An area that might also help here is (social) business process modeling (BPM) [16], since we encounter a situation similar to the "process modeling vs. process mining" debate. The challenge is to look at scalability in a network; the domain is more probabilistic and has other characteristics that these methodologies need to be adapted to. Petri nets could be applied, maybe in combination with hidden Markov chains. Also, we require an appropriate level of abstraction where process models can be useful. Past attempts to achieve this have usually challenged process technology.

6.2 Suggestions for Further Research

Over the next few years, network research will be particularly interesting, since many additional research prospects will emerge. However, for a range of reasons, interdisciplinary research will make little progress. These reasons include funding as well as challenges of mixed methods and/or disciplines. In the meantime, the number and scale of problems related to networks will be greater, so the gap will become bigger. There will be a growing number of research on singular, specific network-related topics, but less interdisciplinary than we, as a group, would appreciate.

Social networks comprise many different research topics in design, implementation, and application. We found several analogies in different sciences that could be used to understand and describe those networks (e.g., graph theory, neurology, sociologic models, etc.). Any cross-disciplinary research so far suffers from a missing common model (including glossary, models and toolkits) that grants a common ground for all different kind of research. Indeed, vocabularies and process models can be seen as two components of a "methodologies toolkit" that should be put together in an attempt to cross the boundaries of individual scientific areas. Other pieces to be included in this toolkit are network analysis, visualization, social simulation (focusing on trust, e.g., in a supply chain), methods to create trust, and provenance. Also, it is important to look at networks over time, not just at a particular instance or point in time.

Social scientists use psychological profiles and combine it with structural data, then use the result as a predictor. An example from [8] is cyber-bullying behavior at schools (where

[10] http://www.gwap.com; see also http://duolingo.com/, http://en.wikipedia.org/wiki/Human-based_computation_game

structural information is about friends and their relationships). A general question is how to design systems (e.g., learning systems such as Backstage [13]) in such a way that people would voluntarily use them since they can see the benefit for themselves. Yet, the real benefit is when you can initiate some kind of social dynamics and sustain it. There are cross-cutting topics where the various disciplines represented in this workshop can contribute, e.g. trust. Network dynamics could use a mechanism design to make social networks more resistant to tampering so that they cannot be manipulated (e.g., by first creating trust, then abusing it).

In five years from now, more people as well as organizations will use the infrastructure, and the richness of infrastructure will rise (Facebook and other social networking platforms will have more content; ERP will be a part of the infrastructure and no longer a specific part of a specific company). This infrastructure will be more vulnerable, and its governance structure will not be solved. Intellectual property and privacy will be even bigger issues in Western countries, but might be no issue at all in China. There will also be demographic change relating to medical infrastructure, decentralized usage of information systems, house systems, speed of growing new developments (e.g., the German Pirates party), and we will have more social norms instead of legal regulations.

6.3 Other Suggestions

Funding: As distinct disciplines grow together and embark on collaborative research, it is also important to convince funding agencies that multi-disciplinary research should arrive on their agendas.

Teaching: Web sciences need to be developed as a field, and also need to be integrated into teaching (as it is done in the WeST Institute at Koblenz,[11] for example). This will most likely lead to novel curricula which receive their content from multiple disciplines in a balanced way.

7 Concluding Remarks

We are at the dawn of a new way of doing research, namely detached from the fact that "I belong to a particular department", "I need to publish in certain journals", "I get funding only in my field". Instead, we are overcoming field boundaries and diving into other areas together with new people. That applies even to the workshop acceptance criteria at Dagstuhl itself. We need a problem-oriented approach to get away from silo thinking: What is the problem? What expertise is needed to solve it?

Web-based systems will transform society: large numbers of users can interact; the available technology enables communities to build and run their own social machines. For a platform to be successful, it should not crack or allow for a bad experience, which requires more than a research prototype. Instead, we need professional software developers. We also need advertising, which is not funded either. The emerging area of *Web Science*[12] has apparently recognized this need, and is working in various ways on bridging the gaps between disciplines [11].

[11] http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab
[12] http://webscience.org, http://eprints.soton.ac.uk/265186/1/metadataisthemessage.pdf, http://journals.cambridge.org/action/displaySpecialPage?pageId=3656

Intuitions we have on certain aspects seem to be wrong most of the time; this is what we see in the network domain (e.g. growth of Facebook). Now that fields are starting to converge, this is even more true. Therefore we need to start looking into the real problem. Economics enjoy modeling, but at the price of complexity reduction. In the micro/meso/macro layer setting, we need to analyze the dynamics between these layers. Studying dynamics becomes even harder that way. A good example is human-computer interaction (HCI). Interfaces were studied for years because they stood still during that time; that is not the case anymore. Now you have to know your users in advance when you build a system. HCI has developed methodologies, which have broken down. This example is about speed of change, not networking. Also in other examples, speed of change is a big issue.

We consider to arrange another workshop to discuss the various methodologies that are used in the different fields of research.

Acknowledgments. This report is the result of Dagstuhl Perspectives Workshop 12182 "Social, Supply-Chain, Administrative, Business, Commerce, Political Networks: A Multi-Discipline Perspective". The editors are grateful to the other participants Jörg Becker, Daniel Beverungen, François Bry, Clemens Cap, Ingo Dahm, Stuart Dillon, Emese Domahidi, Jim Hendler, Bernd Hellingrath, Stefan Klein, Nicolas Pflanzl and Michael Räckers, who have all contributed to this report; we are also grateful to Dagstuhl for providing such a stimulating atmosphere.

References

1 Enrique Estellés Arolas and Fernando González-Ladrón de Guevara. Towards an integrated crowdsourcing definition. *J. Information Science*, 38(2):189–200, 2012.
2 Jörg Becker, Daniel Beverungen, Patrick Delfmann, and Michael Räckers. Network e-volution. *Working Paper No. 11, European Research Center for Information Systems, Münster, Germany*, 2011.
3 Tim Berners-Lee and Mark Fischetti. *Weaving the web - the original design and ultimate destiny of the World Wide Web by its inventor.* HarperBusiness, 2000.
4 Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.
5 Peter Buneman, James Cheney, Sam Lindley, and Heiko Müller. Dbwiki: a structured wiki for curated data and collaborative data management. In Sellis et al. [17], pages 1335–1338.
6 Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Why and where: A characterization of data provenance. In Jan Van den Bussche and Victor Vianu, editors, *ICDT*, volume 1973 of *Lecture Notes in Computer Science*, pages 316–330. Springer, 2001.
7 Peter Buneman and Wang Chiew Tan. Provenance in databases. In Chee Yong Chan, Beng Chin Ooi, and Aoying Zhou, editors, *SIGMOD Conference*, pages 1171–1173. ACM, 2007.
8 Ruth Festl and Thorsten Quandt. Social relations and cyberbullying: The influence of individual and structural attributes on victimization and perpetration via the internet. *Human Communication Research*, (accepted for publication), 2013.
9 Michael J. Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh, and Reynold Xin. Crowddb: answering queries with crowdsourcing. In Sellis et al. [17], pages 61–72.
10 Matthias Häsel. Opensocial: an enabler for social applications on the web. *Commun. ACM*, 54(1):139–144, January 2011.
11 James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. Web science: an interdisciplinary approach to understanding the web. *Commun. ACM*, 51(7):60–69, July 2008.

12 Jim Hendler and Tim Berners-Lee. From the semantic web to social machines: A research challenge for ai on the world wide web. *Artif. Intell.*, 174(2):156–161, 2010.
13 Alexander Pohl, Vera Gehlen-Baum, and François Bry. Introducing backstage - a digital backchannel for large class lectures. *Interact. Techn. Smart Edu.*, 8(3):186–200, 2011.
14 Thorsten Quandt. What's left of trust in a network society? an evolutionary model and critical discussion of trust and societal communication. *The European Journal of Communication*, 27(1):7–21, 2012.
15 Stefano Rizzi. Data warehouse. In Benjamin W. Wah, editor, *Wiley Encyclopedia of Computer Science and Engineering*. John Wiley & Sons, Inc., 2008.
16 Frank Schönthaler, Gottfried Vossen, Andreas Oberweis, and Thomas Karle, editors. *Business Processes for Business Communities - Modeling Languages, Methods, Tools*. Springer, 2012.
17 Timos K. Sellis, Renée J. Miller, Anastasios Kementsietsidis, and Yannis Velegrakis, editors. *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011*. ACM, 2011.
18 Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneißl, and François Bry. Karido: A gwap for telling artworks apart. In Quasim H. Mehdi, Adel Said Elmaghraby, Robert Moreton, Roman V. Yampolskiy, and Julia Chariker, editors, *CGAMES*, pages 193–200. IEEE, 2011.
19 Gottfried Vossen and Stephan Hagemann. *Unleashing Web 2.0: From Concepts to Creativity*. Morgan Kaufman, 2007.

Participants

- Jörg Becker
Universität Münster, DE
- Daniel Beverungen
Universität Münster, DE
- Franççois Bry
LMU München, DE
- Clemens Cap
Universität Rostock, DE
- Ingo Dahm
Deutsche Telekom – Bonn, DE
- Stuart Dillon
University of Waikato, NZ
- Emese Domahidi
Universität Hohenheim, DE
- Matthias Häsel
Otto Group – Hamburg, DE
- Bernd Hellingrath
Universität Münster, DE
- James A. Hendler
Rensselaer Polytechnic, US
- Stefan Klein
Universität Münster, DE
- Nicolas Pflanzl
Universität Münster, DE
- Thorsten Quandt
Universität Hohenheim, DE
- Michael Räckers
Universität Münster, DE
- Gottfried Vossen
Universität Münster, DE

Report from Dagstuhl Seminar 12191

Artificial and Computational Intelligence in Games

Edited by

Simon M. Lucas[1], Michael Mateas[2], Mike Preuss[3], Pieter Spronck[4], and Julian Togelius[5]

1 **University of Essex, GB,** sml@essex.ac.uk
2 **University of California – Santa Cruz, US,** michaelm@cs.ucsc.edu
3 **TU Dortmund, DE,** mike.preuss@tu-dortmund.de
4 **Tilburg University, NL,** p.spronck@uvt.nl
5 **TU Dortmund, DE,** julian@togelius.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 12191 "Artificial and Computational Intelligence in Games". The aim for the seminar was to bring together creative experts in an intensive meeting with the common goals of gaining a deeper understanding of various aspects of artificial and computational intelligence in games, to help identify the main challenges in game AI research and the most promising venues to deal with them. This was accomplished mainly by means of workgroups on 14 different topics (ranging from search, learning, and modeling to architectures, narratives, and evaluation), and plenary discussions on the results of the workgroups. This report presents the conclusions that each of the workgroups reached. We also added short descriptions of the few talks that were unrelated to any of the workgroups.

Seminar 6.–11. May, 2012 – www.dagstuhl.de/12191
1998 ACM Subject Classification I.2.1 Applications and Expert Systems.
Keywords and phrases Artificial Intelligence, Computational Intelligence, Computer Games
Digital Object Identifier 10.4230/DagRep.2.5.43
Edited in cooperation with Nicola Hochstrate

1 Executive Summary

Simon M. Lucas
Michael Mateas
Mike Preuss
Pieter Spronck
Julian Togelius

The video game industry is the largest of the entertainment industries and growing rapidly. The foundations of this industry are techniques from computer science. New developments within video games pose fresh challenges to computer scientists. Around the world, the number of dedicated study programs producing the workforce of the game industry is increasing steadily, as is the number of computer science academics dedicating their careers to solving problems and developing algorithms related to video games. Such problems often require domain knowledge from various research domains, such as psychology and the arts, leading to an inherently interdisciplinary research field.

Artificial intelligence (AI) and computational intelligence (CI), in one form or another, can be found at the heart of almost any video game, controlling the non-player characters (NPCs) as well as many aspects of the game world. They are also used throughout the game design and development process. Academic research within these domains in games aims to solve problems and enable innovation, pertaining to game design, game development, and gameplay. A main focus is on solving algorithmic problems to make game mechanisms more intelligent and efficient, thus making games more immersive, interesting, and entertaining. In the context of serious and educational games, such improvements enable these games to fulfill their societal objectives better.

Artificial intelligence seeks to simulate intelligent behavior in any possible way with human intelligence as a paradigm. Computational intelligence is an umbrella term for nature-inspired computational methods for optimization, learning and controlling. The main methods are evolutionary algorithms, artificial neural networks, fuzzy logic, swarm intelligence, and artificial immune systems. Nowadays, the borders between both disciplines are blurred, and state-of-the-art solutions use hybrid techniques combining elements of symbolical AI systems, CI algorithms and methods from statistical machine learning.

The aim for the *Dagstuhl Seminar on Artificial and Computational Intelligence in Games* was to bring together creative experts in an intensive meeting with the common goals of gaining a deeper understanding of various aspects of games, and of further improving games. It was meant to enforce the communication of different communities and the collaboration with the games industry. The exchange of different views and competencies was to help identify the main challenges in game AI research and the most promising venues to deal with them. This could lead to a common vision of what kind of games could be made possible in the future.

The Seminar was held from Sunday, May 6, 2012, until Friday, May 11, 2012. Over 40 researchers came together at Schloß Dagstuhl, many of them highly-respected and well-known researchers in their field, but also several talented young researchers and even a few representatives from the AI specialists of the game industry. In contrast to what is common for such gatherings, very little time was spent on plenary talks. Instead, the focus was on workgroups which discussed particular topics. However, several plenary sessions were held in which the workgroups reported on their results, and new topics for discussion were brought up. To allow researchers to present their recent work, a poster session was held during the second day of the Seminar, and the posters remained up until the end.

The topics of the workgroups, in alphabetical order, were the following:

- AI Architectures for Commercial Games
- AI Clearing House
- AI for Modern Board Games
- Computational Narratives
- Evaluating Game Research
- Game AI for Mobile Devices
- General Game Playing
- Learning in Games
- Pathfinding
- Player Modeling
- Procedural Content Generation
- Search
- Social Simulation Games
- Video Game Description Languages

In the remainder of this report, abstracts for all the workgroups, describing their discussions and results, are given. Several researchers wrote a short report on their poster, which are included too. We aim to bring a full reports of all the workgroups in the form of proceedings later.

As organizers we are really pleased with how the Seminar turned out. It proved to be the stimulating and inspirational environment that we had hoped for. We found that most, if not all participants agreed with us on that. A lot of this success is due to the excellent facilities provided by the people of Schloß Dagstuhl. We are highly grateful for having had the opportunity to be their guests for the Seminar. We definitely hope to return in the future.

Simon M. Lucas
Michael Mateas
Mike Preuss
Pieter Spronck
Julian Togelius

2 Table of Contents

Overview of Short Talks

3 Overview of Workgroups

This section contains an overview of the results of each of the 14 workgroups.

3.1 Pathfinding in Games

Adi Botea, Christian Bauckhage, Bruno Bouzy, Michael Buro, and Dana Nau

Pathfinding remains one of the most important applications of artificial intelligence in commercial games. The problem has received a significant attention in recent years, both from academic researchers and game developers. In the past, A* search, combined with a heuristic function such as the Manhattan distance, and performed on grid-based map representations, used to be a common approach to computing paths in video games.

In recent years, there has been significant progress in this area: much faster path computation methods based on hierarchical abstractions, more informed, memory-based heuristic functions, symmetry reduction, triangulation-based map representations, and compressed databases with all-pairs shortest paths information. Some of the novel techniques, such as contributions to hierarchical abstraction and triangular map encoding, have already been adopted in video games.

Despite these impressive advances, there are many challenges to be addressed in pathfinding in video games. Previous work often made a number of restrictive assumptions, such as a fully known, static environment in which units are often treated as independent from each other. Such a narrow, single-agent view shows its limits in both collaborative multi-agent pathfinding, and adversarial pathfinding. For example, units can be trapped in deadlocks or collide with each other during path execution. Also, taking physical constraints, such as the speed and minimal turning angle of units, into account, can greatly increase the realism of generated paths. Moreover, ever growing map sizes and speed and memory limitations of game consoles demand more efficient pathfinding algorithms.

Another challenge is to increase academia–industry collaboration. Progress in this direction could be facilitated by ensuring that benchmark problems that are used to evaluate academic research overlap more with problems that are truly relevant to game developers. An example is multi-agent pathfinding, where recent academic contributions have focused on either providing optimal solutions or scaling suboptimal methods to hundreds or even thousands of units. While important on their own, such objectives are not necessarily the most critical ones in current games.

The pathfinding workgroup at the Dagstuhl seminar 12191 has focused on acquiring a snapshot of current pathfinding approaches and identifying important topics for a future research roadmap. In this abstract we have summarized the main conclusions of the workgroup.

3.2 Search in Real-Time Video Games

Peter Cowling, Michael Buro, Bruno Bouzy, Dana Nau, Martin Butz, Philip Hingston, Adi Botea, Moshe Sipper, Hector Muñoz-Avila, and Michal Bida

It is fascinating to speculate as to whether the search approaches that have been so successful for Chess, Go, Checkers, etc. can be used to create strong / interesting / robust players for video games. This large and very active / lively group met several times during the Dagstuhl seminar week to investigate the current state of the art, the benefits and challenges in this area, and promising techniques and directions for future research. This report provides a brief overview of the main points of our discussion, which will soon be extended to a much more detailed summary and road map for future research.

The benefits of using search in creating AI players for turn-based board and card games are readily apparent – with chess programs running on personal computers not losing against world-champions anymore, and Go programs playing at expert level. Real-time video games, however, provide a range of new challenges, and effective search-based players for these games could provide a new dimension in gameplay. For example, a new generation of AI players may develop for complex, real-time strategy and first-person games. Restricting the information available to the AI player to the information a human player may be able to gather in the game, and restricting the actions of the AI player to human player actions (executed in time and space) will make AI players more human-like and also more appealing to human players. Developing search techniques, in such game scenarios, yielding AI players that continue to be effective and interesting against any strategy adopted by a human opponent currently poses a fundamental challenge to the community.

In addition to the potential of transferring search techniques from game to game, greatly reducing development time, new game components may be developed in which lower-level game tasks are accomplished by AI-based search techniques. For instance, using a hierarchy of search levels, a command hierarchy could be simulated for games where a player controls many agents, giving high-level strategy, and recognizable tactics to work alongside effective and realistic low-level behaviors.

In general, the issue for real-time strategy and first-person games is to provide “strong” non-cheating players, and the questions whether a player is "strong" and whether gameplay is “interesting” are closely linked (by the fact that strong players can be adapted to lower playing strength). Hence, the focus of the group was on maximizing AI player strength – which may also provide an easier set of metrics for judging progress in AI in games research. Techniques developed for strategic video games also have much potential for serious games and military simulations.

The use of search allows strategy to emerge from reusable components:

- search algorithms
- knowledge (such as tactical notions) embodied in state evaluation functions and scripted policies
- tuning approaches such as evolutionary algorithms, reinforcement learning, and data mining

Search Techniques

The state of the art in search algorithms includes Monte Carlo Tree Search (MCTS), which shows tremendous promise for games where it is difficult to estimate the value of an intermediate state, such as Go and Hex, as well as showing great promise for hidden information card and board games. Alpha-Beta minimax search has yielded world-champion level players for a range of perfect-information board games including Chess, Checkers, and Othello, as well as high level play in games with hidden information such as Bridge. A* search is the algorithm of choice for a range of pathfinding and other planning problems. Heuristics and Evolutionary Algorithms have used analogies with human and natural problem solving approaches to achieve excellent results for optimization problems and single-player games.

