

Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey

Petar Ristoski, Heiko Paulheim
Data and Web Science Group,
University of Mannheim, B6, 26,
68159 Mannheim

Article 01

Semantic Web Mining State of the Art and Future Directions

Gerd Stumme, Andreas Hotho,
Bettina Berendt

Article 02

WEB SEMANTICS

89% SCOPUS

Index

Summary of each article

What is common between the two articles

the difference between the two articles

Purpose of the articles

How each article proposed and exposed its subject

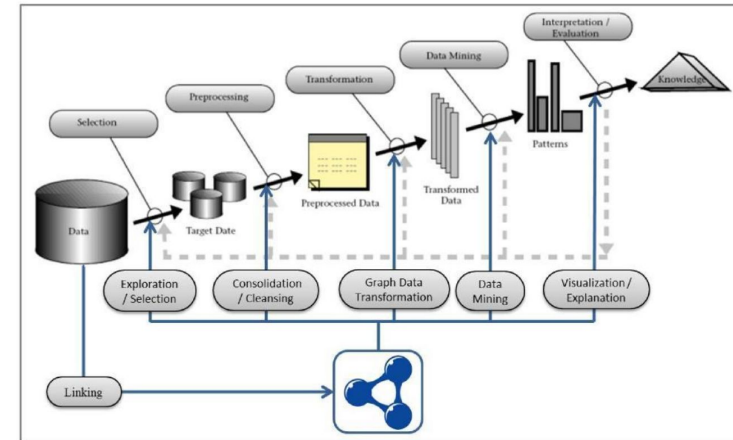
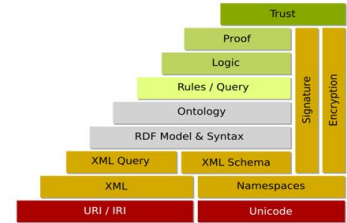
What each article concluded

Conclusion

Abstract: article 01

Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey

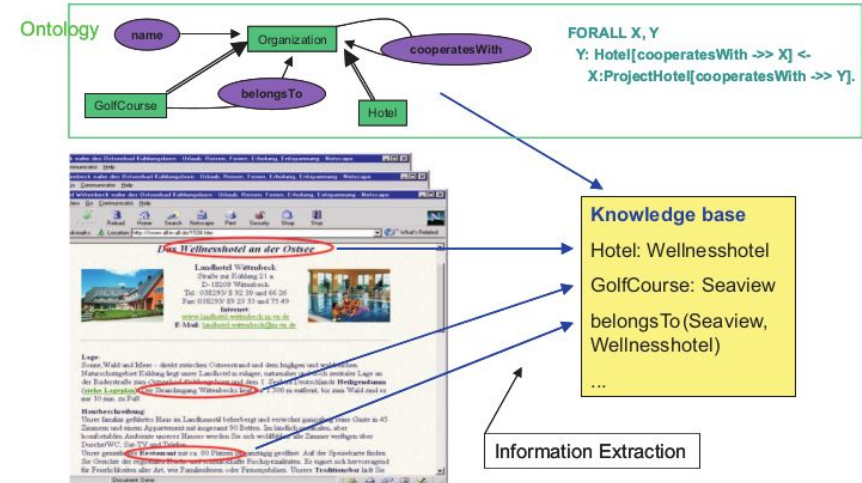
Data mining and knowledge discovery in databases (KDD) is a field of research concerned with obtaining top-level information from data. Tasks performed in this field are knowledge intensive and can generally benefit from using additional knowledge from various sources. Therefore, many approaches have been proposed in this area that combine data from the Semantic Web with the process of data mining and knowledge discovery. This research article provides a comprehensive overview of these approaches at different stages of the knowledge discovery process. As an example, it shows how Linked Open Data can be used in various stages to create content-based recommendation systems. The research shows that, while there are numerous interesting research papers done, the full potential of the Semantic Web and Linked Open Data for data mining and KDD is yet to be discovered.



Abstract: article 02

Semantic Web Mining State of the Art and Future Directions

The second article explains the combination of what Semantic Web Mining is, that is, it shows the combination of two areas of rapid development research, SemanticWeb and Web Mining. The article analyzes the convergence of trends in both areas: due to the growing number of researchers who are working to improve the results of Web Mining, exploring semantic structures on the Web, and they use Web Mining techniques to build the Semantic Web. Finally, it explains how these techniques can be used to mine the Semantic Web itself.



What is common between the two articles?

