Semantic Navigator: Use of Semantic Data in Web Navigation

Jan Michelfeit and Tomáš Knap

Faculty of Mathematics and Physics,
Charles University in Prague, Czech Republic,
michelfeit.jan@gmail.com

Abstract. Semantic web search engines can take advantage of machine-understandable data published on the Web to provide more precise search results and advanced query capabilities. Semantic data embedded in web documents (serialized as RDFa or microformats, for example) can be used in conjunction with a semantic web search engine to provide a better web navigation experience. We present Semantic Navigator,[1] a Mozilla Firefox extension than brings advantages of semantic search to ordinary users. Semantic Navigator detects semantic data in web documents and with the aid of a semantic web search engine, it enables users to easily navigate to related documents containing information about a selected resource or about one of its properties.

Keywords: Semantic Web, Web navigation, Semantic search

1 Introduction

Traditional web search engines usually offer a keyword-based search. With increasing number of documents on the Web, it becomes more difficult to find the information we look for among many documents referring to similar search keywords. Problems associated with keyword-based search include keyword ambiguity or the fact that it may be difficult to search for a property of a resource. The same property is expressed in different words in different documents. However, search engines for the Semantic Web can cope with these problems better because they understand the meaning of what the user searches for.

In this paper, we introduce Semantic Navigator, a tool that utilizes structured data in web documents in order to bring advantages of semantic web search engines to ordinary users and provide an easy way to navigate between documents referring to similar or related concepts. Semantic Navigator is integrated in the Mozilla Firefox web browser as an extension so as to provide a quick access to search functionality and simple interface for users not familiar with technical details of the Semantic Web. Furthermore, other technologies of the Semantic Web Stack, such as ontologies, are utilized.

Semantic Navigator detects presence of RDF data or microformats in web pages and such data can be highlighted in the current web page. The user can

[1] The installation package and a screencast can be found at http://goo.gl/TxlXl

select a resource, a literal, a microformat, or a property of a resource and search for documents that contain more information about the selected element. The Sindice search service [1] is used to obtain search results. Semantic Navigator also provides access to information summaries on Sig.ma [2].

This approach brings advantages of semantic web search to users while they can adhere to the traditional web search paradigm – search for relevant documents and then find information in them manually. A tool facilitating access to a semantic web search engine is indeed necessary in this process. Resource URIs are typically too long and not user-friendly[2] and they are not directly visible to users. Semantic Navigator provides a user-friendly way how to extract URIs from a web page and search for related documents using these URIs.

2 Motivating Example

Suppose the user navigates to a web page about the film Casino. If the user wants to find more information about the film, hyperlinks leading to related web pages can be followed or a web search engine can be used. Both approaches have their limitations. Hyperlinks lead only to documents that were known to the author at the time of writing. Search engine results may not always be relevant because the word "casino" has other meanings than a film title.

If the web page contains RDF data, Semantic Navigator can extract the URI that identifies the film Casino and search for documents that refer to this film. The user may also be interested in a property of the film, such as the release year. If the web page contains information that a resource labeled "Casino" belongs to class "film" and this class has a property "release year" in the respective ontology, Semantic Navigator will let the user know about it and provide a list of pages containing information about the release year.

More examples of how this approach can be used are listed in [3].

3 Prototype description

While the user browses the Web, Semantic Navigator shows its presence only with an icon in the browser statusbar. When it detects semantic data in a document, it is indicated by a change of the icon. Semantic Navigator currently detects RDFa, microformats and external RDF documents linked to the (X)HTML document by a `<link>` tag.

The semantic data in the document can be accessed through the statusbar context menu or highlighted in the document. Semantic Navigator highlights elements that represent RDFa resources or literals and microformats (we currently support *hCard*, *geo*, *adr* and *hCalendar*). In addition to this, literals from the RDF model are annotated and highlighted in the text of the current web page. This feature is useful particularly when semantic data are stored separately in

[2] For example, the music band Nightwish is identified on MusicBrainz with URI `http://musicbrainz.org/mm-2.1/artist/00a9f935-ba93-4fc8-a33a-993abe9c936b`.

an external document, referenced by a `<link>` element, rather then embedded in the (X)HTML document.

A small icon appears next to highlighted elements. It can be clicked in order to display a context menu which contains items for search methods described below. In case of resources, the menu also contains all properties of the selected resource and their values can be listed. Classes the resource belongs to may also be listed and the user can browse properties that have the selected class as its domain.

