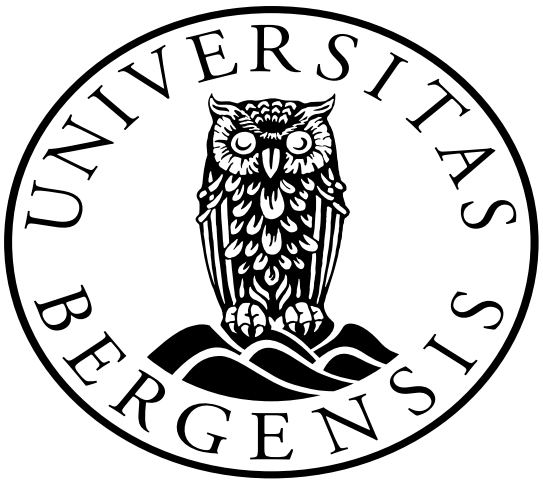
The Semantic Web from a humanities perspectives -

Using The Discovery Project as an annotation

to the semantic landscape.



Pia Jannike Virmalainen Jøsendal

Master thesis

Department of Humanistic Informatics

UNIVERSITY OF BERGEN

Spring 2008

Introduction – scope and objective of the thesis

Why attending to this field of research?

This thesis will address the semantic web, not from a technical view, but from a proximity and interaction view reflecting upon the purpose and usability of the semantic web and what it can offer. It is a field that has mainly been developed in and focused on in engineering, enterprises or within computer sciences, for barely twenty years. However, most of the work done in the beginning was in the form of research, and only within the recent five to ten years has there been an upswing in the production of practical solutions. Practical research on how to approach the humanities is still young, as is the entire field of semantic web. Considering the technical challenges of capturing semantics it is interesting to examine how the humanities can benefit from using the new available technologies.

The landscape

The purpose of this thesis is to investigate and shed some light on a project that may or may not echo the humanities' search for improved ways of collecting, retrieving and sharing information and knowledge. Restructuring material to improve knowledge creation and making information retrieval more accurate and efficient has affected most academic fields, also within the humanities. Digitalising material alone will not significantly contribute to or generate higher levels of knowledge. Firstly, information tends to be swallowed in the vast amounts of information available; an issue that is not outdated, nor new. Secondly, regular database searches set significant demands on the user in terms of recovery. Making use of the materials through alternative structuring and additional applications might, on the other hand, shed some light on an increasingly bigger problem: How to make use of all the available material online?

There is also the problem of building and re-shaping information as the knowledge levels increase. Creating a web of data instead of documents is one solution to this problem, which has provided several theories about a *semantic web*. The semantic web package is not the only solution, but one attempt among others. Topic Maps, ISO standards and UN Core components are all such other ways of "improving" the web and information retrieval. Some of the

solutions focus more on the web than others; some on a specific technology, but these approaches are exluded from the semantic web discussion of this thesis. Research and practical projects on the semantic web have mostly concentrated on materialising results for the social sciences or natural sciences. *Discussing the Discovery Project in the light of this allows us to look at the semantic web from a humanities´ perspective*. This project, which I will elaborate on page 7 and analyse in the final chapter, will try to provide a solution for the field of philosophy on categorising difficult ideas and mindsets and providing a framework for other humanities to hopefully follow or learn from.

Constructing a new structure raises the need for technological development that can provide a good supportive framework for computers in which content can be semantically structured. One suggestion is a combination of using OWL (Ontology Web Language as frameset for knowledge representation) together with RDF (Resource Description Framework for metadata) and XML (Extensible Markup Language functioning like grammar for metadata) to facilitate the structure, vocabulary and semantics needed by computers to process content and not only display content after a query. It has also been suggested that computers can access necessary and relevant information that makes structuring data more efficient if data is supplemented by metadata and contextual data and links. This however presents the problem of having to, for instance, categorise content and classify complex concepts. It also introduces the central and much debated issue of whether technological inventions can convey meaning or knowledge to human users.

Finally, *the semantic web cannot be dealt with autonomously* because there are many influences that effect and change how the range of the technology is perceived. It is important to distinguish and outline these influential surrounding fields in order to constructively be able to treat the concept in the right setting and consequently not to involuntarily simplify the interpretation of it by adding only one meaning to a complex concept. Several movements on the web, in organisations, in society, learning and knowledge acquiring situations have helped stir the notion of knowledge supportive structures. This view of autonomy stretches further than just placing an object in the right setting.

