SEMINAR
SEMANTICS & SEARCH
April 28th, 2006
Location: **F1**

09.15-
09.20 Opening of seminar
Jon Atle Gulla, NTNU

09.20-
09.50 The Semantic Web and ontologies: what assumptions, what justifications?
Karen Sparck-Jones, Cambridge

Abstract. A great deal is written about the Semantic Web as if it already materially exists. There is also a widespread presumption that the Semantic Web requires an ontology, or some relatable set of ontologies, of a thoroughly logical kind. In my talk I will examine this presumption in relation to the manipulation of, and especially access to, information expressed in natural language. The realities of natural language information processing tasks show that aggressively formal ontologies are effective only for limited domains and communities, and that useful broad-cover, general-purpose ontologies have to be text-derived or at least text-endorsed, and will necessarily be soft and imperfectly logical.

Short CV: Karen Sparck Jones is emeritus Professor of Computers and Information at the Computer Laboratory, University of Cambridge. She has worked in automatic language and information processing research since the late fifties, and has many publications including nine books. She is a Fellow of the British Academy and of the American Association for Artificial Intelligence. She has received three awards for information retrieval research as well as, in 2004, the Association for Computational Linguistics' Lifetime Achievement Award. Her more recent research has been on information retrieval models and practice, on automatic summarising, and on system evaluation, where she is involved in international programmes.
Publicity information:
http://www.cl.cam.ac.uk/~ksj

09.50-
10.20 Emerging Semantic Web Trends: Transparent and Trustworthy Applications
Deborah L. McGuinness, Stanford

Abstract. As web applications proliferate, more users (both people and agents) find themselves faced with decisions about when and why to trust application advice. In order to trust information obtained from arbitrary applications, users need to understand how the information was obtained and what it depended upon. Particularly in web applications that may use question answering systems that may be heuristic or incomplete or data that is either of unknown origin or may be out of date, it becomes more important to have information about how answers were obtained. Emerging web systems will return answers augmented with Meta information about how answers were obtained. In this talk, Deborah McGuinness will describe an approach that can improve trust in answers generated from web applications by making the answer process more transparent. The added information is aimed to provide users (humans or agents) with answers to questions of trust, reliability, recency, and applicability. While this is an area of active research, there are technologies and implementations that can be used today to increase application trustability. The talk will include descriptions of a few representative applications using this approach.
Publicity information:
http://www.ksl.stanford.edu/people/dlm/publicity.html

10.20 Coffee Break (30 min)

10.50-
11.20 XML full-text search and entity extraction
Aleksander Øhrn, Fast Search & Transfer

Abstract: I'll present some of the development activities done at Fast Search & Transfer where scalable search, XML technologies and entity extraction are brought together in order to form powerful search and discovery mechanisms. The combination of being able to express queries that involve semantically meaningful entities (and requirements on how they interrelate) plus the ability to analyze the returned result sets, enables some interesting applications. Some open problems will also be discussed.

11.20-
11.50 Semantics for Personalized Search - An Example
Per Gunnar Auran, Yahoo!

Abstract: This presentation will introduce and example on how simple semantic relations can be used for a personalized search experience. The speech will discuss a prototype search application that was developed by Yahoo! Technologies Norway in 2004-2005 by a team lead by the author.

Short CV: Per Gunnar Auran is a Senior Research Scientist at Yahoo! Technologies Norway AS, where his main responsibility is search relevancy for Yahoo!'s vertical search platform, Vespa. He is the technical lead for research releated to document analysis and ranking, query analysis and semantics, personalized and community search for Vespa.
From April 2000 to April 2003, Per Gunnar Auran was the R&D Manager of the Data Analysis Group, at Fast Search & Transfer ASA (FAST), focusing on web search data analysis and search relevancy for AllTheWeb, a leading web search engine of the time. Prior to that he was a research Scientist at the Norwegian Paper and Pulp Research Institute. He holds MSc and PhD degrees from the Norwegian Institute of Technology, specializing in engineering cybernetics.