Semantic Navigator provides the following search methods:

- *Resource search.* List of documents that contain the selected resource URI is presented to the user. The user can also choose to display information about the resource gathered from multiple sources using Sig.ma.
- *Property search.* Documents that contain a *resource-property-*$*$ triple for the selected resource and property are returned. In addition, the Sindice SPARQL endpoint can be queried for a list of possible values of the property.
- *Literal search.* Literals are searched with a keyword-based query. Either the list of documents from Sindice search or a Sig.ma summary can be displayed. The search can be refined by selecting a property and searching for the triple pattern $*$-*property-literal*.
- *Microformat search.* Microformats are similar to blank nodes in RDF as they do not have an identifying URI. Thus only chosen relevant values for each type of microformat are searched. For example, in case of the *geo* microformat, we can search for documents that contain the same values of `vcard:latitude` and `vcard:longitude` properties as the values in the microformat the user selected for search.

4 Implementation

Semantic Navigator is implemented as a Mozilla Firefox extension. The interface is written in JavaScript and XUL (XML User Interface Language), manipulation with RDF data is implemented in Java so that existing libraries can be used. The Any23 library[3] is used to parse RDF data and microformats. The data is then processed with the Jena framework[4], which also provides ontology reasoning capabilities. Sindice was chosen as the search engine because it is currently the only search engine that meets requirements of Semantic Navigator.

Semantic Navigator operates in several steps. The first step is detection of semantic data. RDFa and microformats are recognized in web pages; RDF/XML, Notation3 and Turtle are supported in documents linked with a `<link>` tag. The second step is extraction of RDF triples from data sources identified in the previous step. Third, used ontologies are imported and analyzed. Ontologies are fetched as described in [1] and cached locally. The optional fourth step is highlighting of RDFa data, microformats and highlighting of RDF literals in the

[3] `http://developers.any23.org/`
[4] `http://jena.sourceforge.net/`

text. In the last step, the Sindice service is queried and search results presented to the user.

5 Conclusion and Future Work

In the work at hand, we presented Semantic Navigator, a tool that enables users to search for documents with information about selected resources or about their properties. Unlike other Linked Data browsers, Semantic Navigator doesn't provide access to pure data but to whole documents. While few other specialized projects that use technologies of the Semantic Web for navigation in (web) documents exist (such as Semantic Browser [3] or Scooner[5]), the contribution of our tool is in its versatility. It is not bound to any specific dataset nor ontology.

The practical application of Semantic Navigator largely depends on three factors – relevance of search results returned by Sindice, availability and quality of web documents with embedded semantic data and identifier re-use. There are currently also some technical limitations[6] of Sindice search.

Semantic Navigator could provide a basis for interesting new applications. It could be adjusted to a specific dataset and serve as a ready easy-to-use interface for navigation and searching within the dataset. For example, it could provide the same functionality as Semantic Browser.

Another possible application is a navigation scheme similar to links between Wikipedia articles. Suppose we have a set of articles that are not interlinked. If we export the list of article names in the RDF/XML format and we link it to each article with a `<link>` element, the user can use Semantic Navigator to highlight terms in the text of an article and search for more information about the term. The search could be limited to a selected domain.

As a whole, even though the practical usability of Semantic Navigator is limited due to the low number of web documents with RDFa markup, the tool successfully demonstrates how advantages of Linked Data can be applied to the traditional web of human-readable documents.

Acknowledgement. This work was supported by the grant SVV-2011-263312.

References

1. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G.: Sindice.com: A Document-oriented Lookup Index for Open Linked Data. International Journal of Metadata, Semantics and Ontologies, 3(1), 37–52 (2008)
2. Tummarello, G., Cyganiak, R., Catasta, M., Danielczyk, S., Delbru, R., Decker, S.: Sig.ma: Live views on the Web of Data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web - Volume 8, Issue 4, November 2010, Pages 355–364
3. Gonen, B.: Semantic Browser. Master's thesis, Sakarya University (2002)

[5] `http://wiki.knoesis.org/index.php/Scooner`

[6] Unreliable RDFa detection (`http://goo.gl/uwLvU`) or limited query capabilities when searching for two properties of the same resource (`http://goo.gl/7OCKm`).