A number of approaches have been developed to capture strategy and knowledge, and to deal with the real-time nature of decision making. These include Hierarchical Task Networks (HTNs), which show much promise as they capture strategy in a form within which search algorithms can be applied. Time slicing is a simple approach for dealing with real-time decision making, which has shown effectiveness in simple real-time domains. Hierarchical and level-of-detail approaches allow the computational budget to be allocated where most needed. Dynamic Scripting and "Black Box" approaches, which use learning and search techniques to choose between a range of scripted approaches, also show promise.

Challenges

The challenges in this area are manifold. Particularly in comparison with turn-based games, the most obvious challenge is the massive branching factor and depth that comes from simulating a continuous environment, where decisions can be made at any time (or at least at 20 frames per second). Manually capturing an abstraction of states and actions in such a space is difficult, and automatically finding abstractions poses even a greater challenge. Any abstractions should ideally be "continuous", i.e., preserving the property that nearby intermediate states lead to similar outcomes. In order to use search, abstractions must also allow simulations forward in time. Moreover, most strategic video games have multiple players and asymmetric hidden information, requiring the investigation of new search algorithms. It is also highly desirable that the decisions made by search are robust (in terms of varying opponents) and explainable to convince video companies to integrate search-based techniques in their products. Finally, the search must use hardware that, although becoming increasingly powerful, provides stringent limits on how complex the applied abstractions may be and how deeply/broadly the game tree may be searched while still adhering to the real-time game constraints.

A number of existing techniques represent promising directions for future research. Hierarchical Task Networks (HTNs) can be used to reason and represent high-level decisions while other representational/reasoning methods can be used for low-level decision-making. HTNs can be obtained using a combination of manual approaches and machine learning. The use of search trees with simple atomic actions (such as Group() and Move()) provides a framework for thinking about these lower level problems as well as providing a controllable level of complexity for pathfinding and tactics. Considering game theoretic techniques at an appropriate level of complexity allows for an effective way of handling hidden information and opponent models. Maintaining the complete tree (at least at an abstract level) but partitioning it into "similar" states provides a promising general- purpose direction for dealing with complexity and hidden information. Maintaining trees of strategic choices between other "black box" strategies (e.g. manually generated rule-based ones) also shows much promise.

Further investigations in the areas discussed in this paragraph are likely to indicate many other promising directions in turn.

Outlook

In the future we can imagine video games where AI strategies based on search provide strong, interesting gameplay against any possible strategy, removing the possibility of uninteresting tactics typically used to exploit AI limitations. These general-purpose approaches are expected to be easily transferable from game to game, providing bottom-up abstractions to find high-level strategies, given only information about game rules and physics. For success, however, hybrid and hierarchical approaches will likely be needed. Fast simulations of game states will need to be available as part of this general-purpose approach as well, yielding diverse controllable strategies through parametrization, and explainability through information about the part of the tree searched and simulation outcomes. Machine learning, particularly mining the data from games played between humans, should allow the automatic capture of strategies using tools such as Hierarchical Task Networks. Evolutionary, machine learning, and heuristic approaches will allow opponent strategies to be modeled and exotic new strategies to be found – also in order to extend human gameplay interest. Rather than AI players that follow rather predictable strategies and that were designed in advance, search will allow for flexible AI whose decisions will be as interesting and varied as human expert players. Difficulty levels will be tuned using varying search time and abstraction properties.

To achieve research success, the development and widespread adoption of competitions and benchmarks of increasing complexity will be necessary. There are promising developments here – competitions such as the current Starcraft AI competition [1] provide a platform for researchers to interact with a commercial video game. It is hoped that more competitions will be developed and more video games will allow open access to their game information.

There are links here with other groups at Dagstuhl: the modern board games group (which is investigating highly complex, multiplayer, hidden information games, but with discrete turns), the General Video Game Playing group (which is investigating a platform for studying general- purpose techniques for video games), and the Learning group (which focuses even more on the learning challenge in game environments).

3.3 AI and CI Games for Mobile Devices

Philip F. Hingston, Clare Bates Congdon, and Graham Kendall

This group considered the possibilities of creating new and innovative games that are targeted at mobile devices, such as smart phones and tablets, and that showcase AI (Artificial Intelligence) and CI (Computational Intelligence) approaches. Such games might take advantage of the sensors and facilities that are not available on other platforms, or might simply rely on the “app culture” to facilitate getting the games into users’ hands. While these games might be profitable in themselves, the focus of our discussion was on the benefits and challenges of developing AI and CI games for mobile devices.

[1] http://webdocs.cs.ualberta.ca/ cdavid/starcraftaicomp/

We recognize that mobile games are easier to bring to market than commercial (large scale) video games. This makes them a practical choice for development and study in an academic environment, using relatively small teams of academics and students, who are able to work on relatively low budgets. For example, the relatively small screen size and lack of powerful graphics hardware typical of mobile devices means that simple graphics, often only 2 or 2.5 dimensional are expected, so that large teams of highly skilled artists and 3D modelers are not required. Furthermore, mobile devices usually provide a wider variety of input data (touch, location, images, video, sound, acceleration, orientation, personal data, data from/about other users etc.) and offer more output options (images, video, animation, sound, vibration, wireless, bluetooth, infrared) than is normally available on a desktop or laptop computer. In addition, the popularity of mobile devices provides a means to recruit large numbers of casual users, which provides another potentially large data source. Novel game mechanics and interaction methods might be made possible by processing these input data using AI and CI methodologies.

Computational power and battery life present two potential obstacles to intensive AI/CI-based games, and some potential designs will require offloading some of the computation to servers. It might also be difficult to implement large-scale, complex game worlds due to the limited resources that are available. There are also significant challenges in developing AI/CI libraries that can work with low memory, limited battery power etc., adapting or developing AI/CI methods to work effectively in games that are played in short bursts, using unreliable communications, and providing real-time responses. However, these constraints provide significant research opportunities.

In summary, our group felt that mobile devices are still "young" enough to provide opportunities for developers to implement innovative products without having to employ large specialist teams (e.g. graphic designers, musicians etc.) This opportunity will not last long (perhaps less than two years) so those who are interested in this area might want to explore it now. Moreover, CI/AI both provide significant opportunities both in terms of research challenges and also to make the games more interesting and more fun to play. We would like to see the research community take up the challenge to showcase what can be done with the limited resources available on mobile devices, but also utilizing the larger number of sensors (e.g. movement detection) and other options (e.g. location awareness) which are not available on traditional "living room" game consoles.

3.4 Generalized Video Game Playing

John Levine, Clare Bates Congdon, Michal Bida, Marc Ebner, Graham Kendall, Simon Lucas, Risto Miikkulainen, Tom Schaul, and Tommy Thompson

One of the grand challenges of AI is to create general intelligence: an agent that can excel at many tasks, not just one. In the area of games, this has given rise to the challenge of General Game Playing (GGP). In GGP, the game (typically a turn-taking board game) is defined declaratively in terms of the logic of the game (what happens when a move is made, how the scoring system works, how the winner is declared, and so on). The AI player then has to work out how to play the game and how to win. In this work, we seek to extend the idea of General Game Playing into the realm of video games, thus forming the area of

General Video Game Playing (GVGP). In GVGP, computational agents will be asked to play video games that they have not seen before. At the minimum, the agent will be given the current state of the world and told what actions are applicable. Every game tick the agent will have to decide on its action, and the state will be updated, taking into account the actions of the other agents in the game and the game physics. We envisage running a competition based on GVGP playing, using arcade-style (e.g. similar to Atari 2600) games as our starting point. These games are rich enough to be a formidable challenge to a GVGP agent, without introducing unnecessary complexity. The competition that we envisage could have a number of tracks, based on the form of the state (frame buffer or object model) and whether or not a forward model of action execution is available. We propose that the existing Physical Traveling Salesman (PTSP) software could be extended for our purposes and that a variety of GVGP games could be created in this framework by AI and Games students and other developers. Beyond this, we envisage the development of a Video Game Description Language (VGDL) as a way of concisely specifying video games. For the competition, we see this as being an interesting challenge in terms of deliberative search, machine learning and transfer of existing knowledge into new domains.

3.5 Artificial Intelligence Architectures for Games

Daniele Loiacono, Michal Bida, and Alex Champandard

Despite the differences in technology and the unique designs of specific games, the Artificial Intelligence architectures for commercial titles are based on similar principles. Although in the literature several efforts have been done to define a standard AI architecture for games, the field suffers from a lack of consensus about the best approach to follow. There may not even be a single best approach to follow, but by identifying underlying patterns we can make progress nonetheless.

In particular, two major approaches can be identified when it comes to defining AI architectures for games: an input-output taxonomy and a design based on different abstraction layers. Both the approaches involve modular and layered architectures. In the first approach, the components are typically organized with (i) an input layer, which processes the data from the game environment; (ii) a middle layer (or more middle layers), which elaborates processed data and takes decisions; (iii) an output layer, which map decisions into actions in the game environment. In contrast, the second approach consists of organizing the components of the architectures in different abstraction layers: (i) the bottom layer deals directly with the game engine; (ii) the middle layer(s) provides a convenient abstraction of the game environment; (iii) the top layer deals with reasoning and high-level decisions.

Rather than introducing an alternative approach to design AI architectures for games, we propose a taxonomy of the typical components they involve. To this purpose, we combined the underlying principles of the approaches previously described: components of the architecture are classified with respect to both their level of abstraction and their role in the input-output pipeline. Our aim is to support the analysis and the understanding of the existing middleware, commercial games and academic frameworks. The proposed taxonomy might be used to understand which components are provided by a given middleware, which is the standard technology to implement a specific component, how the gameplay affects the design of the AI architecture, etc. In addition, it might also allow to identify the major challenges in specific areas and to discover opportunities for the game research.

As an example, we selected a set of typical components involved in an AI architecture for games and discussed about the major challenges they provide; then, we identified the most promising areas in our taxonomy not yet fully exploited in existing frameworks; finally, we used our taxonomy to classify and to compare popular middleware and games.

3.6 Believable Agents and Social Simulations

Michael Mateas, Elizabeth André, Ruth Aylett, Mirjam P. Eladhari, Richard Evans, Ana Paiva, Mike Preuss, and Michael Young

The topics for this workgroup were believable agents and social simulations in digital games. The environments discussed were interactive short form dramas and massively multi-player online (MMO) games. For the MMOs our vision was to populate them with rich characters, massive interaction affordances, and drama in every situation. When envisioning short form dramas we saw massively repayable dramatic situations as well as many actions available at any moment. We discussed aspects of the experience, and games where the player has rich, near human bandwidth (symmetric interaction). When discussing affective engagement we focused on ways to evoke a much broader range of emotions.

Our discussions addressed three main topics: scaling up, believable characters, and social modeling. For each of these topics we identified the main challenges and opportunities.

Scaling up – The Authoring Wall

When using traditional game development methods the authoring effort becomes unmanageable if the game world becomes very big or complex as illustrated in Figure 1. It becomes difficult to control the dynamism so that the players' experiences are enjoyable and make sense. An important challenge for research in AI is to find ways to overcome the authoring wall, allowing increased dynamism in future games. A challenge for researchers in this area is that, for only moderately dynamic experiences, traditional game development techniques (finite state machines, scripting, etc.) take less authoring effort than using generative AI techniques. So in building research games, it can be difficult to demonstrate that the AI is enabling a previously impossible game experience except by building a highly dynamic, fully playable experience, something many research teams lack the resources and experience to do.

Opportunities for moving beyond the authoring wall include developing techniques for crowd-sourcing and supporting user-generated AI content. Authoring tool for non-experts is another approach, with Inform 7 serving as an existence proof that a highly declarative, rule-based authoring paradigm can be made accessible to writers without computer science training. We noted that declarative representations scale more readily than procedural ones. There is the opportunity to learn lessons from Knowledge Engineering tools in the 1980s, which encoded domain structure to guide knowledge acquisition (authoring).

Believable Characters

We identified the current state of the art in the development of believable characters to be:
- use of planning models – reactive and traditional;

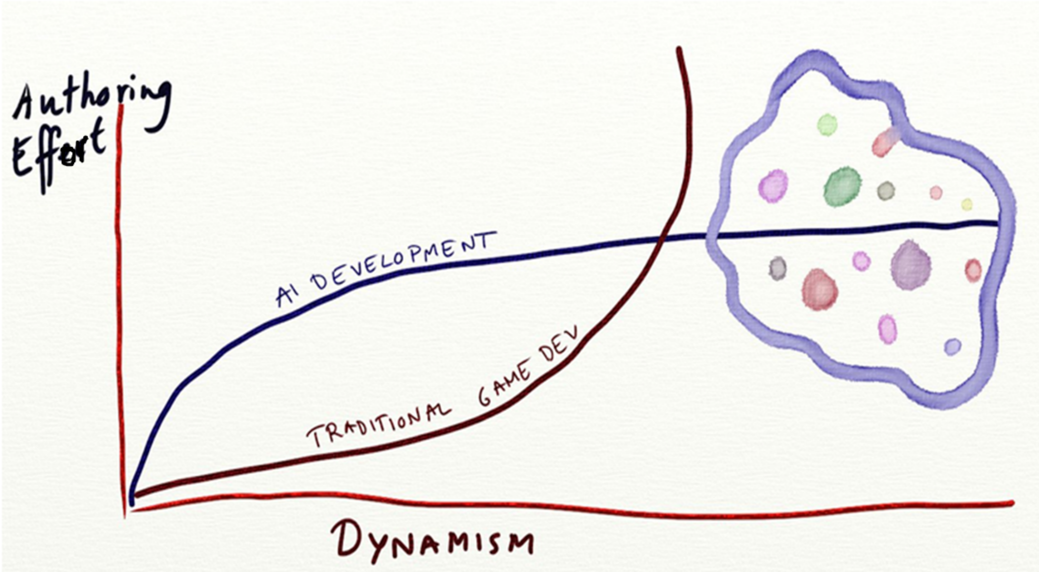

Figure 1 The Authoring Wall

- use of plan-based representations of physical, social and language actions (domain specific hierarchical planning);
- emotion modeling with convergence on cognitive appraisal architectures;
- FSMs (returning for use in dialogue- moving towards performance of interactive dialogue);
- personality captured through traits that modulate behavior and expression;
- personality-trait models less directly tied to expression but strongly correlated with multiple cognitive processes (so are more generally applicable across capabilities);
- mood modeled as a function of emotions (sums, averages) and with decay;
- modeling mood through spreading activation. When a character is reminded of something that is emotionally significant, there is an emotional echo; and
- episodic memory supporting temporal queries, evaluative labeling, and generalization.

We noted that most agent learning focuses on easy-to-evaluate criteria. For believable characters, significant challenges remain for making learning personality-specific and maintaining believability while online learning engages in exploration.

Opportunities for future development of believable characters include:

- to combine statistical language model for style with semantic or symbolic content models;
- dynamically generated character soundtrack communicating character and social state;
- two directions for more information-rich signals from:
  - player: high-bandwidth naturalistic interaction (gestures, gaze, prosody); new communication modes (biometrics, out-of-band interactions like music selection, player modeling);
  - purposefully set up choices for characters to allow them
    - to express personality through choice, and
    - conversely set up player choices that give information about player;
- take advantage of low-cost motion capture to correlate prosodics and gesture features in generative models (has to be parameterized by emotion and personality and social context); and
- using explainable AI to drive interface elements (text explanations) or thought bubbles.

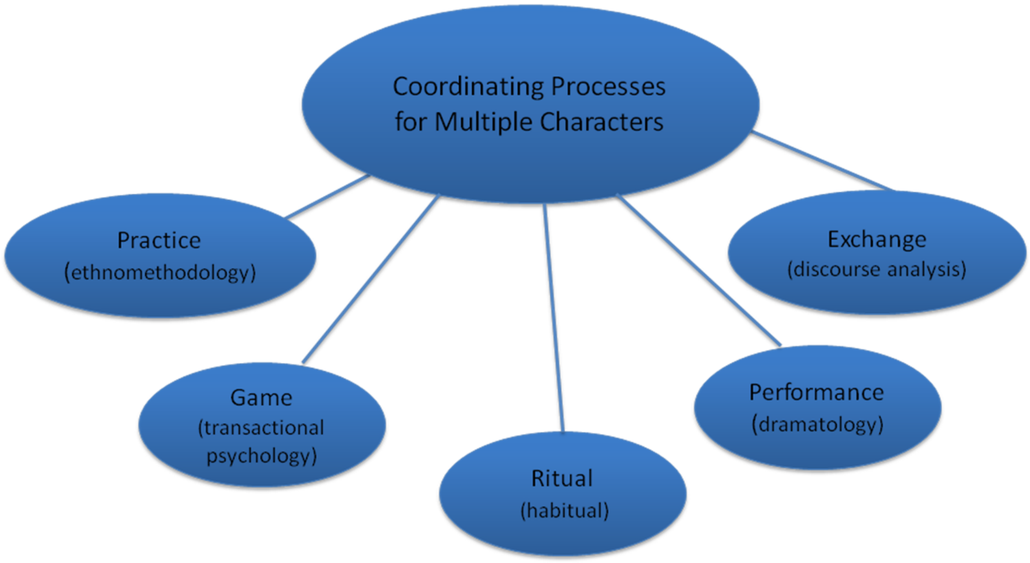

Figure 2 Coordinating Processes of Multiple Characters

Regarding expression of believable agents we recognized several limitations:

- the rich emotional models that are computed reside in the side-scenes and are not currently being expressed to players;
- multi-modal expression often results in inconsistencies between modes (uncanny); and
- handcrafted art- and audio assets currently work best (as opposed to procedurally created content), but create an authoring bottleneck.

Opportunities we recognized regarding expression were:

- to go beyond naturalistic world simulation affordances to express character state;
- reifying state in characters and objects;
- behavior explanation (HUD); and
- use of stylized comic-inspired forms, music, abstract visuals.

Social Modelling

In the current state of the art of social modeling we are converging on architectures with multiple social practices above the agent level, but using different approaches of which some are noted in Figure 2. In general, work on social modeling is far less developed than work on believable characters who engage in limited social interactions.

Challenges in the field of social modeling include addressing that we are unable to share social practices between systems and that there is no common theoretical basis for social practices (and no common representational formalism).

Opportunities for social modeling include:

- make connections with agent-based social simulations;
- make a detailed comparison between different social models to compare how they represent the same social practice; and
- use a situation which has been tagged with social-annotations as a common target for comparing the different formalisms.

3.7 Learning and Game AI

Hector Muñoz-Avila, Michal Bida, Graham Kendall, Christian Bauckhage, and Clare Bates Congdon

We focused on learning aspects in current computer games, challenges, and opportunities for future applications. Unlike the General Video Game Playing group, which focused on an unified environment for a class of games, we looked at the broader issues with learning independently of languages and platforms. Unlike the planning and search group, we considered learning even in the event in which no state-action transition models are given including topics related to data mining and pattern recognition of playing logs.

State of the art

In our discussions we observed that there are very few applications of machine learning to commercial games. Those of us who teach courses or give lectures on the topic, tend to focus on the few well- documented success stories (e.g., the use of induction of decision trees in the commercial game Black and White). There is a number of noncommercial applications of machine learning to game (e.g., the use of reinforcement learning to play Backgammon). The use of machine learning to find patterns from network and log data has demonstrated significant potential (insert- Christian-paper-here). Also there is significant research, demonstrating the use of Monte Carlo techniques and evolutionary computation to evolve rules for high-performance for arcade games (cite-Clare-paper(s)-here), to learn

Challenges

We identify five kinds of challenges.

