Metodología y Documentación Científica

Trabajo Final

Adam Westerski

1. **Research work title.**
   Usability study of application development and information aggregation (mashup) in the Linked Data and Semantic Web area.

2. **Introduction.**

Semantic Web

The explanation of Semantic Web and its fields of usage has evolved since the initial contribution of Tim Berners-Lee in 1998 [8] . Research has progressed a lot and people try to adjust the original concepts to the changing situation. The simplest W3C definition [2] states that *Semantic Web is a metadata based infrastructure for reasoning on the Web*. Although this is a short explanation it captures the key aspects of this topic.

Current web is created in an easy to understand way for humans. The Semantic Web initiative tries to deliver means for understanding the web content by computers. According to the W3C tutorial [2], Semantic Web is not about delivering any sophisticated artificial intelligence algorithms. Although AI might be connected to the described subject it is not what Semantic Web is about. The key idea is much less complex and ambitious then giving birth to intelligent machines. In Semantic Web computers are to reason about the resources content through the analysis of the metadata. Currently in the World Wide Web most information is stored only in a human readable form without any particular order. People use natural language for describing content and do it in many different languages and ways. The huge diversity of this metadata makes it very hard to analyze.

One of the goals of the Semantic Web on the metadata level is to create a universal description method to annotate resources. In order to accomplish this goal the ontologies are proposed. Their task is to define concepts and relationships that refer to a certain domain.

When the required descriptions are available then they can be used in a number of ways by certain program logic to reason about relations between annotated resources. This process is about choosing any two resources (possibly very complex) and being able to determine whether they are similar, parts of each other or perhaps connected in some other way. In fact this is the actual meaning of earlier discussed process of machines being able to understand web content. Such ability brings a vast new world of possibilities.

Web of data and Linked Data

The Web of Data also referred to as Web of Linked Data is the next step in the evolution of Semantic Web idea. It is not what the initial trends envisioned for Semantic Web to become, nevertheless Web of Data is where Semantic Web is at the moment.

There are many mature Semantic Web projects or even standards already published. All of those were supposed to build the foundations of the new Internet but they are not popular neither widely accepted. Some point out the overwhelming complexity as reason for such situation, while others remark small and unrealistic appliances of what has been released so far. The originally envisioned and hyped Artificial Intelligence for the Web has not arrived therefore a simpler solution is taken into account as the first step on the road to reaching Intelligent Internet. This first step is called Web of Data. One of the key concepts of this trend is to provide not only annotations for the Web but also means to interconnect the annotated resources. The outcome of initial Semantic Web activity is a situation where the metadata and annotations of web resources are already out in the Internet but data is still not interlinked and therefore hard to use in practice. Selected websites or particular groups of interest have their web resources annotated but outside of those communities the information is not usable or no one has any interest of using it. Therefore the Web of Data initiative is also about reaching out to the people -the Internet community. Not some selected, enclosed groups like before but more massive audience. The goal is to introduce

easy to use paradigms for annotation of the web content[3][4] and functional applications that take advantage of those annotations (semantic search engines[5][6][7][8], Semantic Web based portals like Twine[9]. By some Web of Data is also referred to as Web 3.0 because it harvests on the idea that was the corner stone of Web 2.0 success -social networking. A great number of the main stream activities within Web of Data put most emphasis on internet users, user activities and virtual societies in the modern Internet. There is a lot of discussion and research about annotations in collaborative environments that dominate the contemporary Internet environment. In the end the ultimate goal remains the same Web of Data, similarly as the aforementioned Semantic Web, is supposed to deliver more user friendly Internet through appliance of machine understandable content and mediation between human and machine.

Semanticly Interlinked Online Communities (SIOC)

In the context of this proposal and the appliance of Semantic Web technologies in social networks SIOC[10] is a very important initiative. It proposes annotations of community resources of the internet such as: blogs, boards, mailings lists etc. SIOC defines how and where the machine readable information about communities should be put. It specificity points out how to describe each of the aforementioned social networks. Furthermore it delivers means to interlink all of those sources of information. For instance with the help of SIOC the same user can be identified across many systems and linked to resources such as blog posts and mailing list posts at the same time. Additionally with the use of further discussed semantic search engines applications can extract all this data and present it in a human readable form(for example after login a user on the forum can see his/hers blog posts and mailing lists entries from other portals across the entire Internet). Currently SIOC has been published as a W3C submission[11].

Semantic Search Engines

The SIOC initiative provides the annotations but not the means to organize or utilize it. Currently SIOC is mostly used among some selected communities but it has occurred that the amount of the information is massive even if produced by such a little number of sources. In time it became clear that it is impossible to control such a stream of information with simple tools therefore in order to index, efficiently search and browse SIOC and other Semantic Web data semantic search engines have been proposed.

In general this branch of Semantic Web research is supposed to provide server-side technology that would provide similar search services as Google for the current Web. The difference is that Semantic Web search engines are not supposed to be used by humans but by other applications. As envisioned by many[12] those engines will be the bridge between user applications and the huge amount of information stored in the Web of Data. The most active among current research attempts to develop Semantic Web search engines are: Sindice[5], Semantic Web Search Engine(SWSE)[6], Watson[7] and Falcons[8].

