A Workshop Report: Mining for and from the Semantic Web at KDD 2004

Andreas Hotho
KDE Group, University of Kassel
hotho@cs.uni-kassel.de

York Sure
Insitute AIFB, University of Karlsruhe
sure@aifb.uni-karlsruhe.de

Lise Getoor
Computer Science Dept/UMIACS at University of Maryland
getoor@cs.umd.edu

ABSTRACT

The international workshop on Mining for and from the Semantic Web (MSW) at the KDD 2004 successfully brought together people from the communities Semantic Web and Knowledge Discovery. The goal of the workshop was to strengthen the communication and interaction between these communities as there is currently an agreement that both can benefit largely from each other. Overall the contributions and discussions showed that today mining *for* the semantic web is the primarily targeted research area. However, with a growing amount of available semantic web data also mining *from* the semantic web will be more prominent.

1. INTRODUCTION

The international workshop on *Mining for and from the Semantic Web* was located at the 10th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2004) on Sunday, the 22nd August 2004, in Seattle, WA, USA.

The workshop successfully brought together researchers from the research areas Semantic Web and Knowledge Discovery. According to T. Berners-Lee the Semantic Web is "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation". Current standardization efforts include e.g. the W3C recommendation for the Web Ontology Language[1] (OWL). Knowledge Discovery is defined by U.M. Fayyad as "the nontrivial process of identifying valid, previously unknown, potentially useful patterns in data". The combination of these two fast developing research areas is often called semantic web mining and addresses research questions trying to combine methods and technologies from both research areas.

The workshop followed a series of related workshops on *Ontology Learning* at ECAI 2000 and *Semantic Web Mining* at ECML/PKDD 2001 and 2002. Its call-for-papers generated contributions from North America, Europe and Asia. The nine accepted papers which were accompanied by an invited talk covered a wide range of issues falling into categories such as Text Mining and Semantics, Mining Structures and Tools, Frameworks and Applications.

[1]see `http://www.w3.org/TR/owl-features/`

2. TEXT MINING AND SEMANTICS

Currently there is a lot of textual data available on the Web and in intranets which is frequently being used to extract concepts and relations based on natural language processing (NLP) techniques. The two talks *The Terascale Challenge* and *Large-Scale Extraction of Fine-Grained Semantic Relations between Verbs* given by Timothy Chklovski presented results in this direction. Another approach for using textual data was presented in the talk *Sentiment Extraction from Unstructured Text using Tabu Search-Enhanced Markov Blanket* given by Edoardo Airoldi. Such sentiments can e.g. be used to derive customer preferences.

3. MINING STRUCTURES

We thank especially Alon Y. Halevy for stimulating discussions during the workshop and afterwards by giving his brilliant invited talk on *Mining Structures to Predict Semantics*. He identified the reconciliation of the semantics of disparate information sources as a key challenge in building the Semantic Web. Alon presented an approach to the reconciliation problem that is based on mining large corpora of schemas and semantic mappings and illustrated several examples where this approach has been successfully applied.

With a growing amount of semantically represented information the analysis of the structural properties comes into interest. Shawn R. Wolfe talked about *Exploiting Recurring Structure in a Semantic Network* and sketched ideas to use recurring structures in semantic networks.

4. TOOLS, FRAMEWORKS AND APPLICATIONS

SEMEX is a tool for management of personal information and was presented in the talk *SEMEX: Mining for Personal Information Integration* given by Xin Dong. An approach for integrating different methods from the inductive logic programming community to mine semantic web data was presented by York Sure in the talk *A Knowledge Discovery Workbench for the Semantic Web*. Last, but not least, the framework presented in the talk *A Framework for Image Annotation Using Semantic Web* given by Latifur Khan headed towards analyzing image data supported by ontologies.

5. JOINT SESSION WITH WEBKDD

Three of the papers were accepted and presented jointly with the WebKDD workshop.

The idea of using formally represented background knowledge in a text classification task was presented in *Boosting for Text Classification with Semantic Features* given by Andreas Hotho. Different linguistic ontologies were used as basis for an empirical evaluation with well-known corpora.
The two talks *Integrating Web Conceptual Modeling and Web Usage Mining* by Rosa Meo and *Discovering Links between Lexical and Surface Features in Questions and Answers* by Soumen Chakrabarti were mainly contributions to the WebKDD workshop. Rosa Meo presented an approach to use a conceptual log based on WebML which allows for easy analysis of usage behavior. Soumen Chakrabati presented a data driven machine learning approach to retrieve short passages from html pages for giving answers to given questions. The approach relied on linguistic methods to improve the quality of the model.

6. CONCLUSION

Overall the contribution and discussions showed that today mining *for* the semantic web is the primarily targeted research area. However, with a growing amount of available semantic web data also mining *from* the semantic web, e.g. by using approaches such as ILP and relational data mining, will be more prominent.
We thank all presenters and participants for the stimulating environment and interesting discussions. We hope that all benefited and enjoyed the workshop as much as we did and we are looking forward to continuing the exciting work in the shared research space of the communities on Semantic Web and Knowledge Discovery.
More detailed information about the workshop topics and the accepted papers are available at the workshop Web site (see `http://km.aifb.uni-karlsruhe.de/ws/msw2004`).