Editors' Introduction to the Special Issue "Learning in Web Search"

Introduction

The emerging world of search we see is one which makes increasing use of information extraction, gradually blends in semantic web technology and peer to peer systems, and employs grid style computing resources for information extraction and learning. This Informatica special issue explores the theory and application of machine learning in this context for the internet, intranets, the emerging semantic web and peer to peer search.

Search can also be viewed as a knowledge sharing service on the Web, as an interface to the Semantic Web. While some automation in building the Semantic Web has been achieved, it remains a labour intensive annotation process with problems in scaling up to the full free-text Web. A partial implementation of semantic-based search is possible where hierarchical concept spaces rather than full ontologies are used, and where information extraction and learning tools in the search engine perform approximate tagging of concepts. This partial semantic-based search could be viewed as a key infrastructure for more complete Semantic Web development, and arguably, as a safety net for it.

Overview of the issue

The articles in this special issue originate from two different backgrounds. Two articles are from the EU IST project ALVIS[1], which aims to bring web search infrastructure closer to the vision of the semantic web by automating some of the labor intensive annotation processes. The article *Semantic Search in Tabular Structures* by *Aleksander Pivk, Matjaz Gams and Mitja Lustrek* explores techniques for making tables and their content the subject of search while also considering the implicit semantics in the tabular structure. In the contribution *Beyond term indexing: A P2P framework for Web information retrieval*, the authors *Ivana Podnar, Martin Rajman, Toan Luu, Fabius Klemm and Karl Aberer* present a new framework for full-text information retrieval in P2P overlay networks and introduce a novel retrieval model based on highly discriminative keys.

Two further papers are selected papers from the workshop "Learning in Web Search" held at the International Conference on Machine Learning (ICML) in 2006 organized by the editors of this issue. The article *A Semantic Kernel to classify Texts with very few Training Examples* by *Roberto Basili, Marco Cammisa and Alessandro Moschitti* contributes to the field of using semantic background knowledge in the context of kernel methods. The article *Sailing the Web with Captain Nemo: a Personalized Metasearch Engine* by *Stefanos Souldatos, Theodore Dalamagas and Timos Sellis* presents the implementation of a metasearch engine that exploits personal user search spaces.

Stephan Bloehdorn, Wray Buntine and Andreas Hotho

---

[1] http://www.alvis.info/