Semantic Web Mining for Building Information Portals

Jens Hartmann, Andreas Hotho and Gerd Stumme
Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany

Abstract

Semantic Web Mining combines two fast-deve-
loping research areas: Semantic Web and Web Mining.
In this connection, the intention is to improve,
on the one hand, Web Mining methods by new for-
mal 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 sem-
antic 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 ma-
icine processable information which is organized on
different levels. The aim of the Semantic Web is to
annotate web resources with semantic information.
As a result, web resources can be accessed and pro-
cessed by a knowledge driven mechanism whereby
information is accessible both for humans and for
web services.

As the Semantic Web enhances the first gener-
ation of the WWW with formal semantics, it offers
new chances and challenges for Web Mining. Con-
tents of web documents are represented in a formal
way and links of documents are semantically anno-
tated, 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 tradi-
tional 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 con-
tent and structure mining techniques (e.g., text
mining, PageRank, Hubs & Authorities) can be ex-
tended by using Relational Data Mining techniques
based on Inductive Logic Programming, using a for-
mal knowledge representation and inference mech-
nism [6]. However, the scalability of such meth-
ods has to be improved, considering the expected
growth of the Semantic Web.

Usage Mining can also be enhanced further, for
instance for discovering web usage patterns on a se-
monic 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 ex-
traction of semantics [4, 3] by machine learning
techniques for supporting the process of Ontol-
ogy 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.

1see 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 planned 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. **(1) Building an Ontology for the desired domain.**
2. **(2) Development of Scalable Storing, Processing and Querying methods for the Knowledge Management System.**
3. **(3) Enhancement of Web Mining approaches to use within a semantic based framework.**
4. **(4) Visualization and Browsing of Complex Data Inventories.**
5. **(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


