Semantic Web Mining for Building Information Portals

Jens Hartmann, Andreas Hotho and Gerd Stumme
Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany
{jhart, stumme, hotho}@aifb.uni-karlsruhe.de
http://www.aifb.uni-karlsruhe.de/WBS

Abstract

Semantic Web Mining combines two fast-developing research areas: Semantic Web and Web Mining. In this connection, the intention is to improve, on the one hand, Web Mining methods by new formal semantics represented in the Web and to utilize Web Mining results, on the other hand, for building up the Semantic Web. We also present the SemI-Port project, in which methods and tools for semantic information portals will be developed.

1 Semantic Web Mining

The Semantic Web is based on a vision of Tim Berners-Lee. He suggests to enrich the web by machine processable information which is organized on different levels.[1] The aim of the Semantic Web is to annotate web resources with semantic information. As a result, web resources can be accessed and processed by a knowledge driven mechanism whereby information is accessible both for humans and for web services.

As the Semantic Web enhances the first generation of the WWW with formal semantics, it offers new chances and challenges for Web Mining. Contents of web documents are represented in a formal way and links of documents are semantically annotated, which enables a contextual structure analysis and search for related documents.

The intention of Semantic Web Mining [2] is to improve the results of Web Mining by exploiting the new semantic structures in the web. As in traditional Web Mining, one can distinguish between content, structure, and usage mining. However, in the Semantic Web, content and structure are strongly interwoven, hence a distinction between Semantic Web Content Mining and Semantic Web Structure Mining blurs.

For mining the Semantic Web, traditional content and structure mining techniques (e. g., text mining; PageRank, Hubs & Authorities) can be extended by using Relational Data Mining techniques based on Inductive Logic Programming, using a formal knowledge representation and inference mechanism [6]. However, the scalability of such methods has to be improved, considering the excepted growth of the Semantic Web.

Usage Mining can also be enhanced further, for instance for discovering web usage patterns on a semantic level based on ontologies. A system storing such semantic log files has been developed at the AIFB [5].

As Semantic Web can improve Web Mining, Web Mining can on the other hand also be utilized to help building the Semantic Web. The aim is to develop methods for extracting semantics from the Web and to use them to build up the Semantic Web. Ontology Learning concerns the semi-automatic extraction of semantics [4, 3] by machine learning techniques for supporting the process of Ontology Engineering. Instance Learning techniques can then be applied to fill the ontologies. By using the resulting ontologies for mining the Web again, the loop can be closed.

Research activities as for instance the Semantic Web Mining Workshops [7, 1] at the conferences ECML/PKDD, co-organized by the AIFB, reflect the growing interest and point out further activities in this emerging area.

---

[1] see http://www.w3.org/DesignIssues/Semantic.html

2 SemIPort

The intention of the project *Semantic Methods and Tools for Information Portals (SemIPort)* is to evolve a set of methods and tools for representing and accessing information within a semantically structured information portal, while offering the possibility to integrate own information.

The participants of the research project are the Institute for Applied Informatics and Formal Description Methods (AIFB) at the University of Karlsruhe (coordinator), the German Research Center for Artificial Intelligence (DFKI), the Fraunhofer Institute for Integrated Publication and Information Systems (IPSI), and the Data Bases and Information Systems Group at the University of Trier. The project is funded by the German Ministry for Education and Research bmb+f.

For testing purposes, the developed approaches will be evaluated on data from the online bibliography DBLP, and the tools are planed to be integrated into the competency and service network portal of the German Informatics Society (GI) which is currently under construction.

The research tasks within this project are:
(1) Building an Ontology for the desired domain.
(2) Development of Scalable Storing, Processing and Querying methods for the Knowledge Management System. (3) Enhancement of Web Mining approaches to use within a semantic based framework. (4) Visualization and Browsing of Complex Data Inventories. (5) Personalization and Agent-based Interaction.

The third task covers two Semantic Web Mining topics as separate work packages: Semantic Web Usage Mining and Semantic Web Content/Structure Mining. In the presentation, we will discuss these topics in more detail.

3 Conclusion

We presented Semantic Web Mining as intersection of two fast-developing research areas Semantic Web and Web Mining. We discussed how the Semantic Web can improve Web Mining approaches and how Web Mining can be used to build the Semantic Web by using additional semantic information. In the presentation, we will include a more detailed overview about Semantic Web Mining and will give

a report of the SemIPort project.

References

[1] B. Berendt, A. Hotho, and G. Stumme (Eds.). *Semantic Web Mining*. Proc. of the Semantic Web Mining Workshop of the 13th Europ. Conf. on Machine Learning (ECML'02) / 6th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'02), Helsinki, August 2002.

[2] Bettina Berendt, Andreas Hotho, and Gerd Stumme. Towards semantic web mining. In I. Horrocks and J. A. Hendler, editors, *ISWC 2002, First International Semantic Web Conference*, pages 264–278, 2002.

[3] A. Maedche. *Ontology Learning for the Semantic Web*. Kluwer, 2002.

[4] A. Maedche and S. Staab. Discovering conceptual relations from text. In *ECAI-2000 - European Conference on Artificial Intelligence. Proceedings of the 13th European Conference on Artificial Intelligence*, pages 321–325. IOS Press, Amsterdam, 2000.

[5] D. Oberle. *Semantic Community Web Portals - Personalization, Studienarbeit*. Universität Karlsruhe, 2000.

[6] Heiner Stuckenschmidt, Jens Hartmann, and Frank van Harmelen. Learning structural classification rules for web-page categorization. In S. Haller and G. Simmons, editors, *Proceedings of FLAIRS 2002, special track on Semantic Web*, pages 440–444, May 2002.

[7] G. Stumme, A. Hotho, and B. Berendt (Eds.). *Semantic Web Mining*. Proc. of the Semantic Web Mining Workshop of the 12th Europ. Conf. on Machine Learning (ECML'01) / 5th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases(PKDD'01), Freiburg, September 2001.