With this autonomy in mind, there are two specific implications that separate the theory in this view. If a semantic web is supposedly performing autonomous actions to withdraw or emulate knowledge at any level, it is no longer dependent on user participation. The system

acts on a foundation of its own set of defined activities, transferring all responsibility of deduction and reasoning to the system.

First order logic starts to resemble some sort of techno-autonomy where reasoning and axioms provide the only necessary framework for deduction and conclusion. The search for an automatic feature of production of meaning creates significant implications. The notion of autonomous systems carries with it the idea that people are redundant as decision makers and that the abilities of reasoning, semantics and logic are transferable qualities and applicable to technologies. Such macro systems elude the complexity of a context to the benefit of reasoning power and elicit a system where its purpose to support or help people becomes superfluous.

The alternative perspective treats the semantic web in a manner more similar to a neural network or connectionism. People are not reduced to bystanders, but are instead providers, moulders, producers, consumers and recipients to such an extent that on one level people are the connective links of knowledge and the semantic web is a foundation to work upon.

There is thus a tension between an idea of a system that is acting on behalf of people, where activities are removed and transferred away from people in order to be incorporated as machine transactions. This autonomous thinking has its counterweight in the functional approach that reviews psychological, social, cognitive, socio-economic, cultural and other perspectives to take them into account when weighing connective relevance.

Linnaeus and Buffon is a good example of the strained relation between respectively the rigid and systemic versus the relative contextual dependent sense-making (Lambe 2007). The functional perspective of an object's dependency of its contextual placement was to Buffon the answer to a taxonomy/taxonomian view of classification and mechanically structured objects. This tension between Linnaeus and Buffon may be applicable in the semantic web landscape today. Sense making is, exponentially to the number of recently initialised projects a reoccurring focal point to the discussion, but it should be noticed that it is a spectre of critical angles and not merely a distinction between the automatic and social.

The psychological, social, cognitive and cultural aspect can be imagined as opposite of autonomy. However, there are some that see automatic processes as an extension or even

symbiotic embedded quality that lies within man. The idea of “machinic thinking” is presented by Felix Guattari, as a quality of humans too, not only machines. Our cognitive minds are conducting machinic actions to induce responses in social environments. People are the consolidating factor of merging sets of processes and things. They themselves have the ability to coincide this in the same manner as computers of today merge things together on behalf of people. So to speak, people are acting as machines and the development of computers is a materialisation of this evolution.

Indifferent of the perspective, I argue that the tension between the two views has turned the landscape into a multifaceted landscape of which tensions grow stronger, but also build bridges between the various disciplines and merge into novel approaches.

In addition there is a needed change from the theoretical aspect to the operational in this respect. Realising the semantic web has required a turnover of ideas into real world settings which has perhaps resulted in losses in translation. Deployment of the humanistic view in the operational is my focus. It is in this multifaceted landscape that the Discovery Project has the real challenge of facing some of the issues that concretely will reveal themselves on an operational level. How the project answers to these challenges and what approach they choose to have in the process is what I hopefully will be able to elucidate at the end of this thesis. My arguments descend from the middle of the two counterpoints of autonomy, with an emphasis on the social involvement.

An introduction to the Discovery Project

Functioning as an observer in this landscape I find it useful for the argument’s sake to have practical examples from real world projects. Text encoding and library sciences have significantly contributed to the acknowledged need and usability of digitalised structured text and discourse. The Discovery Project is the latest emergence of former such projects. The relations founded between engineers and humanities generated new ambitions and needs to cover mining and mapping of texts. As a growing outcome of this advanced field, the Discovery is among the first that attempts to semantically enrich corpuses of this size and with a content of such intricate character as philosophical texts have.

The Discovery Project is a collaborative project funded by the EU`s *e*Content*plus* programme. This programme was initiated in 2005 to make digital archives, libraries, content of scholarly, cultural and scientific character, more "accessible, usable and exploitable". Until 2009 the Discovery Project attempts to construct a website that collects, indexes, expands and shares various *philosophical* texts and other multimedia on a platform that encourages knowledge sharing amongst interested parties. The targeted users are scholars and professionals doing research on philosophy.

