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