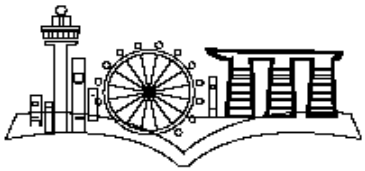

Submitted on: May 31, 2013

Building interfaces on a networked graph

Martin Malmsten
LIBRIS, Kungl. Biblioteket, Stockholm, Sweden.
martin.malmsten@kb.se

Niklas Lindström
LIBRIS, Kungl. Biblioteket, Stockholm, Sweden.
niklas.lindstrom@kb.se

Abstract:

The National Library of Sweden, The Swedish Cultural Heritage Board and the Swedish National Archive has during 2013 held a number of workshops with the purpose of both exposing Linked Data and creating interfaces built on top of multiple datasets from the participating organisations and others such as DBPedia.

Having the ability to work with data from multiple stakeholders does, however, create great expectations when it comes to creating interfaces with a cohesive user experience. Historically, doing such a thing would have involved aggregating the data to one single database, which goes against the underlying principles of the architecture of the web. However, Linked Data technologies have evolved enough to provide the means of creating such interfaces by directly interacting with the live datasets, be they local or remote.

The release of SPARQL 1.1 has proved a game changer in this area. Specifically, it is now possible to create federated queries across multiple datasets in a standardised and vendor independent way. In other words, SPARQL-enabled parts of the web of data can now be queried in real-time, furthering the vision of Linked Data, turning the web into a database.

On these workshops we have created a new authority view for Libris based on these technologies, drawing information from multiple sources to create a view that is more in the context of the user than in the context of the source material. This also allows for a more natural partition of responsibility regarding data. For example, influencedBy-relationships would be hard to maintain in the national catalogue, but are a natural part of Wikipedia. In the other hand, a complete bibliography and list of authors is the responsibility of the national library. Using SPARQL, a query such as "books by authors influenced by Strindberg" is easily answered providing the user with extended information that would otherwise have been impossible or costly to maintain by a single party.

This quite simple query can of course be solved by other means, but one of the points of SPARQL is

that it queries the data itself, you do not have to expose a specific API for each type of query. Queries can also easily become arbitrarily complex.

The obvious risk to this approach is inherent in the distribution of data and server infrastructure. However, since the approach mirrors that of the larger web itself, by being an actual part of it, these problems are well-known and solutions such as cacheing and fault-tolerance are ubiquitous today.

Keywords: Linked Data, SPARQL, Federated query