The platform is based on semantic technologies such as ontologies, RDF and XML and will be offering users the opportunity to help moulding the size and shape of the content based on their knowledge and experience with the raw material (e.g. texts). In this way it is intended and favoured that the structure and the content provided will with this solution be "constantly improved and augmented by the scholars who use it" (The Discovery Project 2008). In order to do so, any "user will be able to contribute to the project by submitting scholarly essays, reviews, commentaries, annotations, metadata, or even new ontologies" (The Discovery Project 2008). Defined as a semantic web technology project, The Discovery Project attempts to enable better communication in its construction by increasing the efficiency of knowledge transfer between computers. This way the results presented to a user should be relevant without requiring the same effort of performing searches as with the existing web.

In the description of Discovery at EU's respective website of ongoing projects the Discovery Project is categorised under "Cultural and scientific/scholarly content". In the online abstract it can be read that the project aims to collect digital content to enable and "perform sophisticated queries, apply inference rules and above all to semantically enrich the data" for a peer network to subsequently apply new knowledge on the existing content. The purpose of the project is thus to enable an efficient environment in which philosophers *can work with and share their knowledge* over the web and not merely be presented with the content in a scholarly manner (EU online).

Several instances are involved in the Discovery Project, among them RaiNet, who have a large collection of various multimedia objects that are also included in the semantic enrichment process. ITEM is a leader in the adoption of digital technology for modern literary manuscript studies, ILIESI specialises on philosophical lexica, Net7 specialise on Open Source and DEIT recent focus is on semantic web projects and work on multimedia,

knowledge representation and general informatics to mention some. These are all specialised groups within their fields that work on the Discovery Project.

I have conducted my research with the project group situated at the Wittgenstein archives in Bergen (WAB), a part of the The Research Group for Text Technologies, UNIFOB AKSIS and any conclusions that can be drawn from this work are based on this departments work as *a part of* the Discovery project. Again, it is based on my experiences at WAB and I am not necessarily drawing conclusions on behalf of the entire European region of the Discovery Project. WAB works on preparing the Wittgenstein corpus, 5000 pages of the Nachlass, to make it available online. It is already digitally available, but is now under reconstruction in Discovery to provide a better use of the material.

Considering the number of people theorising the issues concerning semantic web development, there has been done surprisingly little within the field of humanities. The Discovery project is a project of larger dimension than many of its predecessors in semantic web research as well as one of the pioneers within the humanities. I am looking at this project as one of the pieces in a puzzle of creating a different web.

I will focus on the theoretical aspects of enabling possibilities for creating a semantically structured web, and will explore the Discovery Project as an extension to general tendencies in the landscape. Through my research I have also explored how the project group located in Bergen attends to the respective challenges. There will be two focal points in this paper concerning these challenges. First there is the altered concept of knowledge in technical environments and organisational contexts. The second focus is on the idea, materialization and impact of a semantic web. Both will solicitude the viewpoints on the Discovery project as well as make up the framework of this paper.

The Discovery Project will not be further commented on in the final chapters of this thesis. I will base my findings in this project on results from the theoretical discussions about the semantic web. I find it necessary to have all the perspectives at hand before making any comments about the presence of the Discovery Project in this landscape.

Explorative research of theory and method

Due to the complexity and vast landscape of theory that the semantic web touches upon there are no clear linear ways to acquire the information needed for this particular project. Therefore I have used a style and method which enable the best possible search results based on the limitations of my research frame. By proposing various theories based on continuous assumptions and acquired knowledge throughout the process it has been crucial for new theories and directions to be reviewed.

As mentioned, the foundation for this paper is the theoretical perspectives and coverage of ongoing discussions within the semantic web community. I am conducting theoretical research on the field of semantic web to create an overview and understanding of which contextual forces that influence the development. This will provide a general setting for placing theories and ideas, sufficient to form a landscape in which I attempt to provide an understanding of the Discovery Project. I have been questioning some of the perspectives, and commenting on the versatile or contradictory theories that affect the interpretation of what the semantic web implicates. Technologies that are in use have been briefly outlined, but not further discussed in detail as it is not a main concern in this thesis. However, some technical aspects will be covered as the work of preparing texts in the Discovery Project and parts of the difficulties within the semantic web concern the technological reach.