12.00-
13.00 Lunch

SEMINAR
SEMANTICS & SEARCH
April 28th, 2006
Location: **ITV-454**

13.00-13.20 Web Spam and the Search Engines
Per Holager, NTNU

Abstract: Web spam is material put on the WWW mainly to promote client sites towards the top of the results listings of the search engines. The amount of such spam is so great that it is a major problem: One study showed that about half sites represented in one search engine data set were spam. This threatens the quality of results, which should interest anyone using these services. This presentation discusses how web spam works and describes how spam styles have developed through the last 15 years. It also presents some counter-measures that the engines may apply. One surprising approach is commercial deals where the search engine companies supply the spam contents.

13.20-13.40 Document space adapted ontology. Application in IR
Stein L. Tomassen, NTNU

Abstract: Retrieval of correct and precise information at the right time is essential in knowledge intensive tasks requiring quick decision-making. In this talk, a method for utilizing ontologies to enhance the quality of information retrieval (IR) by query enrichment will be discussed. The focus is tuning a retrieval system by adapting ontologies to provide both an in-depth understanding of the user's needs as well as an easy integration with standard vector-space retrieval systems. The ontology concepts are adapted to the domain terminology by computing a feature vector for each concept. Then, the feature vector is used to enrich a provided query.

13.40-14.00 Folksonomies and Ontologies
Csaba Veres, NTNU

Abstract: Two of the most exciting themes to unite activity for transforming the Internet are the Semantic Web, and Web2.0. While they differ in the formality of their inception and the scope of the problems they aim to solve, they are linked by their common reliance on metadata. The formal ontologies of the Semantic Web and the wild, emergent folksonomies of Web2.0 have been pitted against one another as competing means for managing the vast array of data on the Internet. Advocates of folksonomy are particularly vocal in claiming victory in this battle, pointing to the rapid adoption and demonstrable popularity of Web2.0 services as opposed to the lack of widely deployed Semantic Web applications. My perspective is that the link between the two should be exploited to benefit both. My claim is that emergent folksonomies can be best understood as products of human cognitive processes, and their analysis should focus on the way our mental architecture performs classification. I present a technique that can be used to expose the latent structure behind folksonomies, which goes a long way towards creating useful ontologies. This is useful for the Semantic Web because it is a rich source of cheap, relevant, structured data to enable Semantic Web applications, and it is useful for Web2.0 in enabling interoperability and rich search facilities across Web2.0 applications

14.00 Coffee Break (30 min)

14.30-14.50 AlViz ontology alignment visualization tool
Jennifer Sampson, NTNU

Abstract: One of the main reasons we need to align ontologies is to share a common understanding of the structure of information among people or software agents. Ontology alignment is the process where for each entity in one ontology we try to find a corresponding entity in the second ontology with the same or the closest meaning. While full automation is the ultimate goal, not everything can be done by machine, user interaction is still essential in order to control, approve and optimize the alignment results. We are developing a theoretical framework for understanding ontology alignment quality and through this work we propose a number of tools and techniques for achieving quality alignment results. I will briefly describe one of these tools, AlViz, our new visual ontology alignment tool for facilitating user understanding of the alignment results.

14.50-15.10 Performance Semantics in a Continuously Changing Enterprise
Jon Espen Ingvaldsen, NTNU

Abstract: Every thing changes. Continuously. Markets change. Products change. Your competitors and customers change. Laws and regulations change. Available technologies change. The continuous changing environment requires that process-aware information systems are very flexible with respect to coordination of applications, integration to business partners, and performance monitoring.
This presentation describes status and future directions for research work that aims at providing dynamic performance information through a simple natural langue query interface.

15.10-15.30 Ontology Value in Information Management
Darijus Strasunskas, NTNU

Abstract: Ontology is applied in wide range of application areas, e.g., semantic interoperability, view alignment, etc. Therefore, the quality of ontologies is a delicate topic, e.g. an appropriate level of granularity is application specific. In this presentation I discuss the application of ontology to information management in general and analyze ontology quality facets essential for improvement of information retrieval in particular.

15.30-16.00 Summarizing discussion & closing of seminar
Jon Atle Gulla, NTNU