F20BD Coursework 2
Jamie McCulloch
H00189648

What is the objective of the Semantic Web?

The Semantic Web is a "variation of the Web that includes documents which describe explicit relationships between things" [1]. The aim is to have all documents on the web relate to each other, to allow users to discover more information with less effort. It is the theory that, if every document on the Web was linked, then humans and machine can work in unison with ease through giving all information on the Web well defined meaning. According to Tim Berners-Lee, who invented the concept of the World Wide Web, the Semantic Web would be a continuation of the Web we know today, not a new one. This brings up the matter of translating/transferring all the data already existing on the Web into a well defined, reachable format. The format used for linking data is Resource Description Framework (RDF).
The key property of the Semantic Web that it focusses on shared data. By defining what type of data something is, it is also relating it to other data of that 'type', creating a Web of semantics (hence the name). The Semantic Web is completely decentralized and so trust is a major component to its success. The users must know how credible the source of information is, and if they are suddenly all connected to each other, and one source is incorrect it can cause the whole collection of information to be false.

What are the key technologies for the Semantic Web? How do they relate?

RDF – RDF is a graph database that stores data in a triple store format, often with a *Subject, Predicate, Object* format. RDF uses Unique Resource Identifiers (URIs) to reference and identify resources. The most important factor of RDF document is that it can be linked to other RDF documents that it relates to. The Semantic Web uses data that is stored in RDF, or can be mapped to RDF, it is fundamental to the Semantic Web as we know it. The RDF model enables a link between two resources, which can be furthered strengthened with ontologies [2].

Ontology – An onotology is a file or document that sets out and defines relations among terms. The most common type of ontology is the use of a taxonomy and a set of inference rules. The taxonomy defines classes of objects, and all of the relations among them, while the inference rules supply further power. Through setting rules for program can work out certain 'facts' and manipulate terms more effectively for the user. The more rules – the more accurate the manipulations. Ontologies can also be used to solve merging conflicts when linking two resources. For example if in one resource the field is called firstName and in the second it is referred to by fName, then the Web must know that these should be linked and the same field.

OWL – Web Ontology Language (OWL) is used to when the application needs to know *about* the data, and not just process and display it [2]. OWL provides additional vocabularies and semantics which allow for more accurate judgement. Through this, it provides more accurate machine interpretations of web content against XML and RDF. OWL also offers a standard means of describing relationships between RDF, an example of this being through *rdfs:subPropertyOf*.

SPARQL- SPARQL is a declarative query language used to extract data from RDF knowledge bases[6]. The web contains hundreds of thousands of data documents but there is only 260 SPARQL endpoints [5]. A SPARQL endpoint allows execution of a query through HTTP, allowing it to take place on the server. Due to the highly individualised nature of these queries, the cachability of SPARQL is low – there will probably not be queries the same, so no there is no advantage to caching these. This may contribute to the lack of endpoints, however endpoints also tend to be unstable, and not comply with standards[5]. To mitigate this, endpoints may restrict usage, or query execution time.

What is the difference between Linked Data and the Semantic Web?

The main difference between Linked Data (LD) and Semantic Web (SW) is that the SW is the *goal*, that's what we are trying to achieve, LD is *how* we are going to link information together[4]. SW is an extension of LD, without LD there would be no SW. There are still other ways they differ however. The SW provides an environment where applications can query data and draw inferences from it, as detailed above. To make this a reality though, we need the data to be in a standard format, and it needs to be reachable and manageable by SW tools [3]. The SW needs data, but it also needs relationships within the data to create a 'Web'. These interrelated datasets are collectively known as 'Linked Data'.
LD follows four main principles [7], as conceptualised by Tim Berners-Lee:

(i) Use URIs as names for things
(ii) Use HTTP URIs to allow people to look up those names
(iii) When a URI is looked up, provide useful information using recommended standards (like RDF or SPARQL)
(iv) Include links to other URIs so more data can be discovered.

To ensure LD, all data needs to be available in RDF, or be given immediate access through RDF to existing databases. LD should also set up query endpoints (SPARQL) which would allow data access quicker and more efficiently.

Where are Semantic Technologies being employed, and what benefits do they provide for those use cases?

Semantic technologies are appearing more and more common within society. With web examples such as search engines using semantic technology to display more information to the user that is relevant to a search, but not exclusive to. For example – if a user searches a movie, a search engine will display ratings for the movie, the main actors, the director, the budget, related movies, etc. The search engine pulls all of this data from different sources, and merges it into one, allowing the user to see lots of information without leaving the search engine and navigating to another site. It is not just search engines on the web however, DBPedia has mapped the entire contents of every Wikipedia page to RDF format, allowing this to be linked and used in other ways [3].
More scientific data was stored and generated between 2009-2014 than in the entirety of mankind [2]. With this much scientific data, semantic technologies could be used to efficiently link and display lots of information from different databases and documents. It is reported that sequencing a human genome has dropped in price from $10,000 in 2007, to just $1000 in 2012 [2]. With the rise of Big Data, and the ability to sort through massive amounts of data quickly and extract what we need, the price will continue to drop. We can store this data in RDF, and link it together with other data, perhaps resulting in better healthcare in that it may be easier to diagnose illnesses based on patterns not already known due to the inability to trawl through petabytes worth of data.

Another avenue in which semantic technologies are used is within chemical engineering. The drug discovery business in now heavily reliant on management, curation and integration of massive amounts of potentially useful data. By using semantics the information can be interpreted and knowledge can be derived from this [8].

References

[1] SWAG - Semantic Web Agreement Group (2001, May 7). What is the Semantic Web [Online]. Available: http://swag.webns.net/whatIsSW

[2] Wu, Hongyan, and Atsuko Yamaguchi. "Semantic Web Technologies for the Big Data in Life Sciences." *Bioscience Trends* 8, no. 4 (2014): 192-201.

[3] W3 (2005). Linked Data [Online]. Available: https://www.w3.org/standards/semanticweb/data

[4] Vadim Eisenberg (2011, Oct. 29). On the difference between Linked Data and the Semantic Web [Online]. Available: http://vadimeisenberg.blogspot.co.uk/2011/10/on-difference-between-linked-data-and.html

[5] Rietveld, Laurens V., Wouter Beek, Stefan Schlobach, Ruben Verborgh, and Miel Vander Sande. "Linked Data-as-a-service: The Semantic Web Redeployed." *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9088 (2015): 471-87.

[6] - Hasnain, Ali, Qaiser Mehmood, Syeda Sana E Zainab, and Aidan Hogan. "SPORTAL: Profiling the Content of Public SPARQL Endpoints." 2016, Hasnain, A., Mehmood, Q., Sana E Zainab, S., & Hogan, A. (2016). SPORTAL: Profiling the Content of Public SPARQL Endpoints. International Journal on Semantic Web and Information Systems (IJSWIS), 12(3), 134-163. Doi:10.4018/IJSWIS.2016070105.

[7] Tim Berners-Lee (2006, July 27). Linked Data [Online]. Available https://www.w3.org/DesignIssues/LinkedData.html

[8] Borkum, M.I., and J.G. Frey. "Usage and Applications of Semantic Web Techniques and Technologies to Support Chemistry Research." Acta Veterinaria Scandinavica 6, no. 1 (2014): 18.