Due to the practical example of the Discovery Project I needed a method of comparing this concrete practical example to the theoretical landscape. The biggest concern when attempting to draw conclusions was consequently how to compare these two considerably different sizes and dimensions. Since I did not wish to bring any assumptions to the outcome, it can be considered to be an explorative style of approach to the subject matter. Grounded theory could best be seen as a valuable approach to this thesis because "grounded theory depends on methods taken by the researcher into and close to the real world so that the results and findings are grounded in the empirical world" (Patton 2002, 125). I spent the majority of my time searching through, reading, including and abandoning what are considered to be central texts and articles as I found that the semantic web discipline is highly fragmented. Many theoretical directions were discarded as irrelevant to the practical example at WAB, but at the same time, it shaped the focus of the aim of this thesis and gave me input on what to look for at WAB. Based on my research at WAB I was able to confirm theories that were relevant to

my research and it also shaped the choice of theory which I considered as necessary in order to evaluate the project in the appropriate theoretical context. In addition, because the thesis is somewhat interdisciplinary of nature as is the field of humanistic informatics, there were uncertainties connected to what theory that would be applicable.

I conducted field research at WAB, but again limiting my research to only touching upon certain aspects of the project. This is due to several practical issues concerning the time frame of both my research project as well as the Discovery Project. My time frame for research has been limited to one year as this is a master thesis and not a doctoral dissertation, which naturally limits the scope of my research. In addition the Discovery Project is a three-year project only to be finished in 2009 and I conducted my research in the middle of the production process. Naturally, the project is not yet a finished product, and I had to find other focal points to research. Finally, I did not have access to sufficient resources to conduct research that encompasses the entire Discovery Project, since its participants are located all over Europe. A natural limitation is to concentrate on the participants at WAB for most of the examples.

Also, since the Discovery project is a work in progress I could not analyse its state from the viewpoint of a finished result, nor had I any final reports at hand that could describe final research results. The best way to produce findings in this case was to continuously observe the participants and have a few informal conversations and conduct interviews as a supplementary understanding of the project development. Considering that the project is doing construction work in an unknown territory, observing the participants handle possible obstructions was a good opportunity to evaluate how they themselves interpreted the project. I tried to capture their understanding of where the project placed itself in the semantic web landscape and I also wanted to see how construction issues were approached. Finally I attempt to analyse the projects setting in the context of a semantic web.

To solve the above mentioned limitations to the research aim of this thesis, which are the issues of a limited time frame, two considerable different dimensions, the research object as a work in progress and few available resources, I decided upon an economical approach using a “debriefing method”. Subsequently, I used the Discovery Project as an “annotation” to what is happening within the field. While an annotation normally is meant as a critical, descriptive or evaluative note often adding or being a supplement to a citation or bibliography, I used it as

an evaluative note to highlight the distinctive features within humanities that could draw attention to certain interesting elements of the semantic web landscape that are widely debated, but often dismissed to the benefit of practical solutions. All this means that I use only pieces of the Discovery project that can be compared with the landscape and discuss critical issues regarding the semantic web. Finding information about problem areas and concerns within the project group about these issues was also partially available to me in the groupware (a community discussion forum containing posts, reports, comments etc) used by the entire Discovery project's participants.

As the Discovery project is not finalised yet, it is a special opportunity to be able to view some of the dynamics and opinions in motion during the various stages of development. Conducting most of my research as an observer of the participants of the Discovery project, I have attended only one meeting held at WAB, because there was only one meeting arranged from the time I got involved. I still got an idea of what their objectives were for the project at that time and at the same time I could observe both what issues were considered to be important to discuss within the group and the group dynamics. After establishing some relationships with project members at WAB, both during this meeting as well as on more casual occasions, I chose four participants for further interviews. My sources are kept anonymous under a promise of not using direct quotations from the debriefings. As the project is a work in progress, internal opinions may harm the group's dynamics and integrity. Questions need to be answered under safe circumstances and I also need to avoid incrimination of peers of any kind. I was lucky to be introduced to several participants through a reliable source in the project. This helped when I needed to conversationally ask questions and get information during my writing process. I have conducted informal conversational debriefings which were recorded, but unfortunately due to a misfortune two interviews were lost and partially covered. From these interviews, I have used secondary sources of personal notes. In addition to the interviews and informal conversations, I observed parts of the groupware in use in order to follow the progress and to verify any possible barriers. This latter method was used to be able to follow the progress of the project for my own benefit and not as a central part of the analysis. The only material I had access to in the groupware were reports and internal notes on partial elements of the project, which have been used as secondary observation material to establish a foundation for basic understanding on the project and its participants.