1. Need to explain decisions. It will be desirable for machine learning algorithms to explain its decisions. Whereas in non-commercial applications such as determining if a credit application to be granted or not, it is fine if the algorithm doesn't explain itself (because a generic explanation to the customer suffices), in the gaming context, poor performance in a small number of situations will be magnified through social media and player-to-player communication. In addition, assessing quality is very difficult because the search space in many commercial games (e.g., a real-time strategy game) is too large.
2. Selecting machine learning technique. Another challenge is that there is no simple answer if a game developer asks the question about which machine learning to use. In other areas (e.g., pathfinding), the capabilities and limitations of some techniques relative to others is well understood.
3. Obtaining the input data. Getting the data for input to test the machine learning algorithms can be difficult. There is no clear value added for a commercial company to gather and share the data.
4. Need to demonstrate value added. If a gaming company has money to invest in a game, aspects such as graphics will get prioritized simply because it is unclear what the benefit from investing on machine learning is.
5. In some situations adaptable AI might not be desirable. In games that for reasons such as commercial considerations, the expected time the game will be played is bounded to, say, 20 hours, having adaptable AI can make it replayable for a long time and hence it might be undesirable to have those capabilities.

Opportunities

We identify a number of opportunities for machine learning techniques including:
1. Balancing gaming elements. Many games have different elements such as factions in a real-time strategy games (e.g., humans versus orcs) or classes in a role-playing game (e.g., mages versus warriors). Machine learning could help with balancing these elements.
2. Balancing game difficulty. Games, particularly those that are open- ended such as massive multiplayer online (MMO) games. A difficulty is how to tailor the game simultaneously towards dedicated players (e.g., players who play 20+ hours per week) and casual players (e.g., players who play 10 hours or less a week).
3. Finding loopholes in games. Pattern recognition techniques can be used to first detect usual patterns in game logs and then use these patterns to detect outliers. Such techniques will enable to detect anomalies (e.g. exploits in MMOs) much faster than it is currently done, which is manually for the most part.
4. In some situations adaptable AI is highly desirable. In open ended games such as MMOs there is a need to extend the playtime. Providing adaptable AI can be an important contributing factor to extend the game AI.

3.8 AI for Modern Board Games

Pieter Spronck, Peter Cowling, Alex Champandard, Pier Luca Lanzi, and Ana Paiva

Modern board games are recently developed board or card games, which usually involve more than 2 players, and often contain non-deterministic elements and/or imperfect information. Examples of such games are Monopoly, Poker, Settlers of Catan, Dominion, Puerto Rico, Agricola, and Power Grid. In the last decade, tens of thousands of new board games have been published. Information on all those games, including rulesets, can be found on such websites as boardgamegeek.com and pagat.com.

Research into AI for classic, 2-player, deterministic, perfect information games such as Chess and Go in the last 50 years has lead to highly capable computer players which outrank human grandmasters in many games. The techniques developed for this research (in particular, tree-search techniques) have not been applied to AI for modern computer games, as they are considered to be unsuitable. Instead, AI for modern computer games is mostly based on heuristics and probabilistic reasoning.

However, modern board games, which contain elements from both classic board games and modern computer games, can function as a stepping stone to discover new applications for the classic techniques, so that they may find applications in modern computer games as well.

To investigate AI for modern board games, some typical games must be chosen as research subjects. The five requirements we identified for these game selections are:
1. The social aspect should be crucial in the gameplay, even if limited to tactical aspects such as temporary alliances.
2. The game should be sufficiently popular, and should have a relatively simple ruleset, so that it can be easily explained.
3. The game should support 3 or more players, as a major objection to the use of tree-search techniques is that they generally do not support more than two players.

4. The game should contain some imperfect information, so that no "perfect move" can be identified.
5. Non-determinism should have little influence on the outcome of the game, to limit the number of trials that must be run to get statistically significant results.

We have not yet identified a game that meets all these requirements. We have, however, organized a one-day workgroup in which multiple programmers developed an AI for the game "The Resistance". These AIs entered into a competition, in which various teams played the game over at least 10,000 rounds.

"The Resistance" is a game in which players get assigned a role: each player is either a Spy or a Resistance fighter. The roles are secret but the Spies get to know each other at the start of the game. One player is the leader, who has the responsibility to propose a team to go on a mission. The size of the team is determined by the rules. Everybody gets to vote on the team, and if the majority agrees with the leader's selection, the team gets to execute the mission. Regardless of the outcome of the vote, the leader role moves to the next player. Moreover, if the proposed team for a mission is rejected five times, the Spies automatically win the game. During the mission, the members of the team each play a card that indicates whether they support or sabotage the mission. These cards are played in secret, shuffled, and revealed. If one of the cards says that the mission is sabotaged, the mission fails. Otherwise it succeeds. As soon as three missions have succeeded, the Resistance fighters win. If, however, before that time the Spies manage to sabotage three missions, the Spies win.

During the programming workgroup, diverse ideas for the implementation of AIs were investigated. While it is difficult to draw solid conclusions from the experiments, several things were noted:
1. It seems that the Spies have a much better chance at winning than the Resistance fighters. The best Spy AIs won the game almost 50% of the time, while the best Resistance fighter AIs won the game less than 20% of the time.
2. The proportion of victories of Spies compared to Resistance Fighters highly depends on the types of AIs included in the competition. Especially the inclusion of a randomly playing strategy had a high impact on the reported results.
3. Simple strategies tended to work better than more complex strategies. This may no longer hold when new strategies are built.

We now intend to expand the game engine with extra communication possibilities, and organize the competition on a larger scale.

3.9 Player Modeling

Pieter Spronck, Georgios Yannakakis, Christian Bauckhage, Elisabeth André, Daniele Loiacono, and Günter Rudolph

Player modeling concerns the capturing of characteristic features of a game player in a model. Such features may encompass player actions, behaviors, preferences, goals, style, personality, attitudes, and motivations. Player models can be used to let the game adapt automatically to be better able to achieve its goals with respect to the player. Games may adapt to a player without constructing a player model by simply responding to changes to the game world or to biometric data of the player; however, creating a player model as an intermediate step has at least two advantages: (1) it creates an understanding of who the player is, and therefore an argument for making specific adaptations; and (2) a player model allows generalization of adaptations to other games.

Player models can be constructed upon three data domains: (1) gameplay data, which are directly extracted from the game and player interactions with the game world; (2) subjective data, which are collected by means of a questionnaire (e.g. psychometrics, demographics, emotional states, personality tests); and (3) objective data which can be extracted from e.g. biometrical observations.

Two key approaches to developing player models can be distinguished. The first, theory-driven, approach is derived from experimental psychology and the social sciences, which consists of proposing a model based on literature and domain experience, and then validating that model empirically. The second, data-driven, approach is derived from computer science and the natural sciences, which consists of collecting a large dataset of measurements, and then using computational methods to automatically (or semi-automatically) derive models. Comparing the two approaches, we note that the first inherently contains argumentation and understanding for the choices of the model, which the second lacks. However, practice shows that such models often fail to encompass relevant features because of a lack of insight of the model builders. The second approach has the advantage of automatically detecting relevant features; however, it is also prone to detecting meaningless relationships between user attributes, game context and user experience.

In computer games an extensive set of features of player behavior can be extracted and measured. At the same time there is, usually, lack of insight in what these features actually mean, at least at present. Therefore, in the current state of research, the second approach seems most suitable. A typical technique employed for creating player models using the second approach is data clustering. This technique cannot be applied well without sufficient insight into the structure of the data space, however; depending on this structure, different clustering algorithms may give widely different results. Domain-specific knowledge, feature extraction and feature selection are necessary to achieve results that make sense.

Experience has shown that diligent application of clustering techniques may provide insight into group behaviors. However, it remains hard to make predictions about individuals. As such, player models can usually only give broad and fuzzy indications on how a game should adapt to cater to a specific player. One possible solution is to define several possible player models and classify an individual player as one of them. Then, when gameplay is going on, the model can be changed in small steps to fit the player better. I.e., the player model is not determined as a static representation of the player, used to determine how the game should be adapted; rather it is a dynamic representation of a group of players, that changes to highlight the general characteristics of a specific player, and drives the game adaptation dynamically.

Note that player characteristics within a game environment may very well differ from the characteristics of the player when dealing with reality. Thus, validated personality models such as psychology's Five Factor Model might not fit well to game behavior. An interesting research direction in player modeling research is to determine a fundamental personality model for game behavior; such a model will have some correspondence with the Five Factor Model, but will also encompass different characteristics. Moreover, the behavioral clues that can be found in game behavior will be considerable different from those that can be found in reality.

Regardless of the line of research in player modeling chosen, the biggest obstacle right now is a lack of data. What is required is a rich multimodal corpus of gameplay and player data as well as player descriptions. Such a corpus must include detailed gameplay data for several games for a large number of players, including actions, events, locations, timestamps as well as biometrical data that are trivial to obtain in large scales (e.g. camera images and eye-tracking). Demographic data for the players must be available, as well as player information in the form of several questionnaires and structured interview data. Not all this data needs to be available for every subject in the database; several large datasets of gameplay data already exist, and it would be beneficial to include those in the database too.

3.10 Evaluating AI in Games Research

Kenneth O. Stanley, Michal Bida, Paolo Burelli, Risto Miikkulainen, and Georgios N. Yannakakis

An important task for any researcher in artificial intelligence (AI) is to formulate the right experimental setup to evaluate research results. However, this task is particularly challenging when AI is combined with video games because the goal of the research is often unstated and the results may be in part subjective. Nevertheless, the general issue of the proper course for evaluation in this area is rarely discussed, which has sometimes led to confusion when authors and reviewers hold differing assumptions on how evaluation should be conducted. The goal of our group session was to begin to address this issue by surveying available types of evaluation and offering some recommendations for future authors.

We identified two key categories of goals for research in AI and games. The first is "games as AI arenas." In this type of research, the game acts mainly as a benchmark for testing the performance of AI algorithms. The second category is "AI for better games," wherein the goal is to improve the game experience itself through innovative AI technology. An important conclusion of our session was that it is critical that authors should identify into which category their research falls because the proper form of evaluation will depend necessarily upon this distinction.

Participants also explored several other dimensions of research evaluation in this area, all of which ultimately tie into the question of the primary goal of the research. Among the other issues discussed were different types of evaluation (i.e. researchers studies versus evaluation by a reviewer), the expectations of different audiences (i.e. academia, industry, or players), methods of evaluation (e.g. objective versus subjective and quantitative versus qualitative), and evaluation metrics (e.g. human subjects studies, competitions, expert evaluations, benchmarks, statistical analysis, etc.). A general challenge for many studies is to obtain sufficient player data to draw conclusions.

Our main conclusion is a recommendation for authors in this area, particularly those submitting to the IEEE Transactions on Computational Intelligence and AI in Games (TCIAIG) journal: It would help in the evaluation of research by reviewers if authors would specify (1) their research goal (one of the two primary goals) and (2) how their evaluation matches their goal. Our recommendation is to ask authors to make clear their answers to both these questions in any paper that they submit in this area.

3.11 The AIGameResearch.org AI Game Clearinghouse

Kenneth O. Stanley, Alex J. Champandard, Clare B. Congdon, Philip F. Hingston, Graham Kendall, Pier Luca Lanzi, Daniele Loiacono, and Risto Miikkulainen

Many researchers in the AI-in-Games community build their own video games to demonstrate or test new technologies. These games often exhibit innovative features that would appeal to other researchers, the game industry, and even the general public. It is also often the case that the more players an experimental game attracts, the more informative the experimental analysis can be. Yet at present there is no simple or common solution for attracting a sufficient audience to an experimental game. The aim of our session was to begin to organize an "AI-Game Clearinghouse" that will allow the community to pool its collective products to attract a significant audience to the work of the community in general. The result will be a new website, AIGameResearch.org, where anyone on the internet can find a wide range of innovative games based on cutting-edge AI technologies.

Creating such a website can yield several other benefits for the research community. Not only can it attract more players to our games and thereby enhance our research and publicity, but it can also help to demonstrate the significance of our work to the game industry and to funding agencies concerned with relevance and impact. Furthermore, in addition to games, the website can provide a pathway to active experiments, source code, announcements, publications relating to the game collection, and discussion forums for users. Thus it can serve as a clearinghouse for research in this area. Another interesting facet of such a site is that because the public is genuinely interested in playing innovative games, it is possible that the site could attract revenue, which might someday help to support research in AI and games. Research groups might also someday sell their games through the site.

A site offering games to the general public will require moderation to ensure an acceptable level of quality. Otherwise, the public will not ultimately trust the site, jeopardizing its mission. It is likely that many individuals and groups (including non-researchers) will eventually want to submit their games to the site once it becomes widely known. Therefore, we propose to establish an "editor in chief" who makes final decisions on whether to include individual games that are submitted based on the recommendation of reviewers from a permanent "editorial board." Another aim to maximize the site's impact is to establish ties to IEEE TCIAIG, AIGameDev.com, GPEM, and other relevant venues and resources. However, the site will remain a community effort independent of any specific organization.

Participants resolved to continue this effort by establishing the site and editorial board after the conclusion of the Dagstuhl workgroup.

3.12 Video Game Description Language

Tommy Thompson, Simon Lucas, Tom Schaul, John Levine, Marc Ebner, and Julian Togelius

As participants in this Dagstuhl session address the challenge of General Video Game Playing (GVGP), we have recognised the need to create a Video Game Description Language (VGDL).

Unlike General Game Playing, we have envisioned GVGP will not require a prescribed language to facilitate understanding of the logic of the game: requiring the computational agents to ascertain these facts for itself. However, we would still require means to define the wide range of problems the GVGP agents may face for the purposes of classification and categorization. Not only would such a language provide means to encapsulate the features and mechanics of a game for the purposes of human understanding, but also provide context for the evaluation of GVGP agents having completed playing.

Outside of the issues of classification, there is also the opportunity for automatic game generation. Given the intent of the GVGP group to work within the Physical Traveling Salesman Problem (PTSP) framework, we aim to attach a code-base to the VGDL compiler that derives implementations of these games from the definition that can be used in conjunction with GVGP. Implementing such a compiler could provide numerous opportunities; users could modify existing games very quickly, or have a library of existing implementations defined within the language (e.g. an Asteroids ship or a Mario avatar) that have pre-existing, parameterized behaviors that can be customized for the users specific purposes. Provided the language is fit for purpose, automatic game creation could be explored further through experimentation with machine learning algorithms, furthering research in game creation and design.

In order for both of these perceived functions to be realized and to ensure it is suitable for a large user base we recognize that the language carries several key requirements. Not only must it be human-readable, but retain the capability to be both expressive and extensible whilst equally simple as it is general. In our preliminary discussions, we sought to define the key requirements and challenges in constructing a new VGDL that will become part of the GVGP process. From this we have proposed an initial design to the semantics of the language and the components required to define a given game. Furthermore, we applied this approach to represent classic games such as Space Invaders, Lunar Lander and Frogger in an attempt to identify potential problems that may come to light.

In summary, our group has agreed on a series of preliminary language components and are now keen to experiment with forms of implementation for both the language and the attached framework. In future we aim to realize the potential of the VGDL for the purposes of Procedural Content Generation, Automatic Game Design and Transfer Learning and how the roadmap for GVGP can provide opportunities for these areas.

3.13 Procedural Content Generation

Julian Togelius, Alex J. Champandard, Pier Luca Lanzi, Michael Mateas, Ana Paiva, Mike Preuss and Kenneth O. Stanley

The work group on procedural content generation (PCG) focused on the future potential of this young subfield of game AI research. Three ambitious goals for PCG research were stated, and eight medium-term research challenges were identified. The solution to these challenges would constitute good progress towards ultimately solving the grand research goals. Further, six actionable steps were identified; these are concrete research projects which one could start work on immediately and which would contribute to solving some of the research challenges. The three goals stated were:

1. high-quality multi-level, multi-content PCG
2. PCG-based game design and
3. generating complete games from scratch.

The eight research challenges were:
1. non-generic, original content
2. representing designer style
3. general content generators
4. search space construction
5. usable and powerful interfaces for PCG systems
6. overcoming the animation bottleneck
7. interaction and opportunistic control flow and
8. establishing a comprehensive theory and taxonomy of PCG systems.

The six actionable steps were:
1. generating complete Atari 2600 games
2. procedural animation for simple generated creatures (e.g. a fish)
3. co-generating quality quests and maps
4. generating music modulated by game events
5. competent generation of Super Mario Bros levels including macro-scale structure and progression and
6. player-directed generation with model-based selection.

For each of the goals, challenges and actionable steps, the work group identified the state of the art in terms of published papers and games.

3.14 Computational Narrative

R. Michael Young, Ruth S. Aylett, Paolo Burelli, Mirjam P. Eladhari, Richard Evans, and Ana Paiva

The use of automatic generation methods to create narrative elements in games has the potential to create genres of game play that have only been hinted at to date in commercial titles. In pursuit of this capability, a growing number of research efforts are targeting critical representational and algorithmic problems in the area of computational narrative. Current challenges for scientists working in this area range across many problem areas. Some of the challenges we consider both significant and near at hand include:
- the creation of dynamic characters (inducing their dialog and internal personalities)
- the generation of novel quests and story sequences
- the adaptation of a game's story in response to player activity and
- the development of sharable systems for realization. Specifically, the construction of research systems in computational narrative would be accelerated by access to game environments/toolkits that can be used to address multiple research problems, that can be accessed by distinct research sub-systems and that and can be shared across many research groups.

Near-term research opportunities for this community include:
- systems that build stories drawn from MMO game play logs,

- systems that generate tailored story-based support for face-to-face role playing used in corporate training and simulation and
- collaborative support for machinima generation used as pre-visualization for game developers, cinematographers and video/film producers.

We also see the internet as a significant resource to address a number of research problems, including aspects of narrative content creation, crowd-sourced data collection for evaluation and other aspects.

- Longer-term research goals for this community include:
- Systems that, like Star Trek's holodeck,create entire story worlds and dynamically drive them through a user's interactive experience,
- games that blend alternate reality game mechanics, augmented reality capabilities and automatically generated narrative elements,
- Long-lived, drop-in/drop-out narrative-based games that last for months or years and
- Systems that automatically drive hybrid human-robot systems, where human players interact with robots as NPCs

4 Overview of Short Talks

During the seminar about 15 short talks were given. The contents of most of these are incorporated in the workgroup reports (Section 3). The contents of the remaining ones are given below.

4.1 Using Computer Games to Close the Loop in Artificial Visual Information Processing

Marc Ebner (Universität Greifswald, DE)

The human brain is highly complex. Even though significant advancements have been made in recent years, it is still unknown how the brain works as a whole. We are currently unable to simulate an artificial brain which is able to mimic human performance in its entirety. This is in part due to a lack of understanding how various parts interact. In order to create fully autonomous individuals which can interact with human players in computer games or virtual realities we need to be able to simulate a human individual in a way such that it is not apparent to the player whether this individual is artificial or not. Computer games, viewed through screen captures, are an ideal tool to perform research on closing the loop from visual input to the action of a virtual player.

4.2 Social Simulation Games

Mirjam P. Eladhari, Richard Evans, and Michael Mateas

The purpose of this talk was to introduce the topic of social simulation games for discussion at the Dagstuhl Seminar 12191, "Artificial and Computational Intelligence in Games". We wanted to identify why we would want to simulate social interactions in games, and to open the seminar-discussion for identifying the main research challenges in the field.

Well-known examples of social simulation games include Little Computer People[2], The Sims[3], PsychSim [7], Façade [5], and recently, Prom Week [6]. The type of play in social simulation games can be compared to children's play with dollhouses, as with chamber plays and improvisational theater. While the former help children's learning about practices in social behavior, the latter may help us, in later stages of life, to better understand the human condition. This is also the case for well-crafted social simulation games.

