Semantic Classification Search

Contextualizing the Dewey Decimal Classification using Freebase

by Hugo C. Huurdeman, h.c.huurdeman@uva.nl
Oslo University College, Norway / University of Amsterdam, the Netherlands

Extended abstract

Introduction

This work addresses a research aspect of linked open vocabularies related to the question on how repositories of Open Data can interact and be applied to traditional knowledge organization systems such as the Dewey Classification System. It also presents a solution to address this challenge, by describing Semantic Classification Search (SCS)[1], a web application that combines the Linked Data of the Dewey Decimal Classification (DDC)[2], available via Dewey.info and Freebase[3], a semantic collaborative Knowledge Base based on structured open data.

The paper contains a short literature review on the Semantic Web, Linked Data, the DDC and Freebase. Furthermore, it discusses the technology and development process of Semantic Classification Search, and the issues that occurred during the development, related to the underlying structure of the DDC and Freebase. The article concludes by describing the implications of the Semantic Classification Search application, and offers suggestions for further research.

Summary

Semantic Classification Search is an extensible application that uses Freebase, a semantic online encyclopedia based on the Web 2.0 paradigm, to provide context to the search and browsing process in Dewey's Decimal Classification. The prototype of this application has been developed in a relatively short timeframe, aiming to show that it is feasible to take advantage of the newly released Linked Data of the DDC.

The development of Semantic Classification Search has been done using a modular approach, and the resulting application contains several modules, like a Dewey browsing and retrieval module, and a Freebase retrieval module. Different methods are used to optimize the retrieval of results in Freebase. Furthermore, the Dewey and Freebase system are connected using a basic mapping, that maps broad Dewey classes to Freebase domains (categories of information).

To achieve an optimal performance of Semantic Classification Search, there are still some issues that have to be solved, specifically regarding the structure, naming and availability of data. In addition to that, actual user testing has to be done to evaluate and improve the search and browsing interface of SCS. However, the application has shown the potential of connecting Dewey to collaborative (Web 2.0) data repositories like Freebase. Semantic Classification Search also shows that, using relatively simple technical means, the Dewey Decimal Classification can make its first steps into the era of the semantic web.

---

[1] http://www.timelessfuture.com/semanticsearch

[2] http://www.dewey.info

[3] http://www.freebase.com

Biography

Hugo C. Huurdeman is a PhD-student at the Faculty of Humanities of the University of Amsterdam, The Netherlands. He is involved in the WebART[4] (Web Archive Retrieval Tools) project, that aims to improve the value of Web archives for realistic research scenarios.

Hugo has a MSc-degree in Information Sciences (2007), and is expected to receive his Master-degree for the International Master in Digital Library Learning (DILL) in 2012. The topic of his Master's thesis for DILL is related to mappings between formal classification systems and community-driven ontologies.

Appendix: screenshot of Semantic Classification Search
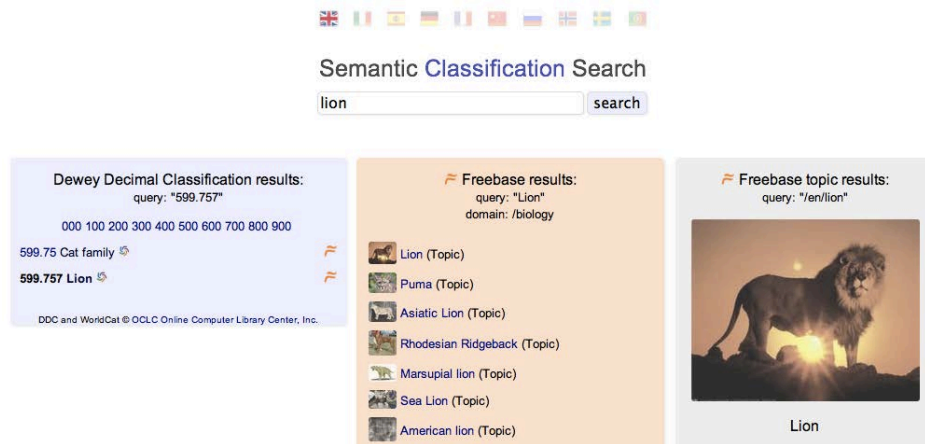

Figure 1: Screenshot of Semantic Classification Search prototype interface (query: Lion)