With this highly dynamic and fruitful interaction between theoretical guidelines and practical insight I have been able to grasp a much wider perspective on problematic issues. Theories have been included, changed, discarded and altered in the process. A lot of the work has been based on the theoretical and recent research performed by the most central actors in the field. I have in addition to theoretical articles and documentation, also attended seminars and participated in a conference in Norway, "Semantic Days 08" where I had conversations with some of the most central actors, producers and theorists on the subject. This has given valuable information and provided me with a fairly updated understanding of the complex motions on the semantic web.

Research questions

After preliminary reading I was left with several questions concerning the surrounding elements of the semantic web, some of which there are no one-dimensional answers to. Dynamics of the society, our history, technological realisation and human behaviour all address fundamentals that cannot possibly be resolved in one master thesis. Some questions I wish to explore in this thesis are general questions about applying semantics to the web:

- What is the purpose of adding semantics to the web?
- What issues would it attempt to deal with?

More specifically, I will attempt to analyse the reach of a semantic web approach in my practical example of WAB as part of the Discovery Project. Writing a thesis in humanistic informatics, it is important to keep focus on the humanistic values and research how these can come through in a technological setting and how new technologies approaching humanistic values can actually be applied to humanistic disciplines.

- How can Semantic web technologies be made useful in humanities?
- Are there any specific challenges in connection to this in which case how may they surface in the Discovery project?

Part 1

Early seeds of the web

> I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.
>
> (Berners-Lee 1999)

This widely quoted paragraph of Tim Berners –Lee stated the semantic web vision, and it was written by the same man who enabled this vision to begin with. The very early stages of the web derived from a need Tim Berners-Lee recognised in 1980 when working on a physics project. Linking note cards together in a tracking system later became the early steps towards recognising a far greater need – a global information exchange application (Berners-Lee 1999). Berners-Lee wrote a proposal when working at CERN in 1989 where he suggested that they should use the Internet as a platform to share information to ease access to documents (CERN 2008). By combining a hypertext structure with the Internet, the combination enabled a scalability and extensibility in which large volumes of documents could be stored, added and also retrieved again. Ted Nelson's description of hypertext is as a nonlinear or nonsequential space made possible by a computer (Noah Wardrip-Fruin 2003). Instead if a traditional tree structure, the hypertext format on the Internet provided a system where documents linked together by keywords would appear and could be used by fellow researchers to track similar projects. This solution would not impose limitations or restrict the content, because unlike a tree structure the content could be found independently of an index.

Extensibility was Berners-Lee's main focus and interlinking like the hypertext system enabled a system which could be extended in terms of content, but equally important in terms of system, language and application versions. The combination of Internet extensibility and linking format would also allow researchers to find what they did not necessarily expect to find (serendipitous), thus expanding the research base. Berners-Lee called it a "multiuser system" in his original proposal and claimed location nor computer systems would appear as

obstacles (Berners-Lee 1989). Anyone could have access from anywhere provided they had a computer and an Internet access. Most importantly he sought it to be a "**pool of knowledge**" (Noah Wardrip-Fruin 2003, 797). What Tim Berners-Lee envisaged in the end of the 80s was meant to be a globally accessible system for everyone who sought to share information on a network of interlinked texts.

Tim Berners-Lee is thus referred to as the inventor of the World Wide Web (in 1990) and later, the initiator of the World Wide Web Consortium (W3C) in 1994. His accomplishments are accredited in nearly every book covering the development of the Internet. Not resting on his achievements, he kept instigating new projects and research to maintain progress for the Web to keep growing and evolving.

Linking together documents to create a web of content became a mind-boggling success. For anyone to produce content, placing it on the web and accessing the web from any location became reality for anyone with an Internet access. In fact, the web expanded at such a rate that document overflow soon became a reoccurring issue when dealing with content retrieval. The web's enabling key feature, extensibility, also caused the web to become too large for people to sort through in order to find satisfying information as searches became inaccurate and did not include all the available documents.

Searching for an approach to the semantic web – from displaying to processing

Today it is not possible to say exactly how many web pages are in existence at any singular moment because of the explosive expansion rate, but they are in the billions. Finding files and documents on the web is consequently not a task for manual search. Search engines (such as Google, Kvasir, Yahoo! and many more) took on the task of sorting the documents and stand today as a crucial and essential tool on the web to perform retrieval. The web has become dependent on these to index the world wide web of documents. However, since the web mostly is based on mark-up languages like HTML (Hyper Text MarkUp Language), information retrieved from search engines is based on the presentation and design information of a web page. Most retrieval complications are faced with the web's construction issue of such presentational formats.