Generally, the term 'social' refers to the interaction of organisms with other organisms and to their collective co-existence. In terms of simulating these interactions we always, when building social simulations, use different models of mind for the system design. In the area of social simulation games, these can be divided into three types. One type of model can be an underlying theory of social behaviors, such as Wittgenstein's notion of social practices [3, 8, 9] that is used in The Sims 3[4], or the use of Goffman's [4] theories in the design of Prom Week. A second type of model for design is that of game genre. Prom Week have similarities with puzzle games, and The Sims with dollhouse play. In the research prototype Pataphysic Institute [2] the conventions of combat in the genre massively multiplayer online role playing games are used as a model for interaction in certain play modes. Yet a third type of model of mind affecting the design of social simulations may derive from the use of an existing or for the purpose invented implementation technique. For example, much of the game design in the Pataphysic Institute revolves around the fact that spreading activation networks are used in the implementation of both autonomous and semi-autonomous entities, while Prom Week and The Sims 3 both use forward chaining rules. Social simulation games are created for various reasons. It can be for their own sake, taking a l'art pour l'art perspective as in the artistic tradition, or for exploring the human condition as in the humanist tradition. It can be for pure entertainment, and the selling of the same, or it can be for a practical purpose such as for teaching about conflict resolution or cross-cultural understanding. No matter the purposes of individual projects, the creation of social simulation games is likely to lead to new advances within the areas of social believable agents as well as in complex systems.

Challenges for future work, identified at the seminar, include:

- defining believability [1];
- negotiating boundaries of the research areas addressing social simulation and believable agents respectively in order to allow technology transfer;
- a single implementation approach that work in both single and multiplayer, for both player- and non player characters;
- further development for use of social practices by:
  - defining an operational language for specifying social norms and practices;

---

[2] David Crane and Rich Gold, Activision, 1985
[3] Will Wright, Electronice Arts, 2000
[4] Electronic Arts, 2009

- modelling examples of full breath of social practices;
- creation of agents who learn new social practices by adapting to an inhabited world or environment;
- solve how to allow agents to participate in simultaneous, concurrent social and hierarchical practices;

- addressing authoring by:
  - creating authoring tools for balancing casts of characters;
  - creating in-character action sequences, or drama management, producing coherent story units;
  - integrating multiple capabilities for coherent agent-performance;
- increasing expression and believability of agents/characters by:
  - facilitation of expression of character by actions and adverbial modifiers;
  - widening "the expression bottle-neck", currently production of audiovisual content is costly even if procedural;
  - increased usage of non naturalistic forms of expression;
  - create believable agents that evolve and change, i.e. with increased persistence, and;
  - create characters who behave as if they have a history.

References

1 J. Bates, The Role of Emotions in Believable Agents, Technical Report CMU-CS-94-136, Carnegie Mellon University, 1994.
2 M. Eladhari, "The Pataphysic Institute", in Proceedings of the Sixth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-10), The AAAI Press, Menlo Park, CA, 2010.
3 H. Garfinkel, Studies in Ethnomethodology (Social & political theory), Polity, 1991.
4 E. Goffman, The Presentation of Self in Everyday Life, Anchor, 1959.
5 M. Mateas and A. Stern, "Procedural Authorship: A Case-Study of the Interactive Drama Façade", in Digital Arts and Culture: Digital Experience: Design, Aesthetics, Practice (DAC 2005), 2005.
6 J. McCoy, M. Treanor, B. Samuel, B. Tearse, M. Mateas, and N. Wardrip-Fruin, "Authoring Game-based Interactive Narrative using Social Games and Comme il Faut", in Proceedings of the 4th International Conference & Festival of the Electronic Literature Organization: Archive & Innovate (ELO 2010), 2010.
7 D. Pynadath, V. David, and S. Marsella, "PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents", in Proceedings of the IJCAI05 Conference, 2005, pp. 1181–1186.
8 L. WittgenStein, Philosophische Untersuchungen / Philosophical Investigations, Basil Blackwell, New York, 1953.
9 L. Wittgenstein, "Preliminary studies for the Philosophical Investigations", in The Blue and Brown Books, Harper & Row, New York, 1958, pp. 1961–1961.

4.3 Believable Agents for Games

Philip F. Hingston (Edith Cowan University, AU)

Humans are interacting more and more with robots, bots, and other agents. Our thesis is that if these agents are more believable/human-like, our interactions are likely to be more successful. For example, in the computer game context, human-like bots or NPC's (non-player characters) are often more fun to play against (see, for example, [1]).

We ask the reader to imagine themselves playing a game against a robot (e.g. tennis) or an NPC (e.g. a fast-paced action game). Suppose you know (or suspect) that your opponent is not human — would this make the experience more or less fun, or would it make no difference to you? Aside from the initial novelty value, there is quite a bit of evidence to suggest that non-humanlike bots may to be less fun to play because: they are too hard to beat (too fast, too accurate etc); or too easy to beat (too stupid, non-adaptive etc); or perhaps because there is no shadenfreud – if you know your opponent is a bot, then there is no joy in causing it pain (for example, Weibel et al. showed that players prefer to play opponents that they believe to be human, even if they are not in fact human [2]).

Suppose then we that we want to make more believable/human-like bots, or more generally, agents. How can we go about it? There are at least three different kinds of approaches: ad-hoc (these usually use some kind of hand-crafted rules, perhaps with randomness added); learning-based (these use various methods to learn competence, or to learn by imitation to be human-like, or to adapt to the opponent/environment); cognitive/psycho-social models (these are the most ambitious, attempting to model human behavior). The last of these is the most recent innovation, and it will be interesting to see how successful it will be going forward.

Whatever means we use to create them, how can we tell if the agents we create are believable? One answer is to design a suitable Turing test. For example, I organize an annual NPC Turing test competition based around the commercial game Unreal Tournament 2004 (a FPS or first-person shooter). In this competition, competitors create and enter AI-based NPC's, and human judges try to decide which of their opponents is a human and which is a bot. To date, the competition has been run in 2008, 2009, 2010 and 2011, and although the NPC's are improving, judges still reliably rate human opponents as more human than NPC opponents (see [3] for a description and analysis of the results up to 2009). Competitors have used all of the approaches listed above, and combinations of them. At present, there is theorizing but no clear understanding of how the judges are able to make this distinction, even in the very limited context of a FPS, where the available actions mainly consist of frenetically running and jumping, shooting at the opponent with various deadly weapons (only in the virtual game world, of course).

The purpose of this talk was to engage the listener to start to consider believability, or the related concept of human-like-ness in intelligent agents. Some seminar participants were keen to disagree with the views put forward in this talk, and that is a good starting point for interesting discussion, argument and questioning!

To conclude, I'd like to put a couple of more philosophical questions:

- should agents merely appear human-like, or should they BE human- like?
- would it be better in some cases to make them unmistakably NOT human?

References

1. B. Soni and P. Hingston, "Bots Trained to Play Like a Human are More Fun", in Proc. of IEEE International Joint Conference on Neural Networks, IJCNN 2008, Hong Kong, June 2008.
2. D. Weibel, Bartholomäus. Wissmath, S. Habegger, Y. Steiner and R. Groner, "Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment", Computers in Human Behavior, 24(5), pp. 2274–2291, 2008.
3. P. Hingston, "A Turing Test for Computer Game Bots", IEEE Transactions on Computational Intelligence and AI In Games, Vol. 1, No. 3, September 2009.

4.4 Adaptive Artificial Intelligence in Games

Pieter Spronck (Tilburg University, NL)

Joint work of Spronck, Pieter; Ponsen, Marc; Sprinkhuizen-Kuyper, Ida; Postma, Eric
Main reference P. Spronck, M. Ponsen, I. Sprinkhuizen-Kuyper, E. Postma. "Adaptive game AI with dynamic scripting," Mach. Learn. 63, 3 (June 2006), 217–248, 2006.
URL http://dx.doi.org/10.1007/s10994-006-6205-6

Adaptive Game AI concerns artificial intelligence (AI) in computer games which adapts to dynamic circumstances. In particular, the AI adapts in response to behavior of human players. The three main goals of adaptive game AI are (1) self-correction, the ability to recover from mistakes; (2) creativity, the ability to invent new tactics; and (3) scalability, the ability to exhibit behaviors appropriate for the human player's skill level.

Commercial game developers have included adaptive game AI in only a very small number of games. Some of the reasons for this lack of interest are the high complexity of modern games, the perceived lack of efficiency of adaptive techniques, and fear of AI learning the wrong lessons.

For game developers to accept adaptive techniques in their games, it is essential that these techniques meet four computational and four functional requirements. The computational requirements are: (1) Speed: the AI must be fast as there is little processing power available for adaptation; (2) Effectiveness: the AI cannot tolerate (much) ineffective behavior, even during learning; (3) Robustness: the AI must be able to deal with the inherent non-determinism that exists in most modern games; and (4) Efficiency: the AI must use as many learning opportunities as possible, to finish learning long before the game is over. The functional requirements are: (1) Clarity: game developers wish to understand what adaptive game AI is actually doing; (2) Variety: the AI should not always exhibit the same behavior; (3) Consistency: the AI should finish learning in a predictable period of time; and (4) Scalability: the AI should take the human player's skills into account.

While traditional adaptation techniques seldom meet all these requirements (e.g., most of them are breaking either the effectiveness or the efficiency requirement), several techniques for adaptive game AI exist which are suitable for commercial modern games – in particular, techniques based on optimization (e.g., hill-climbing), imitation (e.g., case-based reasoning), and reinforcement (e.g., dynamic scripting).

With the ever increasing complexity and realism of virtual game world, the player's freedom to express behavior in games increases as well. The consequence is that the AI has to take into account and interpret an increasing variety of player behaviors. This means that AI that worked well last year, is no longer sufficient for newly released games. We can actually observe a decline in effectiveness of game AI that is developed with classic methods. Adding adaptation to game AI will allow it to become more effective automatically. Therefore, it is not a question if, but when game developers will give their AIs adaptation capabilities by default.

Participants

- Elisabeth André
Universität Augsburg, DE
- Ruth Aylett
Heriot-Watt University Edinburgh, GB
- Christian Bauckhage
Fraunhofer IAIS – St. Augustin, DE
- Michal Bida
Charles University – Prague, CZ
- Adi Botea
IBM Ireland – Dublin, IE
- Bruno Bouzy
Paris Descartes Unversité, FR
- Paolo Burelli
IT Univ. of Copenhagen, DK
- Michael Buro
University of Alberta, CA
- Martin V. Butz
Universität Tübingen, DE
- Alex J. Champandard
AiGameDev.com KG, AT
- Clare Bates Congdon
University of Southern Maine, US
- Peter Cowling
University of Bradford, GB
- Marc Ebner
Universität Greifswald, DE
- Mirjam P. Eladhari
University of Malta, MT
- Richard Evans
Little Text People at Linden Lab, GB
- Philip F. Hingston
Edith Cowan University, AU
- Graham Kendall
University of Nottingham, GB
- Pier Luca Lanzi
Politecnico di Milano, IT
- John M. Levine
University of Strathclyde, GB
- Daniele Loiacono
Politecnico di Milano, IT
- Simon M. Lucas
University of Essex, GB
- Michael Mateas
University of California – Santa Cruz, US
- Risto Miikkulainen
University of Texas – Austin, US
- Hector Munoz-Avila
Lehigh Univ. – Bethlehem, US
- Dana S. Nau
University of Maryland – College Park, US
- Ana Paiva
IST – TU of Lisbon & INESC-ID
- Mike Preuss
TU Dortmund, DE
- Günter Rudolph
TU Dortmund, DE
- Tom Schaul
New York University, US
- Moshe Sipper
Ben Gurion University – Beer Sheva, IL
- Pieter Spronck
Tilburg University, NL
- Kenneth O. Stanley
University of Central Florida – Orlando, US
- Tommy Thompson
The University of Derby, GB
- Julian Togelius
IT Univ. of Copenhagen, DK
- Georgios N. Yannakakis
IT Univ. of Copenhagen, DK
- R. Michael Young
North Carolina State Univ., US

Report from Dagstuhl Perspectives Workshop 12212

Co-Design of Systems and Applications for Exascale

Edited by
Arndt Bode[1], Adolfy Hoisie[2], Dieter Kranzlmüller[3], and Wolfgang E. Nagel[4]

1 **Technische Universität München & Leibniz Supercomputing Centre Garching, DE,** bode@informatik.tu-muenchen.de
2 **Pacific Northwest National Lab., US,** adolfy.hoisie@pnnl.gov
3 **Ludwig-Maximilians-Universität München & Leibniz Supercomputing Centre Garching, DE,** kranzlmueller@nm.ifi.lmu.de
4 **Technische Universität Dresden, DE,** wolfgang.nagel@tu-dresden.de

Abstract

With more and more machines achieving petascale capabilities, the focus is shifting towards the next big barrier, exascale computing and its possibilities and challenges. There is a common agreement that using machines on this level will definitively require co-design of systems and applications, and corresponding actions on different levels of software, hardware, and the infrastructure. Defining the vision of exascale computing for the community as providing capabilities on levels of performance at extreme scales, and identifying the role and mission of the involved experts from computer science has laid the basis for further discussions. By reflecting on the current state of petascale machines and technologies and identifying known bottlenecks and pitfalls looming ahead, this workshop derived the concrete barriers on the road towards exascale and presented some ideas on how to overcome them, as well as raising open issues to be addressed in future leading-edge research on this topic.

Seminar 22.–25. May, 2012 – www.dagstuhl.de/12212
1998 ACM Subject Classification B. Hardware, D.2 Software Engineering
Keywords and phrases Exascale, Co-Design, Scalability, Power Efficiency, Reliability
Digital Object Identifier 10.4230/DagRep.2.5.71
Edited in cooperation with Christian Straube

1 Executive Summary

Arndt Bode
Adolfy Hoisie
Dieter Kranzlmüller
Wolfgang E. Nagel

With Petascale computing being a reality today, the focus of the computational science community is already on the next barrier – exascale computing. With systems even more powerful by orders of magnitude, scientists start thinking about the possibilities and challenges. This workshop addressed the many scientific, technological, and financial challenges of exascale level computing with the hypothesis that exascale computing is only possible by co-designing across different levels of software, hardware, and the surrounding infrastructure.

The workshop program has been composed of a series of short talks, less than 20 minutes on average, and extensive time for discussions. Starting with an overview of the workshop motivation and the general methodologies for co-design, different aspects of co-design have been addressed. This has been followed by talks on modeling, simulation and tools, as well as programming models, runtime support and compilers. The second part addressed the specific problems of system-software for performance, power and reliability and the resulting system architectures, while finally application level aspects of exascale co-design have been discussed between the participating experts from different areas of high performance computing. In all discussions it has been important to tackle a multidimensional combination of major challenges associated with the development of exascale systems and applications from different angles instead of addressing an isolated aspect.

The results of the workshop are manifold[1]: Firstly, the vision based on the requirements of the scientific community is thus "to provide exascale capabilities to scientific and engineering applications", where it is important to notice that exascale means extreme scale or large scale, not the particular barrier of exaflop performance looming ahead. With this vision at hand, the participating experts identified their particular role and mission as follows: "to co-design systems such that they reach exascale capabilities within the given technological and non-technical (social, ...) boundaries". Each expert has been knowledgeable on a distinct layer of the exascale architecture, the mission requires expertise across all layers, and exascale computing requires involvement from all relevant areas of computer science in order to perform exascale co-design of hard- and software, including also different levels of software working closely together with hardware and the interfacing to the environmental infrastructure. This has lead to the definition of co-design, where two or more distinct activities collaborate on and across different layers to design a system architecture for a specific goal.

In summary, the workshop has reflected on the current state of petascale machines providing multiple examples from world-leading machines and using them to derive the barriers on the road towards exascale computing. Looking beyond the current research into the future, where exascale computing will become feasible, we have been trying to identify the exascale roadmap with intermediate goals and pitfalls on the way to exascale, and leveraging the combined forces of computer science to overcome them.

[1] A scientific paper will be created within the next months.

2 Table of Contents

3 Exascale Co-Design Methodologies (Overview)

3.1 Exascale Co-Design Workshop: Overview and Motivation

Adolfy Hoisie (Pacific Northwest National Lab., US)

- Modeling of systems and applications
- Simulation methodologies and frameworks
- Co-Design boundaries: architecture, application, system software, programming models, tools
- Co-Design process

3.2 Co-Design: a Holistic Approach on Energy-Efficient Computing

Arndt Bode (TU München & LRZ Garching, DE)

Co-Design is usually considered to be applied to architecture and applications in order to keep exascale systems programmable and scalable. Based on the experience with SuperMUC and other energy efficient HPC systems we advocate to extend the co-design principle to the entire "HPC production chain": building infrastructure – power supply – climate and cooling – systems hardware architecture – systems software architecture including operating system, programming language, libraries and tools middleware – operating strategies – application and algorithm.

Some arguments for the extension of the co-design paradigm are

- all of the elements of the HPC production chain listed above influence strongly the amount of energy consumed by the solution
- future exascale system cost and technical feasibility are strongly dependent on energy consumption
- many of the elements of the HPC production chain are strongly correlated (example: to what extent does the operating strategy allow the application program to control the clocking of the system)

A discussion on this holistic approach to co-design should bring together all parameters influencing energy consumption and their mutual dependency.

3.3 Another Perspective to Co-Design . . . from one of those who did not raise their hand

Hans Bungartz (TU München, DE)

- Meaning of co-design: not only the "classical" HW-SW one, but also (and maybe even more important) SW-SW co-design, having in mind the complete SW stack for system SW to application software

- How far can/should/must the co-design of systems and applications go? i.e. application-tailored (narrow) systems (MD-Grape, QPACE, ...) vs. application-type systems (BlueGene, e.g.) vs. general-purpose systems
- Maybe a more philosophical issue: From the general application perspective, "exascale" pushes HPC into a narrower niche. Only a few communities have the experience and the codes for petascale, and the number will decrease towards exascale. And only a few communities really need exascale computing, and also this number will decrease towards exascale. Nevertheless, the technologies on the exascale agenda will bring benefit to the non-high-end computing sector, too. We should keep this in mind – in particular with regard to engineering/industrial applications
- Should we make a "funding roundtable" – maybe in the evening? Discussing problems, presenting briefly new initiatives world-wide, reporting the IESP/EESI status, ...?
- algorithmic issues: impact of the architecture on algorithm design, esp. having in mind the big algorithmic paradigms such as hierarchy, adaptivity, dynamics, multi-level – which are all somewhat nasty for exascale ...

3.4 Performance Engineering based on Knowledge instead of Abstraction

Jan Treibig (Universität Erlangen-Nürnberg, DE)

Great improvements of the efficiency of applications are possible using a holistic iterative performance engineering approach guided by a deep understanding of the interaction between software and the hardware.

Exascale requires an extreme effort with regard to the efficient use of available hardware. We propose a holistic iterative optimization process, which takes into account code, machine and runtime analysis in order to iteratively create a diagnostic performance model. This is a white box approach in every aspect, leading to a deeper knowledge of the characteristics of the application code and the properties and capabilities of the underlying hardware. The application developer himself is conducting this effort forcing him to get a deeper understanding of the interaction of his code with the hardware. This is in opposite to the widespread belief that the developer should be safeguarded from these details by software abstraction and tooling through a black box approach. It is our attitude that there is no alternative to deep knowledge about what is happening on all levels. Making this knowledge available in a transparent fashion to all participating disciplines is the key for a sustainable optimization effort. All necessary aspects are already available including static code analysis tools, micro benchmarking, profiling tools and hardware performance monitoring. This analysis allows to create an application- and hardware-specific diagnostic performance model, which explains the resulting performance. Since such a model is based on understanding and not on heuristics or statistics it can identify opportunities for possible optimization strategies and provides reliable statements about the attained efficiency of a code on a given architecture. By proposing and teaching this systematic method of performance engineering a great potential for increased efficiency will be leveraged.