The two articles seek the semantic web to solve a problem in different areas, in other words, in the first article it uses the semantic web to mine databases using the KDD technique in order to obtain knowledge about relational databases where they are often only data in columns but without any meaning between them, but the semantic web has been used to remove useful information for data analysts without necessarily knowing what the database and its information is about.

In the second article, it uses the semantic web to obtain information about the data displayed on web pages and to contextualize that data in a mechanized way.





What's different between the two articles?

The articles differ in the way they present in which one demonstrates examples of semantic web implementation to mine information on the web, while the other shows the step by step to create the KDD and using the semantic web to obtain information about relational bank data of which the content and its applications are not known.





Purpose of the articles


Article 01

The proposal is to show with an example how to apply the semantic web in the data mining process and what the use of it can do to improve the understanding of the results at the end of the mining execution.

Article 02

The purpose of this article is to show studies on the combination of the semantic web and web mining using examples from past research and then provides references for future approaches for people interested in these areas.


How each article proposed and exposed its subject?



In this article, it provides a structured search for data mining approaches. Following the well-known data mining process model proposed by Fayyad, it is also discussed how semantic data is occupied in the different stages of the data mining model. In addition, it analyzes how different characteristics of linked open data, such as the presence of interlinks between data sets and the use of ontologies as schemas for the data, are explored by different approaches.

The rest of this article is structured as follows. Section 2 defines the scope of this research and places it in the context of other research in similar areas. Section 3 describes the knowledge discovery process according to Fayyad et al. In section 4, it presents a general model for data mining using Linked Open Data, followed by a description of approaches using Semantic Web data at the different stages of the knowledge discovery process in sections 5 to 9. In section 10, it shows an example of a LOD-enabled KDD process use case in the domain directory of recommendation systems. It concludes with a summary of the findings and identifies a number of promising directions for future research.

How each article proposed and exposed its subject



The second article uses the first section to introduce the subject and in the next II and III sections it presents brief overviews of the areas of Semantic Web and Web Mining. It then describes how these two areas cooperate today and how that cooperation can be further improved.

And how Web mining techniques can be applied to help create the Semantic Web. Section IV talks about researching approaches to learn ontologies and / or instances and how it can be used to improve the process and results of Web Mining.

Section V discusses these various techniques. In Section VI, he outlines how the loop can be closed: from Web Mining to Semantics. I concluded, in Section VII, that a strong integration of these aspects will greatly increase the comprehensibility of the Web for machines and thus become the basis for future generations of intelligent Web tools. It also shows the two notions of “semantics” and shows a delineate between the strengths, weaknesses and complementary.




What each article concluded

In the first article, he presents research on the use of data from the Semantic Web, for data mining and knowledge discovery. Following Fayyad's classic workflow pipeline, showing examples for using web data


Semantics at all stages of the pipeline, as well as full pipeline support approaches.

And when analyzing the research results, the first observation is that there is a lot of research work in the area and there are applications in many domains. A domain of frequent application is biomedicine and life sciences, but approaches are also transferred to several other domains. In addition, there are some sophisticated applications and tool stacks that go beyond mere research prototypes.



In addition, it shows that there are still some unknown territories in the research scenario of data mining enabled for the Semantic Web. This shows that, although impressive results can already be achieved today, all the potential that the Semantic Web enables in data mining and with KDD can still be discovered.

In the second article, the combination of two areas of rapid development research, Semantic Web and Web Mining, was studied. And it was discussed how Semantic Web Mining can improve the results of Web Mining, exploring the new semantic structures on the Web; and how the construction of the Semantic Web can make use of Web Mining techniques. And it provides an example that shows the possible benefits of further research in this attempt to integrate. And with this deeper investigation of that interaction it has given rise to new research questions and that will stimulate more research on the Web. With an important focus which is to allow search engines and other programs to better understand the content of pages and websites.



And with That reflected in the wealth of research efforts that shape the pages in terms of an ontology of the content, the objects described on these pages. And according to the article they hope that in the future, Web mining methods will increasingly address content, structure and usage in an integrated manner in iterated extraction cycles and using semantics, in order to understand and reshape the Web. iterated cycles, he hopes to see a productive complementarity between those who depend on semantics in the sense of the Semantic Web and those who rely on a more vague notion of semantics.



Conclusion

My conclusion is that each article proposed a different method to expose its research in the first one, a method demonstrating something more applicable and conclusive to the development of a practical project was used, in contrast in the second article it was more a demonstration and explanation on the subject with the purpose of showing research in the area to attract new researchers on the subject and show that there is still much to be researched.

Thank you!

