Web Science and Engineering 2015

Homework 4

Eric Camellini
4494164
e.camellini@student.tudelft.nl

1. HOMEWORK

1) What is Linked Open Data, what are its main principles, and how is it used?

To define Linked open data it is necessary to mention the concept of Liked Data. Linked data is connecting and interlinking the structured semantic information on the Web in order to enable the exploration of the Web of Data through links, like it is possible to do now with the classic Hypertext. The term was coined by Tim Berners-Lee that also defined the guidelines to publish data following the linked data method: use HTTP URIs (Uniform resource identifiers) to identify data, follow the standards for data description (e.g. RDF) and include links to other data. Linked open data (LOD) is linked data published as open content (i.e. with an open license)[3][4].

2) Explain what DBpedia is, and how it is used in web applications.

DBPedia[2] is a crowd-sourced project that has the objective to extract structured content from Wikipedia. This results in the creation of a huge semantic database that contains the semantically relevant information extracted from Wikipedia, and that can then be semantically queried. In other words it can be seen as a project that aims to create a machine readable version of Wikipedia. Tim Berners-Lee indicated it as a perfect example of Linked Data [8].
In web applications DBPedia can be used to link entities in the semantic data embedded into a web page to their already defined DBPedia entry. Doing in this way many different web pages can easily agree on the semantic meaning of a certain entity by linking it to a common DBPedia entry.

3) What are Microformats and RDFa, and how are they used?

RDFa (RDF in attributes)[6] and Microformats[5] are projects that allow to embed semantic metadata in HTML-like documents (e.g. HTML4 and 5, XHTML, XML and etc.). RDFa is a W3C extension of the set of available markup attributes that enables the usage of RDF directly within the tags of almost any markup language. Microformats has the same objective, but differs from RDFa in the way in which the semantic data is embeddable: it is not based on new attributes, but on already present ones like *class* (HTML, XHTML standard). The Microformats project defines predefined formats that can be followed in the markup language in order to express semantic information (e.g. the Geo Microformat is implemented using the values "geo", "latitude" and "longitude" as values for the class attributes). There are other similar projects that have the same objective, such as Microdata, RDFa Lite etc.

4) What are [within the context of the Semantic Web] vocabularies, give a few examples, and describe how those are used?

Vocabularies are schemas that should be followed while creating semantic information (e.g. while embedding semantic metadata to a Web page), so that is possible for many content creators to agree on the definition of certain entities/relationships (the term can also be seen as a synonym of "ontology")[9]. An Example of vocabulary is Schema.org[7], and a practical usage could be following http://schema.org/Person when adding semantic metadata about a Person in a Web page (e.g. using RDFa), that means using the set of attributes defined in the Person schema (e.g. name, birthPlace etc.).

5) What are folksonomies, and how are they used?

A folksonomy is the classification of user generated tags in a system, based on their statistical characteristics (e.g. frequency). In a folksonomy implementation the users are allowed to annotate their content using public tags and then to index, categorize or filter the content of all the other users using these tags (e.g. show all the items tagged with a certain keyword, show the resources that are annotated with the most used tag). Folksonomy implementations are also known as collaborative/social tagging systems.[1] The most famous example of folksonomy is the Twitter[1] hashtags system. Such a feature is useful in the vision of Semantic Web because user content can then be indexed, filtered and analyzed by machines, even if a tag doesn't have a semantic

[1] `https://twitter.com/`

meaning (it is simply a keyword).

6) Describe and explain at least one research challenge related to the Semantic Web.

I think that the most challenging part of the Semantic Web field at the moment is finding a set of well-defined, complete, standards (or to promote their usage, when they already exist) to eventually enable every developer to easily add semantic metadata to the created content and to easily enable the usage of the semantic data through, for example, standardized inference techniques, in a seamless and transparent way. This challenge includes the following ones:

- Convince all the content creators to add semantic information, and provide them tools to do it: they should be able to choose in an easy way which standars to use, which vocabulary/schema and where to link external entities (e.g. this could be achieved through the integration of vocabularies and linked data sources in the IDEs or frameworks used for web developement, with functions for suggesting/recommending semantic data linking and insertion to the programmer);
- Convince data providers to adhere to the Linked Data standards, and to publicly release this data when possible: this means, for example, that there should be data linking functionalities integrated in the tools used to manage data management, persistence and distribution.
- Define standard logical models and inference procedures that can be used to explore semantic metadata available on the Web and to derive new knowledge in an automated way;
- Provide libraries and frameworks that allow developers to use these logic and inference tools in a transparent way: to reason over data and infer new information without the need of knowing the underlying theoretical and logical concepts.

Many standards already exist, but the developers must be encouraged to use them and convinced of the power ad the importance of the Semantic Web revolution. This challenge of standardization and improvement in the usability of Semantic Web tools must be addressed because the success of all the widespread technology that revolutionized the world always came from the fact that there were well-defined standards and tools to easily use them. It happened with the Web (HTML was standard, easy, well-defined, and a first browser was provided), with the Smartphones (Easy-to-use SDKs to buil apps, two main OSs, possibility to distribute the software through a centralized infrastructure), and now the same thing must happen in the Semantic Web field if we want to see a real revolution in knowledge discovery and information retrieval.

References

[1] Collaborative tagging - w3c. `http://www.w3.org/2005/Incubator/mmsem/wiki/Tagging_Use_Case`. Accessed: October 17, 2015.

[2] Dbpedia. `http://wiki.dbpedia.org/`. Accessed: October 17, 2015.

[3] Linked data. `http://linkeddata.org/`. Accessed: October 17, 2015.

[4] Linked data design issue - tim berners-lee. `http://www.w3.org/DesignIssues/LinkedData.html`. Accessed: October 17, 2015.

[5] Microformats. `http://microformats.org/`. Accessed: October 17, 2015.

[6] Rdfa. `https://rdfa.info/`. Accessed: October 17, 2015.

[7] Schema.org. `http://schema.org/`. Accessed: October 17, 2015.

[8] Tim berners-lee: The next web - ted talk. `http://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=en`. Accessed: October 17, 2015.

[9] Vocabularies and ontologies - w3c. `http://www.w3.org/standards/semanticweb/ontology`. Accessed: October 17, 2015.