4 Exascale Co-Design Methodologies (Modeling, Simulation, Tools)

4.1 Tools and Tool Infrastructures for Co-Design

Martin Schulz (LLNL – Livermore, US)

Program development and analysis tools are an essential element of the co-design process. They enable the evaluation of applications and their characteristics and they can help predict the performance impact on future architectures. However, this requires the development of a new generation of tools that allows a more detailed and predictive performance analysis as well as an expansion of the concept of tools to include simulation and emulation environments of future systems to help understand the impact of architectural changes on applications and system software.

The GREMLIN project, which is in its early design stage as part of DOE/Office of Science Co-Design Center ExMatEx, aims at creating such an emulation infrastructure. It consists of a series of small modules, the GREMLINs, that target particular aspects of a system design and either limit targeted resources, such as caches, network bandwidth, or arithmetic units, or introduce faults or process variations. Combined, these GREMLINs will allow us to create exascale conditions on current machines and with that evaluate applications in larger and more realistic settings. The GREMLIN is intended to close the gap between cycle accurate simulation and execution on prototype hardware and is part of ExMatEx's multi-pronged approach to performance and scalability analysis.

The GREMLIN will itself be built on top of a flexible tool infrastructure that allows us to easily extend and combine the various modules. Further, it enables the transparent integration of performance analysis tools on top of the GREMLIN and provides us comparative performance analysis of different system configurations. Such modular infrastructures are critical to support users at exascale and in particular during the co-design process. They enable us to quickly assemble tools from components, provide mechanisms for interoperability and concurrent execution, and allow us to target specific performance questions without having to rewrite each tool from scratch. This talk will highlight recent efforts in this area, which also reflect a larger trend in the tools community, and how they can help to create new tools, including the GREMLIN tool set.

4.2 An Analytical Framework for Algorithm-Architecture Co-Design

Kent Czechowski (Georgia Institute of Technology, US)

Traditional co-design analysis abstracts away algorithm features entirely (e.g., through an Amdahl's Law style analysis) or draws conclusions from specific code artifacts (e.g., fixed benchmarks, traces, and machines). Instead, I advocate for a more general algorithm-architecture analysis that explicitly relates characteristics of an algorithm, such as its inherent parallelism or memory behavior, with parameters of an architecture, such as the number of cores, structure of the memory hierarchy, or network topology.

My approach marries abstract algorithmic complexity analysis with a formal representation of architecture design trade-offs. I believe this will enable us to say precisely and analytically how high-level changes to the architecture might affect the execution time of a computation; and, conversely, identify what classes of computation might best match a given architecture. It will necessarily not have the fidelity of cycle-accurate performance estimates possible through detailed simulation. Instead, the strength of the approach is that freedom from the artifacts of current hardware and software implementations, while still obeying technological constraints, may lead to radically new methods and insights into how to achieve performance and scalability in future exascale systems.

4.3 Co-Design – how to? Generality vs. Specialization in Simulation Frameworks

Sabine Roller (German Resarch School for Simulation Sciences, DE)

Simulation software development will at the same time be more individualized and more generalized. More individualized because leading edge software (beyond commercial solvers) needs to specialize. But to be able to specialize, it should not need to start from scratch, but to setup on a generalized basis. This idea stands behind projects like OpenFoam etc. which offer a (community-specific) toolbox in an open-source framework, which can be modified to the individual needs. Nevertheless, this idea runs into its limits at different points. One point is that generalization typically prevents from taking advantages of individual features, especially if the user doesn't overlook the basics, but uses the framework as a black box. Another point, which holds even for highly experienced developers with a deep knowledge, is the dependency on libraries. Currently, simulation software developers often get trapped in scalability issues which are not due to their own software, but due to the underlying libraries like MPI, ParMetis, HDF5. Thus, the key for the future development of HPC software is to resolve the individual view points of application software, middleware, OS software, and operational issues into a holistic end-to-end approach. We need to train more generalists, in addition to the specialists.

4.4 Reaching the Pacific Ocean – Co-Design as a long term activity

Michael Resch (Universität Stuttgart, DE)

Introduction: In 1893 Frederick Jackson Turner wrote an essay on the significance of the frontier in American history. Referring to a bulletin of the Superintendent of the Census for 1890 he found that the impressive move westwards of the US-American settlers has found an end. In his description of the advance of the frontier Turner identifies 5 barriers that were reached over time – the Alleghenies, the Mississippi, the Missouri, the Rocky Mounties, and finally the Pacific Ocean. With the advent of the settlers at the Pacific Ocean the development of the USA turned mostly inwards and focused on the development of the settled country.

In Supercomputing we have seen a similar breath taking advance. Over only a few decades we have reached megaflops, gigaflops, teraflops, petaflops and are approaching exaflops. While many prepare for the usage of such exascale systems , others start to doubt whether we will be able to reach exaflops or go beyond that barrier. From a technical point of view there is no doubt that exaflops are possible. The driving factor, however, is no longer innovation but rather a massive usage of standard procedures and components.

Rationale: Over about 30 years Moore's law was proven to be right and clock frequencies have increased accordingly. Starting in the early 1990s increase of clock frequency was replaced by increase in level of parallelism. An analysis of the TOP500 shows that this trend is accelerating over the last 5-8 years. A consequence of this hardware race is that software cannot keep pace.

In the discussions about the development of exascale systems co-design plays an important role. By developing software and hardware at the same time one expects to overcome the asynchronies of the two technical development paths. Theoretically this is a reasonable approach. However, the asynchronies remain. The basic fact is that changes in hardware come in steps of 2-4 years. This is about the time horizon for the renewal of a system and an upgrade of existing hardware technology. We are still living in the age of rapidly changing technology when it comes to computers.

Software changes happen at a different speed. A short investigation of system software and application software shows that both types of software follow very similar patterns. First, there is an idea or a basic algorithm. In a second step there is some prototype implementation. Over a time of 3-5 years the software starts to mature. A renewal cycle for basic software typically lasts for about 20 years. Looking at the history of operating systems one might assume even longer cycles. If we take UNIX and Linux together as being based on very similar fundamental ideas, one would say that the life span is in the range of 20-40 years. Similar time frames can be found for application software. Basic concepts get implemented in prototype software. Over time these packages mature and become available to a wider user community. Overall the process of maturing a software approach takes at least about twice as much time as the change in hardware architecture and potentially even longer.

It remains to be seen whether a co-design approach can speed up the software development process. It is certainly going to be beneficial if software developers are very early on involved in the hardware development process. On the other hand, for a long time operating system development of large hardware providers was done in-house, such that an internal co-design was already taking place. Nothing indicates that such an in-house co-design was able to overcome the gap between fast hardware development and slow software development.

From this point of view, a slow down if not a complete stop in increase of speed could be helpful for software development. Just like reaching the Pacific Ocean allowed the US society and economy to focus on internal development, reaching the exascale barrier might allow us to focus on the development of mature software and further improve quality rather than speed of hardware.

5 Exascale System Software (Programming Models, Runtimes, Compilers)

5.1 Exascale System Software: Programming Models, Runtimes, Compilers, and Tools

John Mellor-Crummey (Rice University, US)

Exascale architectures will be dramatically different from today's systems in several ways. First, they will feature several orders of magnitude more parallelism. Second, they will come with severe resource limitations including power, memory capacity and bandwidth, bisection bandwidth, and I/O connectivity. Third, to deliver exascale performance in the face of severe resource limitations, exascale systems will employ a complex array of technologies, including heterogeneous cores and deep hierarchical memory systems. Finally, exascale systems will need to cope with faults; the scale of these systems will mean that errors will be several orders of magnitude more frequent than on systems today.

As a result, harnessing the vast computational power of exascale architectures will require fundamental changes to system software, runtime libraries, programming models, tools, and applications. Each layer of the software stack for exascale systems will need to be redesigned to exploit the explosion in parallelism, manage scarce resources, handle failure, and respond to changing application needs. However, we cannot expect developers of applications or other layers of the software stack to understand the nature and scope of changes needed for their software without information about how their existing software falls short of what is needed. Developers must be able to understand where, how, and why their software fails to use a target platform efficiently. The complexity of exascale systems will mean that software may not meet performance expectations for a variety of reasons. At a high level, an application's performance on a large-scale parallel system will depend upon how well it makes use of available resources for communication, computation, and I/O. At a more detailed level, using an exascale platform effectively will require identifying mismatches between software demands for resources and hardware capabilities. Only with detailed knowledge about the nature of shortcomings and a correlation of problems to the software will developers be able to understand how they need to restructure their code to leverage the strengths of a target platform and reduce consumption of scarce resources.

The complexity of hardware and software for exascale systems, as well as the concomitant myriad of potential causes responsible for an impediment to performance or scalability, will make performance tools indispensable for figuring out what to change to resolve problems that will inevitably arise. Building effective tools will require support from the hardware and all levels of the software stack to enable accurate measurement and attribution of performance problems. Tools must support analysis at all levels of the software stack and not just end-user applications. As a result, new techniques will be needed for measurement, analysis, attribution, and presentation of information about the performance of the software stack on exascale systems. Tools will need to identify code regions with are insufficiently parallel, consume excessive energy, contend for scarce resources, introduce large parallel overhead, or pose scalability bottlenecks. Two kinds of support will be needed. First, tools will need to support post-mortem analysis of executions to identify opportunities for code tuning. Second, performance analysis APIs will need to support on-line introspection and control to enable informed decision making by adaptive runtime systems in response to changing workload characteristics that affect parallelism, data locality, and resource consumption.

5.2 Performance Observation in Exascale Co-Design

Allen Malony (University of Oregon, US)

The exascale co-design challenge demands a new perspective on the role of performance observation as an integral part of the exascale software stack that enables top-down application transformations to be optimized for runtime and system layers by feeding back dynamic information about hardware and software resources from the bottom-up. Performance observation and analysis technology should be an inherent aspect at all exascale levels to make it possible not only to bridge the gap between programming (computation) model semantics and execution model operation, but to deliver opportunities for online, adaptive optimization and control.

The reliance on post-mortem analysis of low-level performance measurements is prohibitive for exascale because of the performance data volume, the primitive basis for performance data attribution, and the inability to reflect back execution dynamics at runtime. With a multi-level exascale programming stack involving high-level transformations, it is necessary to provide richer context for attributions, beyond code locations and simple program events, together with a programmable, hierarchical, and dynamic "performance backplane" with model-driven measurement and analysis, and meaningful mapping back to program performance abstractions.

The perspective can go beyond the exascale software stack to consider how certain performance observation, computational semantics, and feedback support can be implemented in the exascale system architecture and what advantages it may entail. Thinking here could lead to the creation of new hardware technology to specifically to make the exascale machine more performance-aware and performance-adaptive.

5.3 Software Engineering for HPC – A Gap to be Closed

Felix Wolf (German Resarch School for Simulation Sciences, DE)

The idea of co-designing HPC applications together with systems, and system software suggests that the isolated design of applications is already understood – an obvious misconception that becomes clear when browsing standard books on software engineering. Classic plan-based software engineering methodologies usually fail to address the needs of highly research-oriented projects with underspecified functional requirements. In addition, developers of exascale applications will have to satisfy a number of critical non-functional requirements, such as node-level performance, locality, scalability, energy efficiency, and fault-tolerance. Their successful realization presents not only a genuine methodological challenge in itself but aggravates the software engineering crisis of the HPC sector even further. Whereas the development of enterprise software is understood today to a degree where solutions to most of the recurring requirements are embedded in sophisticated generic application containers or platforms such as JavaEE, software engineering for HPC is still in its infancy. We therefore need to rethink our software engineering practices – seeking orientation along recurring problems found across larger classes of applications. This should start with the cartography of those problem domains, ideally followed by the creation of adequate and easily accessible solutions to be co-designed together with other components of the exascale hardware and software ecosystem.

5.4 The classical What-IF-Question: What happens if we are not successful with Hardware or Software on Exascale Level?

Wolfgang E. Nagel (TU Dresden, DE)

Since a couple of years, exascale computing is used as the motivation buzzword to highlight the short-comings we have experienced in the development of HPC systems in many years. We have to recognize that technology is driving hardware development and software is following very slowly the given architectural trends. We had success with addressing Linpack performance, nevertheless we failed in many other areas like scalability, sustained applications performance, and I/O.

The talk will summarize a couple of facts and identify some strategies which could help to make reasonable progress in HPC computing, may be as a side effect even to exascale computing.

5.5 Programming must become easier – not more complex!

Bernd Mohr (Jülich Supercomputing Centre, DE)

The current designs of exascale computer systems show two important trends:

- A huge number of execution units
- A three-level HW hierarchy
  1. Nodes
  2. General purpose CPUs
  3. Performance boosters (Accelerators, SIMD vector units, etc.)

Programming such systems currently requires to use a corresponding hybrid model with different programming models for each of the three levels:

1. MPI
2. Shared-memory multi-threading (OpenMP, Pthreads, ...)
3. OpenCL, CUDA, vector intrinsics, ...

Handling of data and work distribution, communication, and synchronization has to be handled explicitly at every level. In addition, currently there are no standardized interactions/interoperability between the programming models on the different levels. This is way too complex.

What is needed?

- Keep the explicit programming with MPI for expressing data distribution, communication, and synchronization between nodes. There is a lot of experience and codes here that are impossible to ignore. In part, PGAS 1-sided communication could be used where it makes sense and is more efficient. MPI should further evolve and needs to clearly

define interoperability with PGAS and intra-node level multi-threading. Tuning of the inter-node level is also done explicitly (with tool support).
- Hardware levels 2 and 3 should be programmed by one portable programming model. Compiler technology, smart runtime systems, and libraries need to hide the underlying complexity of the hardware. This programming model should be implicit: the programmer expresses opportunities of independent code execution of various granularity. Then compilers and the associated runtimes take care of efficient and fault-tolerant (e.g., by reexecuting failed executions) execution. Tuning on this level is also done implicitly and automatically by the runtime balancing performance and power efficiency.

5.6 Parallel Programming and Execution Models for Exascale – Evolution or Revolution?

Bettina Krammer (University of Versailles, FR)

Are we going to have fundamentally new parallel programming models for exascale in a few years from now? Probably not. It seems naive to throw the past 20 years' experience in parallel programming overboard in the hope to find something new and better, ready for exascale, in the next 5 years. Support has to continue for huge amounts of legacy codes, predominantly written in C, C++ and/or FORTRAN, relying on standard APIs for parallelisation such as MPI, OpenMP or pthreads. Models emerging over the last years, e.g. PGAS (UPC, Co-array FORTRAN, . . .), cilk, OpenCL, openacc, . . . often lack maturity or portability across a wide range of platforms, or still fail to deliver performance on large-scale.

Applications, programming models and, not to forget, the underlying runtime implementations will have to evolve: away from pure MPI codes towards hybrid codes, combining MPI for inter-node communication e.g. with a shared-memory model inside a node or offload directives for hardware accelerators. Interfaces between programming models will have to be well-defined, and pressure on underlying runtime environments will increase to map (hybrid) programming concepts efficiently to the underlying hardware while hiding as much complexity as possible from the user.

5.7 Unifying Scalability Infrastructures

Barton P. Miller (University of Wisconsin – Madison, US)

A key mechanism for the scalability of large (or massive) scale systems is the tree-based communication network, sometimes called a tree-based overlay network (TBON). These TBON's provide the ability to perform command, control, and data gathering and reduction with logarithmic cost. In supercomputers, such networks are used extensively at the hardware level (for diagnostics, initialization, and maintenance), operating system level (for booting the system and launching programs), I/O system level (for scalable file system operations), tool level (for debugging and performance monitoring) and application level (for constructing scalable applications).

Real systems evolve with independent, individual tree-based infrastructures at every level. This is, at best, wasteful and, at worst, generates execution inefficiencies and interference. There is a strong need to a unified, cross-cutting design for such TBON facilities. These facilities need to handle a variety of clients with a variety of topology needs. In addition, different uses of the TBON infrastructure will need different amount of persistence and requirements for fault tolerance. While the challenge for designing such an infrastructure is great, the payoff can be huge.

5.8 Using Application Proxies for Co-Design of Future HPC Computer Systems and Applications

Alice Koniges (Lawrence Berkeley National Laboratory, US)

The high performance computing community is in the midst of disruptive architectural changes. The advent of manycore and heterogeneous computing nodes, increased use of vectorization, light-weight threads and thread concurrency, GPUs, along with concerns about energy and resilience, force us to reconsider every aspect of the computer system, software and application stack, often simultaneously. One important toolset to aid the exploration of this complex design space is application proxies. Although addressing the needs of full-scale applications is the final target of any computer system and tools design effort, working directly with these codes early in the design process is practically limited and time-consuming. In contrast, working with application proxies is much easier and can provide tremendous value, if the proxies are properly designed and results properly interpreted.

There are a wide variety of available application proxies including traditional offerings (NAS Parallel Benchmarks, High Performance Linpack, etc.) that have been used over the years to evaluate system performance. For co-design, a more focused effort of proxies called compact apps and miniapp can permit a broader collection of activities, including completely rewriting them in new and emerging language paradigms. We propose to the community that a broad-based international effort of providing and evaluating compact apps and miniapps can give co-design a strong application base. Included in these studies should be linked websites, test cases, and reporting of results on a larger scale.

5.9 Resource (Energy) Saving System Software Stack for Exascale Systems

Shinji Sumimoto (Fujitsu – Kawasaki, JP)

Predicted hardware specification for exascale systems will be very power sensitive and the amount of hardware resources, such as memory, processor, interconnect and etc, should be minimized. Moreover, processing not related main calculation processing such as OS, house keeping, monitoring, processing related adaptive functions, should be minimized.

However, current software stacks consume much hardware resources because of full function OS, libraries, and the other software. Therefore, current style of software stack must be re-organized and re-structured.

Some of discussion points are as follows:

- Should we provide current style of software stack for exascale systems?
- How should we provide functions needed for exascale systems?
- How should we design and building software stacks for exascale systems?

My idea for this issues are:

- Building software stack layer using less number of stacks.
- Minimizing provided functions only use at runtime.
- Dividing current software stack functions at runtime to pre-execution functions and real runtime functions, and pre-exection functions are realized as runtime optimization tools.

6 Exascale Systems (Co-Design with Apps, System-Software for Performance / Power / Reliability)

6.1 Resource Aware Programming

Michael Gerndt (TU München, DE)

Resource management on current HPC systems is based on space sharing. The cores of the systems are statically partitioned and assigned to jobs. The jobs have exclusive access to those resources for their entire runtime.

Exascale systems will offer millions of cores and only very few applications will be available to use those resources efficiently over their entire execution. To increase the number of applications that benefit from exascale systems, the resource management should adapt the resources to the scalability of application phases.

Within the Special Research Area TRR 89 Invasive Computing, TUM investigates the concept of resource aware programming in the context of HPC. The invasive computing paradigm allows applications to dynamically invade additional resources, to infect those resources with computational tasks, and to retreat from the resources if they are no longer required. This concept will very well match the current application trend to use adaptive algorithms instead of fixed resolution discretisation to improve the science per flop metric.

While the TRR focuses on an X10 based implementation of this concept for manycore processors, TUM is developing iOMP, an extension of OpenMP with invade, retreat, and infect operations. iOMP will be used to proof the advantages of invasive applications for Scientific Computing on NUMA architectures. In the future, a similar implementation will be investigated for distributed memory systems. Whether MPI is the best choice here is unclear. May be: Clouds as a similar large scale infrastructure with the required resource management.