Social Networking Applications based on Semantic Web

The initiative to create user friendly applications that would utilize Semantic Web data is ongoing however still in very early phases. Those applications are perceived as the last missing piece to render Web of Data and ultimately Semantic Web real. All the previously mentioned elements(the Web of Data initiative, the annotation standards like

SIOC, search engines like Sindice) only build the foundations. Unless Internet users are delivered real value they will never take advantage of what the research world refers to as Semantic Web.

Among numerous little products and demonstrators an attempt characteristic for the social networking area is an application called Twine[9]. It is a data centred social portal. Unlike others it is not trying to build a digital community based on social connections that people have in real life(i.e. Facebook, Twitter) but encourage people to create groups of interests and create their social networks from scratch based on areas of interest. Twine both harvests on the Semantic Web data published on the Web and produces the machine understandable data as well. The first option enables users to semi-automatically import the information of their interest into the portal, the letter provides means to push the data further into other applications and use it in personal environments. Although initially criticized and labelled as overrated Twine and other applications continue to grow and are envisioned to take make the final step of bringing the Semantic Web to the users.

Rapid Application development based on mashup creation from Linked Data

The tools for development of data level mashups and Semantic Web mashups in particular are still in a very immature stage. Like most of the Semantic Web tools, the applications are a result of research in progress with some early commercial products just emerging. One of the interesting visual mashup engines is called DERI Pipes[13]. The project is inspired by the Yahoo Pipes[14] and provides an execution engine and an online development tool. With the DERI Pipes it is possible transform various types of Web Data formats like RDF, XML, JSON or Microformats. During the process of creating a mashup, the user connects data import operators with transformation operators to produce a desired output. Depending on the imported format data can be manipulated with different query languages like SPARQL or XQUERY. Similarly like in the aforementioned Yahoo Pipes all interactions are performed in a drag-and-drop manner. The result can be saved online on the DERI Pipes server and retained for later execution just like a normal REST web service. Additionally it is possible to view the mechanics of the created pipe translated into a simple XML language. Therefore, it can be also exported and executed locally with a command-line tool.

Apart of the aforementioned DERI Pipes the most notable examples of other initiatives are Banach[15] and recently developed SPARQLMotion[16]. The first has been created as a part of SIMILE project. It extends the capabilities of the Sesame data store by implementing a transformation pipeline supplied with a number of operators. The SPARQLMotion is quite similar regarding capabilities to DERI Pipes, however it is a commercial product. The most notable difference is the model of script preparation and publish. In contrast to the DERI Pipes, the SPARQLMotion is a desktop application that can be integrated with other products of the same company to deliver a much wider range of capabilities like script testing or deployment.

3. Aspects of research work.

3.1.Objectives.

- recognize the most popular sources of linked data
- recognize the size of linked data cloud and measure the rate in which it is growing
- provide measurements of usage frequency of different types of linked data
- provide a study of the environments and tools that are used for mashup creation
- recognize areas in which mashups are constructed and goals that they fulfil
- recognize most successful mashup and Semantic Web applications that work outside the domain
- propose improvements for mashup development

3.2.Hypothesis.

Despite huge amounts of Linked Data and the current trend to produce even more, the contemporary Semantic Web applications are unable to use the potential of Linked Data. The reason for such state is not the lack of data but practical applications.

3.3.Experimental method.

Investigate the current applications, data sources used and analyze the most frequent problems that occur. Compare the Linked Data and Semantic Web metadata to past successful web annotation and metadata initiatives (i.e. RSS in comparison to SIOC ontology; REST vs. WSMO/SOA). Point out difference and list of advantages and disadvantages. Show the aspects of Linked Data that can be used in theory and the ones that do not work in practice. Measure the size of the linked data cloud and see what types of data grow fastest, what are the most popular ones. The variables regarding linked data size and growth can be measured with help of online applications such and Semantic Web search engines and information services like pingthesemanticweb.com

3.4.**Variables**.

- Size of linked data cloud (amount of triples)
- Amount of successful Semantic Web or Linked Data initiatives/projects used outside the domain
- linked data growth rate (weekly, monthly)

3.5.Schedule.

Phase	Activities	Time
Initial analysis	Search and analyze literature about usability study	2 meses
	Determine main lines of research in the area of Linked Data and Semantic Web	2 meses
	Revise the bibliography in each area	2 meses
	Analyze the state of the art in each area	2 meses
	Analyze the state of the art and bibliography in connected areas about metadata	1 mes
Execution	Choose representative linked data clouds in different area (i.e. dbpedia)	2 meses
	Design a way to monitor the data growth and metrics to describe it properly	1 mes
	Monitor the data changes	2 meses
Analysis	Analyze the linked data cloud monitor results	1 mes
	Based on comparison to the related work from different domains analyze the metadata	1 mes

Phase	Activities	Time
	usability problems of current Linked Data cloud	
	Review the applications created in different areas that use the metadata	1 mes
	Determine the lacks applications and user needs	1 mes
	Determine is there a relation between the amount of data and the usability problems of Semantic Web applications	1 mes
Conclusions and results	Evaluate the gathered knowledge	1 mes
	Propose changes and possible applications	1 mes

3.6.Web Queries.

