QQML 2009
International Conference on
Qualitative and Quantitative Methods in Libraries
26 - 29 of May 2009 Chania Crete Greece   www.isast.org

Libraries in the Semantic Web Era.

Marek Kopel[1] and Aleksander Zgrzywa[2]

[1]Wrocław University of Technology, Poland, [2]Wrocław University of Technology, Poland

marek.kopel@pwr.edu.pl , Aleksander.Zgrzywa@pwr.wroc.pl

The upcoming era of Semantic Web (SW) domination gives the promise for autonomous machine reasoning. But before the promise becomes reality there are a lot of efforts to make public Web data semantically enriched. The semantic enrichment is believed to allow the Web be readable not only by humans, but also by the machines. As Tim Berners-Lee - the author of WWW idea - envisioned, the existence of SW would allow the existence of intelligent agents that can process information from the Web autonomously. Autonomously means that unlike today Web browsers these agents would only need a high level of abstraction commands. For example, getting a command "Find a drugstore that is closest to my way home" the agent would estimate optimal way home based on current GPS read, match it to search results for local drugstores and check whether they are open at the time of coming home. It may optionally create an event in user personal calendar with a reminder set after the scheduled work time. That automatic reasoning needs inferencing based on Web data. To make the Web machine readable the data should be described by easily accessible semantic metadata.

As far as libraries are concerned, metadata have always been a part of cataloguing. Today in digital libraries it is common to represent metadata using standard ontologies, such as Dublic Core (DC). There is also the idea of exposing the metadata for harvesting in order to create one central database, which would allow to search all the available metadata at once. The protocol for that purpose is Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

The OAI-PMH is RESTful, which means it fulfils Representational State Transfer (REST) principles. But even with that architecture, the idea of harvesting the metadata duplicates and a central storage is in a way with the SW idea. For the purpose of interoperability of SW services Berners-Lee created four rules of linking data on the Web. The main idea of Linked Data is to make each piece of information identified by dereferenceable URI and available via HTTP. When fulfilling the rules data can be easily accessed an processed by a query language. The query language used widely for this purpose is SPARQL.

This paper would focus on problems with exposing the digital libraries supporting OAI-PMH to the Semantic Web services. The most wanted thing for OAI-PMH metadata is being querable via SPARQL. This means the metadata in OAI-PMH formats must be converted and served as RDF. Another important shortcoming of OAI-PMH is grouping and linking the metadata for the purposes of more complex processing and inferencing. Some of those issues are addressed in Open Archives Initiative Object Reuse and Exchange (OAI-ORE), but the integration with SW is still a work in progress.