6.2 Relating Exascale to Parallelism, Power and Energy

Zhiwei Xu (Chinese Academy of Sciences, CN)

To realize sustained exascale performance for targeted applications, a co-design team needs to explore a design space with millions of options. In addition to benchmarking and meticulous modeling, a third approach is needed: relating performance to architectural contributing factors with a small set of simple equations. These simple equations (rules of thumb) could serve as a first-order approximation to weed out "obviously" undesirable design decisions and to suggest promising research directions. Traditional rules, such as Little's law, Amdahl's law and other scaling laws, do not consider power and energy consumption. We propose a simple rule as a starting point for discussion. This equation relates the threads per second performance to number of active threads, Watts per thread, and threads per Joule.

7 Exascale Systems

7.1 Characterizing Application-Architecture Co-Design by Suitability Functions

Vladimir Getov (University of Westminster – London, GB)

Building on previous results this work proposes an abstract model for characterizing the couple application-architecture and a methodology for their co-design. Two sets of parameters are first introduced based on a typical high-end computer architecture. A set of suitability functions are then defined using those parameters. These functions could be used for both optimizing the application-architecture co-design as well as for scalability and comparative performance analysis.

7.2 Software Co-Design for Exascale

Erwin Laure (KTH Stockholm, SE)

In order to cope with the challenges of exascale computing, particularly the massive amounts of heterogeneous processing units, the deep memory hierarchies, and the deep and heterogeneous communication facilities, application developers need support in all phases of the application life-cycle, from programming models that allow the construction of efficient, yet portable, applications, over advanced compilation techniques and adaptive runtime environments to online and offline debugging, and performance analysis tools.

Some of the aspects to be tackled include heterogeneous programming models that allow to exploit every bit of parallelism, advanced runtime support through autotuning and dynamic adaptation, and performance tools that not only are capable to handle the enormous data sets resulting from monitoring peta- and exascale applications but also include a full-system view including components shared with other applications like the interconnect and IO-subsystem.

These tool and method developments need to happen in close collaboration with application developers in co-design teams, as prototyped e.g. by the CRESTA and ScalaLife projects.

8 Exascale Architecture (Design, Execution Models, Power, Reliability)

8.1 Co-Design Challenges of Many-Core Systems-on-Chip

Daniel Molka (TU Dresden, DE)

Hardware development is driven by high volume markets (clouds, webserver, etc.) that pay for the R&D. Thus, the influence of the HPC-community on hardware development is limited. Only small changes/additions, that do not jeopardize other areas of application, are realistic. This could include mechanisms to disable/configure certain features (e.g. prefetcher, coherence).

Future HPC systems will likely consist of many compute nodes equipped with manycore processors. Communication characteristics within a node will be totally different from inter-node communication. In order to achieve a high utilization, the fast intra-node data exchange needs to be exploited by software. Therefore a hierarchical approach for programming is necessary, i.e. coarse grained parallelisation to distribute work between the nodes combined with fine grained parallelisation within the (shared memory) node.

Within the nodes cache coherence will probably continue to be the major boundary for scalability. At the moment there is hardly any cooperation of hardware and software in this regard. A lot of knowledge could be extracted from software/runtime regarding the (potential) sharing of certain data. However, the hardware does not use this. Instead operating systems introduce sharing where none is intended (process migration) and the hardware expensively figures out the actual sharing behavior at runtime. A lot of overhead could be avoided if the hardware coherence mechanisms could be bypassed if software can guarantee that the accessed data is not shared.

8.2 Experiences in Co-Design: Tackling the challenges of Performance, Power, and Reliability

Darren Kerbyson (Pacific Northwest National Lab., US)

In a co-design process two or more factors are optimized in concert to achieve a better solution. Though often considered as the joint design of hardware and software, co-design can taken many forms that can include: algorithms, applications, programming models, runtime systems, and hardware architectures. In addition, as we progress towards exascale systems and applications, the power and resiliency requirements in addition to performance define a multidimensional optimization space. No co-design process to date has covered all factors in a comprehensive fashion, however some notable cases have addressed a subset of factors and the corresponding trade-offs.

Our own co-design experiences have enabled improvements to be made in performance, energy efficiency, and reliability. In the co-design for performance, we used performance models to determine the best configuration of an application and the best (out of several choices) of architectures that ultimately became the first petascale system – the IBM Roadrunner. In the co-design for energy efficiency, we used dynamic performance models to identify periods of idleness in applications and coupled this with a runtime that could lower power consumption on available resources. In the co-design for Fault-Tolerance, a programming model was extended and used by applications to expose critical data (data required for subsequent recovery from node-level faults) with a runtime system maintaining multiple copies across the system. These experiences will provide a view on the value of modeling for the co-design of future hardware and software.

8.3 Co-Design & Resiliency for Exascale Computing

Stephen L. Scott (Oak Ridge National Lab., US)

By whatever terms you want to call it – Fault-Tolerance, Resilience, Reliability – this is one broad cross-cutting area where failure has the potential to negatively impact everyone's pursuit of exascale computing.

Thus it should be considered in the co-design of any hardware, system software, or application targeting exascale computing.

(I will use the term "resilience" to encompass all of the above terms from here forward.)

Today, there seems to be more questions surrounding resilience than there are answers. Certainly part of the problem is the unknown – even to the extent that some dispute that failure will be any more of a problem in exascale than we have today in petascale, or yesterday at terascale. We have trudged along living with unreliable systems to this day due to luck and perhaps more so, because of our own ignorance in recognizing failure versus inconsistency and unknown "issues".

From past experience, the resilience research community cannot successfully impact computing reliability to a significant degree without directly engaging the wider computing community. Therefore, I would like this session to initiate a dialogue with a broad cross section of folks regarding where the resilience research community should focus its efforts for the greatest impact throughout the co-design process.

9 Application Level Issues

9.1 Co-Designing for Online Auto Tuning

Jeff Hollingsworth (University of Maryland – College Park, US)

For exascale systems, continuous auto-tuning will be mandatory. In the past, it was possible to auto-tune (or perhaps even manually tune) a code for a specific machine. At exascale, it will be necessary for auto-tuning to be a continuous and ongoing activity throughout each execution of the application.

This requirement is driven by several factors. First, the sheer number of threads (on the order of one billion) will mean that dynamic load balancing will almost certainly be required. Second, the presence of energy limits (either in the form of total consumption or thermal throttling due to heat dissipation) will mean that the performance of cores will be dynamically varying. This will mean that the exact performance of the hardware will not be known until runtime. Finally, it is likely that hardware failures and faults will be the norm and not the exception. This will also contribute to dynamic changes the hardware available to run applications.

The applications themselves will also have increasing needs for auto-tuning. For example, Adaptive Mesh Refinement (AMR) techniques naturally have opportunities to expose choices about meshing parameters and frequency of mesh updates.

Limited storage and data migration will likely drive applications to be "super coupled". Super coupling will mean that not only will multiple physics/chemistry be coupled in a single execution, but previously separate phases of execution such as data staging, analytics, and visualization will be done concurrently with the application execution.

Taken together these trends imply that humans will no longer be able to tune such a system since the reaction time will likely be on the order of milliseconds to seconds and not once per port to a new machine. Thus automation is our only hope.

9.2 Exascale Co-Design – Data

Achim Streit (KIT – SCC, DE)

Scaling HPC systems and its applications beyond petascale towards exascale computing performance is a large challenge for science and industry today. However not only the raw computing capabilities of such systems and the scalability of the applications need to be addressed: big data also needs to be taken into account in the context of exascale systems.

The data generated by exascale simulations will also be in the exascale regime or beyond and in consequence the analysis of this data will need new paradigms, algorithms, methods and tools as well. Arising questions in this context are: How can data efficiently be transferred from the HPC system to the associated storage hierarchy at the HPC site? To what extend must the storage hierarchy expand with the computing capabilities? Are today's commonly used technologies and design principles of storage systems mature enough for exascale computing systems? Do the data management software technologies scale with the increasing demand? How to assure data integrity in light of the large amount of entities (core, memory, disk) in exascale systems?

Focusing more on the users of future exascale systems in the data context brings up the questions of I/O performance to and from the HPC sites (not inside the exascale system): How can users efficiently transfer the presumably also increasing input and output data from remote to the HPC site and vice-versa? Are data transfer and federated data management technologies mature enough for the foreseen large quantities of data to be transferred? Will HPC sites – partly against their will – become data sinks?

As today's and most likely the future HPC ecosystem consists of a heterogeneous set of systems layered in tiers of computing performance, users will always use more than one system during their research lifecycle – due to computing grants, scalability enhancement, or usage of different architectures in multi-scale, multi-physics codes. Consequently one key question is how a federated research data infrastructure looks like? What requirements exist for WAN connections, data movers and federated data management systems? Is it probably easier and faster, if users fedex' disks with large scale data to and from HPC sites?

Finally coming back to the application enabling aspects, one must ask, whether exascale systems will only be usable by an elite set of users and is it enough to focus all efforts only the top of the performance pyramid? Isn't it also mandatory to scale-up the next tier of performance and its users in order to continue the movability of users scaling up their CSE applications – comparable to professional sports vs. grassroot sports (in German: Leistungssport vs. Breitensport).

9.3 Software-Software Co-Design of Applications and Tools

Karl Fürlinger (LMU München, DE)

Co-Design commonly refers to the synoptic development of hardware and software for a specific application domain. In a similar fashion, benefits can be reaped from integrating various software components which are typically developed in isolation from each other. For example, performance and debugging tools are usually only used when necessary to locate a bug or to tune performance while most applications contain crude performance and correctness indicators (timers, counters) put in place by the developers.

We argue that this situation presents an opportunity for software-software co-design. Historically separate components (the application and the tool) can be turned into a unit that performs the desired function better than a loose coupling would allow. The ad-hoc performance or correctness indicators in applications can be replaced by tried-and-true performance tools technology which supports more advanced performance indicators such as hardware performance counters and has controllable overheads.

For this to be feasible, a lightweight always-on monitoring component has to be developed and an interface to access the data in a standardized way needs to be specified. The resulting performance data can then be made available to a variety of consumers. Firstly, the application itself can benefit form the data in through performance introspection. Algorithms can automatically react to load imbalances or other varying performance conditions. Secondly, other middleware layers and the operators of computing centers can benefit from detailed data of their computing center workloads to drive future procurement decisions, for example.

9.4 How Can We Make Co-Design Really Work?

Cherri M. Pancake (Oregon State University, US)

While there are many challenges in moving to exascale, I'm particularly concerned by two issues: (1) ensuring that it will be possible to get sufficient application efficiency to warrant system costs; and (2) coming up with software development models that will allow more than a handful of people to take advantage of the new capabilities.

The workshop topic co-design process is our chance to address the first issue. The only way to ensure that future systems will provide reasonable levels of efficiency on real problems is to engage application developers effectively in the design process. I emphasize the word effectively, because there have been both successes and abject failures with past co-design efforts. As a former anthropologist who has participated on a number of HPC standards groups, I suggest that the most critical aspects of co-design are convening the right groups and ensuring that design truly reflects the issues that each group brings to the table. I would like to discuss ways to involve the "right" users in co-design efforts, as well as how to deal with interpretation and mediation of what are sometimes conflicting viewpoints within a co-design group. Exascale efforts could learn a lot from the experiences of other domains where co-design and standardization already have a long history and are well understood.

My second issue falls under the workshop topics programming models and application level issues. I believe that, rather than relying on our traditional notions of a programming model, it will be necessary to offer software developers a smorgasbord of targeted methods, libraries, and tools. The real challenge is getting users to adopt new methods. Experience at existing scales has shown that they simply are not willing to start from first principles just to use a new model. Instead, users have been attracted to libraries and methods that are clearly related to their application domains, such as higher-level libraries that manage low-level message passing or memory management for them. These provide some level of abstraction – i.e., operations that clearly relate to their application domain – without requiring that they embrace a full scale programming model. They have the added advantage of allowing a single application to combine a mix of techniques. So I think what will be needed for exascale applications is a mix of approaches, where targeted libraries/tools provide higher-level support for different classes of operations such as map/reduce, multiresolution problems, streaming, etc.

Participants

- Arndt Bode
TU München &
LRZ Garching, DE
- Hans-Joachim Bungartz
TU München, DE
- Kent Czechowski
Georgia Inst. of Technology, US
- Karl Fürlinger
LMU München, DE
- Hans Michael Gerndt
TU München, DE
- Vladimir Getov
University of Westminster – London, GB
- Adolfy Hoisie
Pacific Northwest Nat. Lab., US
- Jeffrey K. Hollingsworth
University of Maryland – College Park, US
- Darren Kerbyson
Pacific Northwest Nat. Lab., US
- Alice E. Koniges
Lawrence Berkeley National Laboratory, US
- Bettina Krammer
University of Versailles, FR
- Dieter Kranzlmüller
LMU München &
LRZ Garching, DE
- Erwin Laure
KTH Stockholm, SE
- Allen Malony
University of Oregon, US
- John Mellor-Crummey
Rice University, US
- Barton P. Miller
University of Wisconsin – Madison, US
- Bernd Mohr
Jülich Supercomputing Centre, DE
- Daniel Molka
TU Dresden, DE
- Wolfgang E. Nagel
TU Dresden, DE
- Cherri M. Pancake
Oregon State University, US
- Michael M. Resch
Universität Stuttgart, DE
- Sabine Roller
German Resarch School for Simulation Sciences, DE
- Martin Schulz
LLNL – Livermore, US
- Stephen L. Scott
Oak Ridge National Lab., US
- Christian Straube
LMU München, DE
- Achim Streit
KIT – SCC, DE
- Shinji Sumimoto
Fujitsu – Kawasaki, JP
- Jan Treibig
Univ. Erlangen-Nürnberg, DE
- Felix Wolf
German Resarch School for Simulation Sciences, DE
- Zhiwei Xu
Chinese Academy of Sciences, CN

Report from Dagstuhl Seminar 12221

Cognitive Approaches for the Semantic Web

Edited by

Dedre Gentner[1], Frank van Harmelen[2], Pascal Hitzler[3], Krzysztof Janowicz[4], and Kai-Uwe Kühnberger[5]

1 **Northwestern University, Evanston, US,** gentner@northwestern.edu
2 **Vrije Universiteit Amsterdam, NL,** frank.van.harmelen@cs.vu.nl
3 **Kno.e.sis Center, Wright State University, Dayton, US,** pascal.hitzler@wright.edu
4 **University of California, Santa Barbara, US,** jano@geog.ucsb.edu
5 **Universität Osnabrück, DE,** kkuehnbe@uos.de

Abstract

A major focus in the design of Semantic Web ontology languages used to be on finding a suitable balance between the expressivity of the language and the tractability of reasoning services defined over this language. This focus mirrors the original vision of a Web composed of machine readable and understandable data. Similarly to the classical Web a few years ago, the attention is recently shifting towards a user-centric vision of the Semantic Web. Essentially, the information stored on the Web is from and for humans. This new focus is not only reflected in the fast growing Linked Data Web but also in the increasing influence of research from cognitive science, human computer interaction, and machine-learning. Cognitive aspects emerge as an essential ingredient for future work on knowledge acquisition, representation, reasoning, and interactions on the Semantic Web. Visual interfaces have to support semantic-based retrieval and at the same time hide the complexity of the underlying reasoning machinery from the user. Analogical and similarity-based reasoning should assist users in browsing and navigating through the rapidly increasing amount of information. Instead of pre-defined conceptualizations of the world, the selection and conceptualization of relevant information has to be tailored to the user's context on-the-fly. This involves work on ontology modularization and context-awareness, but also approaches from ecological psychology such as affordance theory which also plays an increasing role in robotics and AI. During the Dagstuhl Seminar 12221 we discussed the most promising ways to move forward on the vision of bringing findings from cognitive science to the Semantic Web, and to create synergies between the different areas of research. While the seminar focused on the use of cognitive engineering for a user-centric Semantic Web, it also discussed the reverse direction, i.e., how can the Semantic Web work on knowledge representation and reasoning feed back to the cognitive science community.

Seminar 28. May–01. June, 2012 – www.dagstuhl.de/12221

1998 ACM Subject Classification I.2 Artificial Intelligence, I.2.4 Knowledge Representation Formalisms and Methods, H.1.2 User/Machine Systems

Keywords and phrases Cognitive methods, Semantic Web, Analogy and similarity-based reasoning, Semantic heterogeneity and context, Symbol grounding, Emerging semantics, Commonsense reasoning

Digital Object Identifier 10.4230/DagRep.2.5.93

Edited in cooperation with Cong Wang

1 Executive Summary

Dedre Gentner
Frank van Harmelen
Pascal Hitzler
Krzysztof Janowicz
Kai-Uwe Kühnberger

The Dagstuhl Seminar 12221 on *Cognitive Appraoches for the Semantic Web* was held from May 28th to June 1st, co-organized by Dedre Gentner, Frank van Harmelen, Pascal Hitzler, Krzysztof Janowicz and Kai-Uwe Kühnberger. The motivation of this seminar was to gather people from Semantic Web and Cognitive Science in order to determine the most promising ways to move forward on the vision of bringing findings from cognitive science to the Semantic Web, and to create synergies between the different areas of research. The seminar mainly focused on the use of cognitive engineering methods towards a more user-centric Semantic Web. However, the reverse direction, i.e., how Semantic Web research on knowledge representation and reasoning can feed back to the cognitive science community, was also discussed. Besides core members of the Semantic Web, artificial intelligence, and cognitive science communities, the researchers from fields that would benefit most from a more human-centric Semantic Web were also present. This especially included experts on Geographic Information Science (GIScience), the bioinformatics, as well as the digital humanities. While the invitations were balanced, most attending participants were from the Semantic Web, cognitive science, and GIScience communities.

The seminar consisted of three alternating blocks, short talks by the participants, work in breakout groups, and reports by the breakout groups followed by discussions among all participants. The short talks presented the participants' research or future ideas and were the inspiration for the topics discussed in the breakout groups. Each day had a distinct subtopic with respect to the combination of presenters and the formed breakout groups. while the task of the breakout groups differed, it was ensured that each of the 5-7 groups consists of members of all major research domains present at the meeting.

On May 29th, the first day of the seminar, Krzysztof Janowicz gave a short opening talk about the structure of the seminar. Next, Frank van Harmelen gave an overview talk about the Semantic Web, while Dedre Gentner introduced the cognitive science perspective focusing on work on analogies. After lunch, the participants, Rob Goldstone, Christian Freksa, Ken Forbus, Kai-Uwe Kühnberger, Alexander Mehler, Ute Schmid, Gudrun Ziegler, and Helmar Gust, all involved in cognitive science research, presented their work in short talks of 10 minutes. After these talks, breakout groups were formed. The task of each group was to develop a research proposal outline and present it to all participants.

On May 30th, the participants presented their results from the breakout groups. This second day was devoted to researchers from GIScience, bioinformatics, and the digital humanities, as well as work of researchers that already bridged between the Semantic Web and cognitive sickness. The presenters were Andrew Frank, Werner Kuhn, Aldo Gangemi, Cory Henson, David Mark, Krzysztof Janowicz, Giancarlo Guizzardi and Simon Scheider. In the afternoon, the participants formed new breakout groups based on the presented topic. The task was to develop user interfaces and user interaction paradigms that exploit Semantic Web reasoning on the one side and analogy and similarity-based reasoning on the other side.

Finally, the groups reported back to all participants and discusses synergies.