The very first search engine, ARCHIE, enabled a user to search on what was stored on the Internet (not the just the World Wide Web) by searching for explicit file names. This was just a beginning and was nothing nearly as powerful as today's search engines. The search engine we know today began with the arrival of the World Wide Web. WebCrawler, Lycos and AltaVista were of the first generation that indexed pages to be searchable with key words and AltaVista introduced natural language query possibilities, which is now a standard feature of search engines. The second category of search engines that developed were the meta-search engines which reuse results from other search engines, but these do not have large market shares. Vertical searches on the other hand are of the most common search type, but as markets have changed, pure algorithms are not enough to induce wanted results. It is no longer only the amount of information that point out an interest in a new direction for searches, but the inaccessible connections of meaning in a document or webpage. A result set based on highly advanced algorithms does not display the meaning of a web page, but displays where the headline should be and in what style, and that the body text should come after the introduction paragraph and is written in text font Arial. Then the searches are performed based on where in the text or headlines a word appears and following displayed to a user in a rated result set of number of appearances of a word or phrase. Even though many search engines today have evolved and possess advanced classification algorithms that index, filtrate and categorise accordingly to belong to name, place, scholarly texts, pictures and so on, these mechanisms are not semantic web developments.

Another part of the problem is that searches need to deal with the multiplicity and selection of formats, languages and applications. Data collections on nearly anything are growing correspondingly to the growth of the web, social applications, multimedia and the freedom of personal expression. Commercials and product sales in addition to new forms of social software campaign for increased market shares, does so that it is not enough to base query results on only algorithms (Wall 2008). Per Gunnar Auran at Yahoo! explained in a presentation at the "Semantic Days 08" conference held in Stavanger, the alteration from a vertical search method towards a hybrid search models that they are attempting to take into use today to provide better searchability. The structure of databases, customised options and small, specialised vertical search features combined with typical web search methods of fixed algorithms and the less structured global document model provides *both precise match as well as scalability*. This sophisticated processing method can to a large degree handle some semantics, but the information is still often extracted from mark-up based web pages. Dealing

with global collections such as the search engines do on the World Wide Web, these methods taken into use are for now manageable and efficient. Auran quoted Hendler in that “a little semantics goes a long way” and believed that semantics added to the searches would improve retrieval, but a pure semantic web approach would not serve its purpose in this context. It would not be able to uphold in this type of large scale search simply because it required too much attention and maintenance for it to be worth it. A “little semantics” view is counter to e.g the database community who argue that search is doing well without any semantics. The separated view is pointed out and accounted for in the Amicalola-report about the database community and the semantic web where a review of database supporters showed scepticism towards the semantic web (Sheth and Robert 2002).

Classified as semantic web search engines are among a few Swoogle, SWSE (SWSE 2008) and Semantic Web Search (Semantic Web Search 2004-2007) which all base their result sets on searches from OWL, RDF files and RSS feeds. Swoogle’s statistics claim that 643,133,742 triples could be parsed from all semantic web documents (Swoogle 2007). Now, these engines hold a lot of potential, but then again, they only index and crawl the web for semantic content and are not all inclusive and global on the web in that concern. Also, current search engines’ user interfaces have a lower threshold for performing searches. From personal attempts of searching Swoogle, this engine is unfortunately far from simple to understand for the average user on the web. Simple searches are fine, but taking advantage of the more complex structure that semantic engines offer, requires a user to write some code. Provided manuals are often complicated and take too much effort for an average user to read. This will most likely change if these engines become more public and commercialised. There is no longer no doubt however, that semantic data do exist on the web, but collecting it and making use of it is still in the workings.

Practical approaches to the semantic web are very much in the process of being realised, but realisation of theoretical ideas does not imply a general accept for the methods used for realising them. The reach of the adoption; the acceptance of the semantic web idea, or the complete rejection of it, are so differentiated and many that the mapping of all these variations create a landscape almost as fragmented as the web itself. The semantic web seems to have spurred a line of technical, practical and ideological views and as increasingly many disciplines sees the possibilities of semantic web, the multidisciplinary involvement increases.