1. IEEE library "usability study" search
Link: http://ieeexplore.ieee.org/search/searchresult.jsp?SortField=Score&SortOrder=desc&ResultCount=25&maxdoc=100&coll1=ieeejrns&coll2=ieejrns&coll3=ieeecnfs&coll4=ieecnfs&coll5=ieeestds&coll6=preprint&coll7=books&coll8=modules&srchres=11&history=yes&queryText=((usability+study)%3CIN%3Emetadata)&oldqrytext=((usability+web)%3Cin%3Emetadata)&imageField.x=0&imageField.y=0&imageField=((usability+web)%3Cin%3Emetadata)&radiobutton=cit

Selected interesting results:
- Ramalingam, S.; Iourinski, D.; "Using fuzzy functions to aggregate usability study data: a novel approach"
- Ricks, K.; Arnoldy, B.A.; "How to conduct your own usability study"
- Iourinski, D.; Ramalingam, S.; "Using Dempster Shafer theory to aggregate usability study data"
- Rosenbaum, S.; Kantner, L.; "Learning about users when you can't go there: Remote attended usability studies"

2. Springer Link search for "semantic web usability"
Link: http://www.springerlink.com/content/?k=semantic+web+usability

Selected interesting links:
- Ahmed Seffah, Michel Desmarais and Eduard Metzker, "HCI, Usability and Software Engineering Integration: Present and Future"
- Anthony Jameson , "Usability and the Semantic Web"

3. Google search for "semantic web problems"
Link: http://www.google.pl/search?hl=pl&q=semantic+web+problems&btnG=Szukaj&lr=

Selected interesting results:

- "Problems of the Semantic Web"
  (www.semanticfocus.com/blog/entry/title/5-problems-of-the-semantic-web/)
- "Semantic Web: Difficulties with the Classic Approach – ReadWriteWeb"
  (www.readwriteweb.com/archives/semantic_web_difficulties_with_classic_approach.php)
- "Semantic Web roadmap"
  (www.w3.org/DesignIssues/Semantic.html)
- "Layering the Semantic Web: Problems and Directions"
  (204.178.16.26/who/pfps/publications/layering.pdf)

4. Google search for "usability study in Semantic Web"
Link: http://www.google.pl/search?hl=pl&q=semantic+web+usability+study&btnG=Szukaj+w+Google&lr=&aq=f&oq=

Selected interesting results:

- "The Semantic Puzzle | Usability"
  (blog.semantic-web.at/tag/usability/)
- "User-Centered Design for the Semantic Web"
  (www.dfki.de/~jameson/iswc07-tutorial/ISWC07-tutorial-UCD.pdf)
- "A Flexible Dialogue System for Enhancing Web Usability"
  (http://data.semanticweb.org/conference/www/2009/paper/164/html)
- "Why Evaluating Semantic Web Applications is Difficult"
  (swui.webscience.org/SWUI2008CHI/vanOssenbruggen.pdf)

3.7.Bibliography.

[1] T. Berners-Lee, Semantic Web Road map, (1998).
[2] I. Herman, Tutorial on Basic SW Technologies, (2004).
[3] RDFa Primier, http://www.w3.org/TR/xhtml-rdfa-primer/
[4] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, G. Tummarello, Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web. In Proceedings of the Proceedings of the 5th European Semantic Web Conference 2008.
[5] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, G. Tummarello, Sindice.com: A document-oriented lookup index for open linked data. In International Journal of Metadata, Semantics and Ontologies, 3(1), 2008.
[6] A. Harth, A. Hogan, J. Umbrich, S. Decker, Building a Semantic Web Search Engine: Challenges and Solutions. In Proceedings of the 3rd XTech Conference, 2008, Dublin, Ireland, 2008.
[7] M. d'Aquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou, and E. Motta. Watson: A Gateway for the Semantic Web. Poster session of the European Semantic Web Conference, ESWC 2007.
[8] Y. Qu, G. Cheng, H. Wu, W. Ge, X. Zhang, Seeking Knowledge with Falcons. In Semantic Web Challenge, 2008
[9] Twine website, http://www.twine.com/
[10] SIOC project website, http://sioc-project.
[11] SIOC W3C submission website, http://www.w3.org/Submission/2007/02/
[12] M. d'Aquin. Building Semantic Web Based Applications with Watson. Developers' Track of the WWW 2008 conference, Beijing China
[13] DERI Pipes website, http://pipes.deri.org/
[14] Yahoo Pipes website, http://pipes.yahoo.com/pipes/
[15] Simile Banach project website, http://simile.mit.edu/wiki/Banach
[16] Sparql motion website, http://www.topquadrant.com/sparqlmotion/