May 31st, started with additional domain talks and was then followed by presentations of core Semantic Web researchers. Presentations were given by Sören Auer, Lael Schooler, Willem Robert van Hage, Zhisheng Huang, Stephan Winter, Christoph Schlieder, Jens Ortmann, Ken Forbus, Alan Bundy, Benjamin Adams, Jérôme Euzenat, Claudia d'Amato, Sebastian Rudolph, Wei Lee Woon and Pascal Hitzler. In the afternoon, the breakout groups were formed to discussed how Cognitive Science can benefit from Semantic Web research. The task was to design an experiment (in most cases involving human participants). Afterwards the breakout group reported back to all participants.

June 1st, last day of the seminar, started with two longer talks (each about 30 min.) that reported back on what Semantic Web researchers learned from cognitive scientists during the meeting as well as the other way around. The first presenter was Jérôme Euzenat representing his view as Semantic Web researcher on the lessons learned. The second presentation was given by Rob Goldstone to illustrate the lessons learned by the cognitive science community. Finally, the seminar concluded with general discussions on future research and feedback about the seminar.

2 Table of Contents

Working Groups

3 Short Talk Abstracts

3.1 Eating Knowledge Soup with a Fork

Alan Bundy (University of Edinburgh, GB)

We describe a project to solve Guesstimation problems by mining numeric information from the Semantic Web. Guesstimation problems are order of magnitude estimates of answers to numerical questions. Recently, we've looked particularly at questions about renewable energy. Proof plans are used to identify the numeric facts needed to answer the question then this information is sought. Various techniques have been developed to reject noise and erroneous data, discover missing information, etc. These include: normalisation to a single significant digit for, i.e., $d * 10^i$, where d is a digit in range 1-9 and i is an integer; clustering values and taking the mode when it dominates, otherwise the median; guessing missing units by exploiting the unique ratios between imperial and metric units. Contexts are used to focus search where the names of individuals are unknown, e.g., makes and models of cars. User interaction is enabled via a drag and drop GUI. Solutions are fallible. We hope to associate uncertainty values with them in future work. although the solution is fallible and human interaction is often required.

3.2 What you say is what I get, what you don't say is what I don't get

Jerôme Euzenat (INRIA Rhône-Alpes, FR)

Main reference J. Euzenat, "Evolving knowledge through negotiation," Dagstuhl Preprint Archive, arXiv:1207.6224v1 [cs.AI], 2012.
URL http://arxiv.org/abs/1207.6224v1

Semantic web information is at the extremities of long pipelines held by human beings. They are at the origin of information and they will consume it either explicitly because the information will be delivered to them in a readable way, or implicitly because the computer processes consuming this information will affect them. Computers are particularly capable of dealing with information the way it is provided to them. However, people may assign to the information they provide a narrower meaning than semantic technologies may consider. This is typically what happens when people do not think their assertions as ambiguous. Model theory, used to provide semantics to the information on the semantic web, is particularly apt at preserving ambiguity and delivering it to the other side of the pipeline. Indeed, it preserves as much interpretations as possible. This quality for reasoning efficiency, becomes a deficiency for accurate communication and meaning preservation. Overcoming it may require either interactive feedback or preservation of the source context.

3.3 Analogical Processing as a Technology for the Semantic Web

Kenneth D. Forbus (Northwestern University – Evanston, US)

The cognitive simulations we have developed for analogical matching, retrieval, and generalization have been used to both explain existing psychological data and make new successful predictions. They also have been engineered for use in performance systems, providing a technology for human-like analogical processing. This talk provides some examples of how these models have been used. It suggests that analogical processing is a natural technology for the Semantic Web, since it uses structured, relational representations and can reason and learn from collections of ground facts. Several issues that should be explored to bridge the gaps between them are also raised

3.4 Spatial Cognition and Commonsense Reasoning

Christian Freksa (Universität Bremen, DE)

Spatial structure is omnipresent in the physical world. This is true for the internal structure of physical objects, for their external relations to one another, for their relation to their environment, and for their relation to an observer inside or outside this environment. Spatial structure also is omnipresent in perception systems across a large variety of modalities, in biological memories, and in the motor mechanisms that cognitive agents and artifacts use for locomotion and for other types of motion, including motion of perceptual organs and motion of information-carrying signals inside and outside the cognitive agents. When motion or other forms of dynamics enter the picture, time and temporal structure are involved in addition: temporal structure is omnipresent in processes and the structure of time places additional constraints on top of the constraints imposed by spatial structures. Constraints impose limitations; they restrict what a system can do. Does this mean that we should avoid the constraints of spatial and temporal structures if we can in order to avoid the limitations? Of course this depends on what we want to do. In my contribution, I discuss approaches to commonsense reasoning in humans and in artificial intelligence. I then present three types of spatial tasks and present different cognitive approaches to solve these tasks. I emphasize the role of spatial and temporal structures to generate simple solution processes.

3.5 Detecting, discovering, and using knowledge patterns on the Semantic Web

Aldo Gangemi (ISTC – CNR – Rome, IT)

The Semantic Web, specially through the Web of Data, is now ready for empirical investigation and practical deployment. One research opportunity consists in how to exploit triple- (or quad-)based knowledge for intelligent/visual analytic tasks, as well as how to detect or discover relevant invariances out of distributed RDF graphs. The talk will present methods based on cognitively-sound knowledge patterns, and some empirical results in using that approach for dataset analysis, Wikipedia data pattern discovery, exploratory search, and robust ontology learning from text.

3.6 The Analogical Mind

Dedre Gentner (Northwestern University – Evanston, US)

Analogical processes are central in human learning and reasoning. Analogical comparison engages a process of structural alignment and mapping that fosters learning and reasoning in at least three distinct ways: it highlights common relational systems; it promotes inferences; and it calls attention to potentially important differences between situations. It can also lead to re-representing the situations in ways that reveal new facets. An important outcome of analogical comparison is that the common relational structure becomes more salient and more available for transfer-in short, a portable abstraction is formed. Thus, structure-mapping processes bootstrap much of human learning. The power of analogy is amplified by language learning. Hearing a common label invites comparison between the referents, and this structure- mapping process yields insight into the meaning of the term. The mutual facilitation of analogical processing and relational language contributes to the power and flexibility of human learning and reasoning.

3.7 Connecting Concepts to the World and Each Other

Robert L. Goldstone (Indiana Univ. – Bloomington, US)

According to an "external grounding" theory of meaning, a concept's meaning depends on its connection to the external world. By a "conceptual web" account, a concept's meaning depends on its relations to other concepts within the same system. We explore one aspect of meaning, the identification of matching concepts across systems (e.g. people, theories, or cultures). We present a computational algorithm called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems for Translation) that uses only within-system similarity relations to find between-system translations. While illustrating the sufficiency of a conceptual web account for translating between systems, simulations of ABSURDIST also indicate powerful synergistic interactions between intrinsic, within-system information and extrinsic information. Preliminary applications of the algorithm to issues in object recognition, shape analysis, automatic translation, human analogy and comparison making, pattern matching, neural network interpretation, and statistical analysis are described. ABSURDIST is then generalized to accommodate labeled, unweighted, and directed graphs. This generalization is then applied to automated database schema alignment. For this application, it is necessary to have an automatic way of creating structured representations. To this end, we created weighted graph edges between schema elements by computing information-based entropy relations, semantic similarity proxies by web search query hit overlap, and lexical overlap among labels via string edit distance. This extended system is able to align databases with respectable accuracy.

3.8 Structure Transfer and Modeling Analogies: The Role of Patterns in Ontology-Driven Conceptual Modeling

Giancarlo Guizzardi (UFES – Vitoria, BR)

Conceptual Modeling in Computer Science is about representing aspects of a given subject domain for the purposes of Communication, Domain Understanding and Learning, and Problem Solving. Ontology-Driven Conceptual Modeling is an area which employs methods, chiefly, from Formal Ontology in Philosophy, but also Cognitive Science, Linguistics and Logics to improve the theory and practice of Conceptual Modeling. In the past decade, theories from the aforementioned areas have been successfully employed to derive a number of engineering tools for conceptual model-ing, including modeling languages and methodologies, computational tools and knowledge patterns. In this talk, I concentrate on the latter, arguing that patterns are a device for structural transferability which affords the use of higher-granularity modeling primitives (or "modeling analogies") that can reduce the complexity of both the tasks of model construction and model understanding. In particular, I will elaborate on four different types of conceptual modeling patterns which can be derived from ontological well-founded theories, namely: (i) modeling patterns (for capturing standard solutions to recurrent modeling problems), analysis patterns (for detecting properties in a model), transformation patterns (for representing design strategies for mapping expressive conceptual models to less expressive but computationally interesting languages), validation anti-patterns (for detecting deviations from sets of possible models and intended models) and pattern languages.

3.9 Creating and Integrating Micro Domain Theories

Helmar Gust (Universität Osnabrück, DE)

Modeling heterogeneous knowledge like human background knowledge or knowledge distributed in the Web is an widely unsolved problem. Current knowledge representation schemata are still quite static. Problems occur when the relevant knowledge needed in a problem solving situation must be determined. Although semantic Web approaches try to support and integrate distributed domain ontologies, this does not reflect the highly dynamic nature of constructing the relevant knowledge needed in a given situation. The presentation tries to grasp the problem and demonstrates some first ideas for (1) modularizing knowledge on a very fine grained scale and (2) integrating the knowledge of the relevant micro domains needed in a problem solving situation on the fly.

3.10 Semantics of Machine Perception

Cory Henson (Wright State University – Dayton, US)

Joint work of Henson, Cory; Sheth, Amit; Thirunarayan, Krishnaprasad
Main reference C. Henson, K. Thirunarayan, A. Sheth, "An Ontological Approach to Focusing Attention and Enhancing Machine Perception on the Web," Applied Ontology, vol. 6(4), pp.345–376, 2011.
URL http://dx.doi.org/10.3233/AO-2011-0100

The acts of observation and perception provide the building blocks for all human knowledge (Locke, 1690); they are the processes from which all ideas are born; and the sole bond connecting ourselves to the world around us. Now, with the advent of sensor networks capable of observation, this world may be directly accessible to machines. Missing from this vision, however, is the ability of machines to glean semantics from observation; to apprehend entities from detected qualities; to perceive. The systematic automation of this ability is the focus of machine perception – the ability of computing machines to sense and interpret the contents of their environment. Despite early successes within narrow domains, analyzing data of a single modality (e.g., facial recognition), a general solution to machine perception remains elusive. This state of affairs is the result of difficult research challenges, such as the ability to model the process of perception in order to efficiently and effectively interpret the growing stream of multimodal (and incomplete) sensor data. People, on the other hand, have evolved sophisticated mechanisms to efficiently perceive their environment; including the use of background knowledge to determine what aspects of the environment to focus attention. Over the years, many cognitive theories of perception have been proposed, evaluated, revised, and evolved within an impressive body of research. These theories present a valuable stepping-stone towards the goal of machine perception, to embody this unique human ability within a computational system. This talk will describe the information processes involved in perception that will serve as an ontological account of knowledge production. The ontology of perception, IntellegO (Greek: "to perceive"), derived from cognitive theories of perception, provides a formal semantics of perception by defining these information processes that enable the conversion of low-level observational data into high-level abstractions. IntellegO is currently being applied within several domain applications, including a weather-alert service, a fire-detecting robot, and a mHealth application to help lower hospital readmission rates for patients with chronic heart disease. We will demonstrate through these examples how massive amounts of multimodal sensory data is converted into contextual knowledge for improved situational awareness.

References

1 C. Henson, K. Thirunarayan, A. Sheth. An Ontological Approach to Focusing Attention and Enhancing Machine Perception on the Web. Applied Ontology, vol. 6(4), pp.345-376, 2011.
2 C. Henson, A. Sheth, K. Thirunarayan. Semantic Perception: Converting Sensory Observations to Abstractions. IEEE Internet Computing, 16(2), pp.26-34, Mar/Apr. 2012.

3.11 Closed World Assumption and Defaults – not the same thing!

Pascal Hitzler (Wright State University – Dayton, US)

Non-monotonic logics used in knowledge representation and reasoning usually provide a uniform mechanism for modeling both, defaults and closed-world features. In the context of ontology modeling, which is fundamentally based on monotonic and open-world logics, it is an ongoing quest how to incorporate defaults and (local) closed world features. We argue that the traditional perspective, which uses one uniform mechanism to provide both these features, leads to unintuitive results. We also provide some preliminary insights, based on our recent work, on how a more satisfactory incorporation of these features could be realized.

3.12 Enabling domain experts to become knowledge engineers

Krzysztof Janowicz (University of California – Santa Barbara, US)

Big Data promises to lead science into a new age in which complex scientific and social questions can be approached in a holistic way by combining multi-thematic and multi-perspective data across different media formats. This makes the integrating of massive amounts of highly heterogeneous data a core challenge for Geographic Information Science. However, Big Data should not be approached by equally big ontologies. Instead, it needs a framework to assist domain experts in becoming knowledge engineers. This talk presents ongoing work on observation-driven ontology engineering that computes semantic signatures as methodology to mine ontological primitives out of observation data and proposes how geo-ontology design patterns may assist domain experts in creating small, data-driven application ontologies. Finally, exploratory user interfaces are discussed to ease access to heterogeneous geo-data.

3.13 Ambient Intelligence, Cognitive Constraints, and Semantics

Kai-Uwe Kühnberger (Universität Osnabrück, DE)

Interaction with computing devices was classically conceived as a dialogue between a user and a computer. Due to the rapid increase of different types of computing devices that can be used for autonomous interactions between such devices as well as user-centered interactions with a network of devices (like the acquisition of knowledge, the controlling of systems, the communication with other agents etc.) the concept of an internet of things is no longer purely visionary. In this short presentation, I will try to speculate about cognitive constraints with respect to the design of such interfaces and the need for cognitively plausible interaction styles, knowledge-intensive systems, and semantically enriched information transfers between different types of devices in order to facilitate the idea of ambient intelligence.

3.14 Image-Schematic Patterns

Werner Kuhn (Universität Münster, DE)

The semantic web is less semantic than most would wish. For example, it believes that mountains have fax numbers (see http://schema.org/Mountain) and that an author is more likely to have a certain age if that number appears as a page number in his or her publications. Yet, a series of ideas from the cognitive sciences promise to allow for a dynamic reconstruction of meaning. In my talk, I argue for the specification of image schemas as ontology design patterns. These schemas can be seen as frame-like structures related to processes. I illustrate the proposal with the PATH schema, realting motion processes to their trajectories, with start and end positions as well as media and surfaces for motion. The hypothesis of this work is that such image-schematic patterns allow for revealing interesting higher level semantics in low-level encodings like those of linked data.

3.15 Conceptual Spaces, Language Evolution & Network Theory

Alexander Mehler (Universität Frankfurt am Main, DE)

Emergent semantics can provide meaningful knowledge representation, and models of language evolution (MoLE) are candidates for it. To date, MoLE is restricted what regards the semantic complexity of predicates. However, we need more expressive models of conceptual spaces in MoLE. Therefore, we consider network theory as a starting point of such representation models.

3.16 Ecological Approaches for the Semantic Web

Jens Ortmann (Universität Münster, DE)

Humans have the unique ability to perceive affordances in their environment and to communicate these affordances to other humans. This ability is vital in decision-making and planning. Yet, the relation between perception and action is hardly reflected in information systems and the observations of affordances have not found their way into the semantic web. In the past three years I have investigated formalizations of affordances and of the results of affordance observations. My objective is to semantically integrate human observations of affordances with each other and with other information sources. Therefore, I have devised a semantic reference system that can account for the subjective observation of affordances. In addition to that, I have formalized the semiotic process of observing affordances in the human-environment system. I believe that an ecological approach, which emphasizes the importance of interactions within systems, is well suited for the semantic web, which can be considered as a system of intelligently linked and sometimes interacting information sources and services. In an ecological approach, the user with her individual capabilities and specific intentions and needs, is a part of the system. The semantics of information sources and services is always taken with reference to the user and reflects the relevance and meaning that these information sources and services have for her. This enable the provisioning of more relevant information that is meaningful with respect to the user's specific opportunities and dangers in her environment.

3.17 Matrix-Space Language Models for Acquisition of Semantic Knowledge

Sebastian Rudolph (KIT – Karlsruhe Institute of Technology, DE)

Joint work of Rudolph, Sebastian; Giesbrecht Eugenie
Main reference S. Rudolph, E. Giesbrecht, "Compositional Matrix-Space Models of Language," in J. Haji;, S. Carberry, S. Clark, J. Nivre (eds.), Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), 907916. Association for Computational Linguistics, 2010.
URL http://www.aclweb.org/anthology/P10-1093

We propose a novel type of generic compositional models for syntactic and semantic aspects of natural language, based on matrix multiplication. We argue for the structural and cognitive plausibility of this model and show that it is able to cover and combine various common compositional NLP approaches ranging from statistical word space models to symbolic grammar formalisms. We speculate on this new paradigm's usefulness in the area of Semantic Technologies.

3.18 The observational roots of reference of the semantic web

Simon Scheider (Universität Münster, DE)

Joint work of Scheider, Simon; Janowicz, Krzysztof; Adams, Benjamin
Main reference S. Scheider, K. Janowicz, B. Adams, "The observational roots of reference of the semantic web," Dagstuhl Preprint Archive, arXiv:1206.6347v1 [cs.AI], 2012.
URL http://arxiv.org/abs/1206.6347v1

Shared reference is an essential aspect of meaning. It is also indispensable for the semantic web, since it enables to weave the global graph, i.e., it allows different users to contribute to an identical referent. For example, an essential kind of referent is a geographic place, to which users may contribute observations. We argue for a human-centric, operational approach towards reference, based on respective human competences. These competences encompass perceptual, cognitive as well as technical ones, and together they allow humans to inter-subjectively refer to a phenomenon in their environment. The technology stack of the semantic web should be extended by such operations. This would allow establishing new kinds of observation-based reference systems that help constrain and integrate the semantic web bottom-up.

3.19 Image-based place models for geographic recommendations

Christoph Schlieder (Universität Bamberg, DE)

Recommender systems for geo-referenced objects exploit the heuristic principle that people who agree in their spatial choices at one place, are likely to agree at other places too. A spatial choice considered in this context is the decision of a tourist to take a photograph from a particular vantage point. Web-based collections of touristic photographs document virtually millions of such choices and constitute a valuable source for training image recommender systems. Understanding which different place models users adopt, is crucial for improving the quality of the recommendations since there is considerable variation in the images associated with urban places such as Amsterdam or Paris. We found that differences in choice frequency need to be taken into account in order to determine how similar two users are with respect to their choices. It turns out that agreement on spatial decisions adopted by only few users constitutes a good predictor for geographic recommendations. This suggests that frequency information (e.g. the number of instances of a class) could also be useful in addressing related problems of semantic modeling.

3.20 Matchmaking – How Similar Is What I Want To What I Get?

Ute Schmid (Universität Bamberg, DE)

Joint work of Schmid, Ute; Michael Munz; Martin Sticht; Klaus Stein
URL http://www.uni-bamberg.de/kogsys/services/forschung/projects/bmbf-project-emn-moves-match-making/

I present part of a newly started BMBF cooperation project in the domain of mobility for senior citizens. Within this project we want to establish a matchmaking service which enables building of mobility chains by matching volunteers, neighbours and senior citizens. Mobility chains start at home, expand into the neighbourhood, the city, and the larger region. Since matchmaking is restricted by spatial nearness, the data base will only contain a moderate amount of data. The main challenge will be to match offers and requests on the level of activities described in different granularities.

3.21 Ranking Query Results from Linked Open Data Using a Simple Cognitive Heuristic

Lael Schooler (MPI für Bildungsforschung, DE)

We address the problem how to select the correct answers to a query from among the partially incorrect answer sets that result from querying the Web of Data. Our hypothesis is that cognitively inspired similarity measures can be exploited to filter the correct answers from the full set of answers. These measures are extremely simple and efficient when compared to those proposed in the literature, while still producing good results. We validate this hypothesis by comparing the performance of our heuristic to human-level performance on a benchmark of queries to Linked Open Data resources. In our experiment, the cognitively inspired similarity heuristic scored within 10% of human performance. This is surprising given the fact that our heuristic is extremely simple and efficient when compared to those proposed in the literature.

3.22 Reasoning gap between human and machine

Cong Wang (Wright State University – Dayton, US)

Many powerful reasoning algorithms have been provided to help machine to be more efficient and intelligent. However, we are still unknown how much difference between machine algorithm and human process, and how to combine them. We try to pursue this by several experiments. 1.Run machine algorithm with a cognitive hint by human, see whether the hint can help machine reasoning. 2.Let human do some reasoning tasks, see whether a hint from machine algorithm can help human. Furthermore, we'd like to see which kinds of reasoning mechanism are most suitable for human. (deductive, inductive, or abductive) Finally, we try to design a system to balance human process and machine algorithm to achieve a better one.

3.23 Taxonomy Generation for Tech-Forecasting

Wei Lee Woon (Masdar Institute – Abu Dhabi, AE)

Tech-mining is the process by which data mining and other automated techniques are used to obtain insights about the growth and evolution of technology. In my research group we have been studying the use of bibliometric techniques for this purpose. One of the problems with these techniques is that growth indices extracted from individual terms can be very noisy as there is often insufficient data; this is particularly true in the case of rarely seen terms – which are frequently the most interesting! The approach which we have taken is to automatically generate keyword taxonomies of research domains, and to use these to aggregate growth indices extracted from multiple keywords. This can help to increase the reliability of the resulting forecasts, and also serves as a useful tool for visualizing the respective research landscapes. In this talk I also plan to discuss a number of challenges faced in this process and the solutions which have been attempted.

3.24 Grouping Semantic Web Query Results: Requirements and Possible Solutions

Claudia d'Amato (University of Bari, IT)

Joint work of d'Amato, Claudia; Fanizzi, NIcola; Lawrinowicz, Agnieszka; Staab, Steffen; Esposito, Floriana

One of the main usages of the Web from a user perspective is querying for searching and finding information. However, queries often result in an overwhelming number of returned answers while typically only a small part of them is relevant. This requires users to perform browsing and exploratory data retrieval of the returned results for finding the real results in which they are interested in.

Querying the Semantic Web rather than the more traditional Web would allow to decrease the number of returned results and increase the number of relevant results. However still they can be numerous for a manual handling. Often, when humans deal with complex tasks or contexts (i.e. made by a large number of information or objects), they do not directly process the single available objects, rather they first create (mental) classes/categories of interest and successively process the elements within the categories. That is, a human factors the task of coping with a complex environment in two different steps: a classification step and a processing step.

The task of understanding (large amount of) retrieved resources/results for distinguishing between relevant and not relevant results with respect to the specified query is a complex task for a human. Moving from the observation above, a natural direction to follow for facilitating the user in browsing retrieved results is to set up and exploit grouping methods and criteria to decompose the problem in:

1. finding the category (or categories) of interest;
2. inspecting the resources belonging to the considered category. The value added of such an approach has been largely recognized in the literature. Indeed, for instance, it is on the ground of some indexing techniques adopted by DBMS.

In this talk, a set of requirements that grouping methods and criteria have to satisfy will be presented. Hence a possible solution, consisting in the exploitation of conceptual clustering methods, will be given. The last part of the talk will concern with discussing two open questions: 1) are there additional requirements that grouping methods and criteria have to satisfy?; 2) are there alternative ways and criteria for grouping?

As regards the envisioned requirements, they are listed in the following. First of all, the created categories cannot be fixed and predefined since one query could be very different from another one and also because the knowledge and information change over the time. Rather, the categories need to be created dynamically and efficiently. Ideally, the categories should be organized in a hierarchy so that the users can easily browse the categories moving from a general to a most specific view and once that the desired category is found the elements of this categories could be inspected. In order to facilitate the browsing of the hierarchy, the categories should be annotated with labels or descriptions summarizing their contents, namely the set on resources belonging to them. The number of semantically annotated resources belonging to a single category (especially a category at a low level in the hierarchy) will be lower than the overall returned results. In this way a minimization of the user efforts and time for inspecting the results can be obtained. Furthermore, resources do not necessarily have to belong to a single category (of a given level of the hierarchy). They may belong to more than one category at the same time. This means that categories (at the same level) do not have to be necessarily disjoint. Lastly, a notion of similarity, that is able to take into account the semantics of the annotated resources, needs be employed for creating the categories.

Conceptual clustering methods [12] satisfy almost all the requirements listed above and as such they could be successfully exploited for the purpose. Clustering algorithms are inductive learning methods that organize collections of objects into meaningful groups (clusters) [7] by the use of a similarity criterion so that the intra-cluster similarity is high and the inter-cluster similarity is low. Conceptual Clustering methods focus on techniques for supplying intensional descriptions of the discovered clusters.

The adoption of clustering methods for grouping semantic web query results has been proposed in [9], where an extension of the SPARQL query language has been proposed. The extension consisted in adding a new grouping clause called CLUSTERED_BY (similarly to the group_by clause of the SQL language) which enables the call of a clustering algorithm for clustering query results. In [2], a similar approach, consisting in extending the SPARQL query language with the CATEGORIZED_BY clause, is proposed. In this latter case, (part of) the is-a hierarchy coming from a reference domain ontology is exploited for grouping the query results. In [3], a conceptual clustering algorithm for grouping semantically annotated resources is exploited for performing automatic and efficient resource retrieval.

Clustering methods [7] may adopt different approaches. The main distinction is between hierarchical (agglomerative or divisional) and partitional approaches. The latter return a flat list of cluster. The former return clusters organized in a structure called *dendrogram* that is a nested grouping of objects and similarity levels at which grouping changes. A dendrongram is mainly a tree that could be broken at different levels to yield different clustering of the data. Hence, hierarchical clustering algorithms have to be considered to satisfy the constraint of having categories organized in a hierarchical structure. Furthermore, the results presented in [3], showed that a hierarchical divisional rather than agglomerative approach should be adopted since the latter one may generate too fine grained hierarchies that may potentially generate an overload of the browsing activity. However, clusters in the dendrogram are generally assumed to be disjoint. In order to be compliant with the requirement that a resource may belong to more than one category (namely a cluster) at the same time, fuzzy clustering methods [7] that are applicable to semantically annotated resources have to be considered [4]. In fuzzy clustering, each instance has a degree of belonging to clusters, rather than belonging completely to just one cluster.

All these algorithms, need a notion of similarity for comparing the annotated resources. A set of semantic similarity measures have been developed [5, 1, 8]. They could be directly plugged into the chosen clustering algorithm.

Last, in order to satisfy all the requirements listed above an intentional cluster description for each cluster belonging to the dendrogram should be provided. For the purpose, inductive methods for learning description logics concept descriptions could be adopted [6, 10].

The existence of several building blocks should in principle make the realization of the presented idea potentially easy. However, many of the presented algorithms require considerable computational effort and as such not immediately usable in a dynamic an run time environment. An important aspect that needs to be investigating concern the realization of these methods in an efficient way. For the purpose incremental learning algorithms [11] could be explored and how to adapt them to an highly dynamic environment should be studied.

References

1 C. d'Amato and S. Staab and N. Fanizzi. On the Influence of Description Logics Ontologies on Conceptual Similarity. In Proc. of the Int. Conf. on Knowledge Engineering: Practice and Patterns (EKAW'08). Vol. 5268. pp. 48–63, Springer, LNCS. (2008).
2 C. d'Amato and N. Fanizzi and A. Lawrynowicz. Categorize by: Deductive Aggregation of Semantic Web Query Results. In Proc. of the Extended Semantic Web Conference (ESWC'10), vol. 6088 pp. 91–105, Springer, LNCS (2010).
3 C. d'Amato and S. Staab and N. Fanizzi and F. Esposito. DL-Link: a Conceptual Clustering Algorithm for Indexing Description Logics Knowledge Bases. Int. Journal of Semantic Computing Vol.4(4) pp. 453–486 (2010).
4 F. Esposito and C. d'Amato and N. Fanizzi. Fuzzy Clustering for Semantic Knowledge Bases. Fundam. Inform. Vol. 99(2) pp. 187–205 (2010)
5 N. Fanizzi and C. d'Amato and F. Esposito. Induction of Optimal Semi-distances for Individuals based on Feature Sets. In Proc. of the Int. Workshop on Description Logics (DL'07). Vol. 250. CEUR Workshop Proceedings. (2007).
6 N. Fanizzi and C. d'Amato and F. Esposito. DL-FOIL Concept Learning in Description Logics. In Proc. of the Int. Conf. on Inductive Logic Programming (ILP'08). Vol. 5194. pp. 107–121 Springer, LNCS (2008).
7 A.K. Jain and M.N. Murty and P.J. Flynn. Data CLustering: A Review. ACM Computing Surveys. Vol.31(3). pp. 264–323. ACM. (1999)
8 K. Janowicz and M. Wilkes SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity. In Proc. of the Extended Semantic Web Conference (ESWC'09), vol. 5554 pp. 353–367, Springer, LNCS (2009).
9 A. Lawrynowicz. Query results clustering by extending SPARQL with CLUSTER BY. In Meersman, R., Herrero, P., Dillon, T.S., eds. OTM Workshops, Vol.5872 pp. 826–835 Springer, LNCS. (2009)
10 J. Lehmann. DL-Learner: Learning Concepts in Description Logics. Journal of Machine Learning Research. Vol. 10 pp. 2639–2642 (2009)
11 T. Mitchell. Machine Learning, McGraw Hill. (1997)
12 R.E. Stepp and R.S. Michalski Conceptual clustering of structured objects: A goal-oriented approach. Journal of Artificial Intelligence. Vol. 28(1) pp. 43–69 (1986)

3.25 Automating Detective Work – discovering story lines on the Web

Willem van Hage (VU University Amsterdam, NL)

In this short talk I outline a task that we will attempt to partially automate in the coming years, the task of revealing the background story behind a current event from information on the Semantic and Word Wide Web. People and computers are notoriously bad at combining facts that are not presented together. To get a good overview of the story behind current events it is necessary to have all the relevant facts in the same place to summarize them. Finding these facts in the first place is a complex task. We will imitate the strategy used by journalists and detectives to step by step explore leads to gather a complete picture.

4 Working Groups

The following subsections present short notes from the working groups.

4.1 The construction and change of representations

Alan Bundy, Frank Jäkel, Helmar Gust, Alexander Mehler, Simon Scheider, and Wei Lee Woon

Our task was to develop a grant proposal about how knowledge representations are constructed, e.g., in the sense of von Glasersfeld, and how representations change, e.g., over time. We discussed the motivation, challenges, and a workplan. A key question discussed was which type of detectable change in the environment is required before it has to be reflected on the conceptual level.

4.2 Heterogeneity and this sort of things

Claudia d'Amato, Gudrun Ziegler, Jérôme Euzenat, and Willem Robert van Hag

Linked open data contains plenty of heterogeneous information from different sources that can be exploited by analogical reasoning. Thus, the following questions are mainly about: 1. Finding as many analogies as possible? 2. Finding the most complete one (largest analogy). 3.Finding largest number of analogous subgraphs. An analogy is a pair (or more) of subgraphs which can be (partially) mapped. Thus they can be used for: 1. identifying matching instances, 2. completing matching instances, 3. deducing instance completion, 4. classifying these subgraphs as an instance of a particular class, 5. suggesting new concepts for ontologies, and so forth.

4.3 Cultural Dependency

Cong Wang, Andrew U. Frank, and Chirstoph Schlieder

Different cultures have different ways of thinking and conceptualizing. We assume these information is not based on hidden knowledge, but on different mechanism. For example, western people usually use deductive reasoning, while Asian people prefer to use inductive or abductive reasoning. Western people act based on plan, but Asian people prefer to be more reluctant to decide. Hence, the questions are how to model culture difference and even how to mine culture difference in huge data.

4.4 Perception and Semantics – Uneasy Bed Fellows

Jens Ortmann, Christian Freksa, Cory Henson, and Wei Lee Woon

How can we design a more intuitive interaction? The key points are based on principles from perceptual theory and use of semantic models of possible solutions. For example, a user would like to find an attractive hiking route near Dagstuhl and has several requirements, e.g., reachable via public transport and achievable with medium fitness. How can the Semantic Web help average user to find a solution which satisfies these? The core idea is to present a small set of prototypes as starting point and let user give feedback, then further specify and generalize possible solutions.

4.5 The long tail (tale) of linked data

Frank van Harmelen, Pascal Hitzler, Christoph Schlieder, and Stefan Winter

Only a small portion of Linked Open Data has been used frequently (mainly provided by crowdsourcing). Therefore, using data about frequency distributions may improve Linked Data algorithms. We need to discover what type of frequency distributions are relevant and how can frequency data be computed on the fly. We can improve Linked Data by measuring quality (highly populated classes are more important), by resource allocation (high frequency $\rightarrow$ high resolution), or by measuring similarity (less popular features $\rightarrow$ more informative).

4.6 Design an Experiment

Alan Bundy, Jérôme Euzenat, Andrew U. Frank, Frank van Hamelen, Cory Henson, Kai-Uwe Kühnberger, Ute Schmid, and Cong Wang

The starting point is to observe how people gather information to plan a travel and learning from this. How can we incorporate such strategies in static ontologies? How can we observe action chains to predict future behavior and use semantic technologies to assist users in performing tasks, e.g., travel. This will need methods from machine learning, e.g., out of semantic trajectories, as well as ontologies and deduction to infer future activities.

4.7 Reasoning-based user interfaces

Gudrun Ziegler, Benjamin Adams, Ken Forbus, Krzysztof Janowicz, Claudia d'Amato, and Pascal Hitzler

It is difficult to take advantage of reasoning-based systems without the ability to ask complex questions. How can users be helped to build queries intuitively. How to design semantics-enabled graphical approaches that support exploratory search and information browsing. A possible set-up is to translate a natural language query into a conjunctive query, together with a graphical representation. In addition, the system should provide support for disambiguation, for selecting predicate names, or navigate based on background ontology. Similarity and analogy based reasoning should be used in addition.

4.8 Imperfection – Feature or Bug?

Helmar Gust, Andrew Frank, Alan Bundy, Lael Shooler, Frank Jäckel, Zhisheng Huang, Cong Wang, Ute Schmid, Christoph Schlieder, and Stephan Winter

Most of the daily data we use is imperfect. There are many types of imperfection, e.g., ambiguity, uncertainty, vagueness, imprecision, granularity, misalignment, mismatches (cultural differences), temporal uncertainty, ignorance or omission, user profiling. Some forms are well researched, while others are not. Certain types of imperfections are more likely to produce problems, e.g., with respect to semantic interoperability. In other cases, the imperfections are well handled by human interpretation. Provenance ontologies should capture the different types of imperfections. We also need a better understanding of which of them are tolerable (and to which degree).

4.9 Knowledge patterns

Benjamin Adams, Aldo Gangemi, Giancarlo Guizzardi, Cory Henson, Krzysztof Janowicz, Werner Kuhn, and Ute Schmid

There are many different ways to think about and define patterns. for instance patterns can be abstract strategies, e.g., logical patterns that help to resolve modeling problems introduced by a particular choice of knowledge representation language, or building block, e.g., to offer a common way to model reoccurring classes and relations such as location or participation in events. What is common to different approaches to patterns is that they are based on some variance in knowledge structures or domain-independent abstraction. In the group we discussed the following questions 1.What are the theories that we use to extract patterns? 2. When to generalize a pattern? 3. What are strategies for learning patterns out of observation data? 4. How many patterns are there to cover most of the common modeling tasks? 5. Are there groups of patterns that usually form in a certain modeling problem?

4.10 Reproducing Data

Sören Auer, Rob Goldstone, and Lael Schooler

Billion of Euro are spend to recreate and reproduce data that could have been reused. Sensor data being a classical example. The value of data increases by copying and reusing. However, there are many challenges to use external data (instead of reproducing it). 1. Coverage: Does the dataset cover the required information at the same spatial, temporal, and topical resolution? 2. Quality and trust: Is the quality of the data sufficient for the use case at hand? do I trust the data source? 3. Structure: Is the syntactic and semantic structure of the data compatible with own datasets? Are there mappings and transformations that can be applied? 4. Fusion: If a single dataset does not cover the needs, is is possible to fuse multiple datasets to obtain the required quality and coverage?

4.11 *Context* Project Proposal

Andrew U. Frank, Christian Freksa, Jens Ortmann, and Kai-Uwe Kühnberger

There are differences between the term context in computer science (and mathematics) on the one hand and context in cognitive science on the other. For instance, in cognitive science, a constructivist view on context is more appropriate. Context determines meaning and can change the interpretation of terms radically. How do we account for context in our ontologies? Is ontology modularization an appropriate approach to contextualization or do we require more flexible and dynamic approaches to typing? A project could investigate how to develop ontologies and KR methods that are more robust to context.

4.12 How people construct trust in Linked Open Data

Jérôme Euzenat, Christoph Schlieder, Simon Scheider, Claudia d'Amato, Giancarlo Guizzardi, and Lael Schooler

How do people establish trust in the quality of Linked Open Data? Do outgoing and incoming links require a different notion of trust. Can such trust models be included as filters into query languages to include or exclude certain parts of a dataset or a dataset federation? Can we develop measures to automatically compute the quality of links?

Participants

- Benjamin Adams
University of California – Santa Barbara, US
- Sören Auer
Universität Leipzig, DE
- Alan Bundy
University of Edinburgh, GB
- Claudia d'Amato
University of Bari, IT
- Jérôme Euzenat
INRIA Rhône-Alpes, FR
- Kenneth D. Forbus
Northwestern University – Evanston, US
- Andrew U. Frank
TU Wien, AT
- Christian Freksa
Universität Bremen, DE
- Aldo Gangemi
ISTC-CNR – Rome, IT
- Dedre Gentner
Northwestern University – Evanston, US
- Robert L. Goldstone
Indiana Univ. – Bloomington, US
- Giancarlo Guizzardi
UFES – Vitoria, BR
- Helmar Gust
Universität Osnabrück, DE
- Cory Henson
Wright State University – Dayton, US
- Pascal Hitzler
Wright State University – Dayton, US
- Zhisheng Huang
VU University Amsterdam, NL
- Frank Jäkel
Universität Osnabrück, DE
- Krzysztof Janowicz
University of California – Santa Barbara, US
- Kai-Uwe Kühnberger
Universität Osnabrück, DE
- Werner Kuhn
Universität Münster, DE
- David M. Mark
SUNY – Buffalo, US
- Alexander Mehler
Goethe-Universität Frankfurt am Main, DE
- Jens Ortmann
SAP – Darmstadt
- Sebastian Rudolph
KIT – Karlsruhe Institute of Technology, DE
- Simon Scheider
Universität Münster, DE
- Christoph Schlieder
Universität Bamberg, DE
- Ute Schmid
Universität Bamberg, DE
- Lael Schooler
MPI für Bildungsforschung, DE
- Willem van Hage
VU – Amsterdam, NL
- Frank van Harmelen
VU – Amsterdam, NL
- Cong Wang
Wright State University – Dayton, US
- Stephan Winter
The University of Melbourne, AU
- Wei Lee Woon
Masdar Inst. – Abu Dhabi, AE
- Gudrun Ziegler
University of Luxembourg, LU