Still, I found it serviceable to have some distinctions in the approach to this fragmentation and I have come up with two categories that encompass most angles taken to the semantic web.

Product oriented view and user oriented view

From the adoption of new technologies and to the widespread use and incorporation of them there are always economic motifs involved somewhere along the development. Enterprises pay close attention to new inventions that may solve practical problems or become a major new product. To them, return of investment is the key. Where Berners-Lee saw a crisis or need he could provide a solution to, enterprises and vendors saw opportunity. Thus, there can be seen two clear views on the semantic web: the *product oriented* that focuses on producing deliverable products that satisfy the market demand for improved information channels and the *user oriented* that hopes to enable better communication between people and people and computers.

The producer view is not in any way restricted to large vendor activity. As with most definitions there exist nuances in between the two extremes. Anyone who produces anything will expect some return of investment, whether it is money, values, rewards or recognition. Also within academia, the development of a product or idea makes induces a producer. In a product oriented view, a market need is created/covered by the introduction of this new product. The idea of a market extend to the humanities context as complex transactions, social relations and values are considered carefully before there is any engagement in the product development stage. Knowledge of and appreciation of the surrounding influences of a business reflect in the highest degree the involvement of an understanding of human behaviour, history and culture, and cannot be neglected in product development. Therefore, the product oriented view also considers a paying customer, but only as a part of the totality of the products context in a market.

However, the user is not a sheer ignorant recipient of products. Motivation and reason for interest of use of a product decides whether a product succeeds or not. Awareness of these motivational factors are important when there are realistic needs or opportunities seen in interacting, supporting, improving any action for a user. Centring the focus of attention on what the user finds useful, takes other consideration into account. The context affects the user's position and disposition to approve or reject a product. From a user oriented view the

user is the main and only focus, where the product becomes secondary in a way. All the influential cognitive contexts that a user can have are what drive the development for user oriented view.

Even though the technology seems the enabling factor of the semantic web, and thus product development would be the central focus, there can increasingly be traced a shift toward or at least an increased awareness of the user centred approach. The primary push for technological development is not happening merely because the previous existing technology was/is terrible or insufficient. Whether this implied shift towards user awareness is realised in actual projects or remains a formulated goal will require further maturity of the technologies. Until now, search engines mostly base their searches on the advanced syntactic relationships between key words, phrases and documents, but not the thematic or semantic relationship between data. Consequently, it has not brought computers any closer to "understanding" the content they retrieved and presented. By making the web better computer accessible, queries would be ranked by usefulness (Decker et al. 2000) which ultimately would relieve people from unnecessary effort. In line with the argumentation, even though increasingly demanding, the technology still becomes secondary. Interestingly, the amount of effort that lies behind the technological development within semantics is perhaps, as Auran is saying, so time-consuming that it overshadows the semantic web's final function – to support users in the jungle of information. This view of "a little semantics" presents the importance of a definition.

Semantic technology versus the semantic web – different aims?

In my research I have found that frequently used terminology in the semantic web context is, in fact, of variable relevancy to the semantic web. While this thesis is mainly focusing on the semantic web, semantic web has been classified as an ancestor of semantic technology in a paper by (Sheth and Ramakrishnan 2003). This paper gives an account of the similarities between the two, but also the distinctive features that separate them.

> Currently in more practical terms, Semantic Web technology also implies the use of standards such as RDF/RDFS, and for some OWL. It is however important to note that while description logic is a center piece for many Semantic Web researchers, it is not a necessary component for *many* applications that exploit semantics. For the Semantic Technology as the term is used here, complex query processing, involving both metadata and ontology takes the center piece, and is where the database technology continues to play a critical role.
>
> (Sheth and Ramakrishnan 2003)

These terms are often confused, but to a large degree kept separated. Semantic technologies are based in a information systems camp which give focus to refinement and improvement to mainly database systems. Semantic web as defined by Berners-Lee opens up to cooperation between computers and humans and consequently opening up to human computer interaction and the following issues it poses. While understanding the importance of separating the terms, I also agree with Sheth and Ramakrishnan in the approach to semantic web technologies with the help of semantic technologies. *The maturity of techniques in semantic technologies should be taken into use to solve shared similar issues, as well as I believe it is equally important that ideas that form the semantic web should be considered when adopting semantic ideas in semantic technology.* I wish to emphasise the importance of not forgetting what the search and development is for. As written in the introduction of Sheth and Ramakrishnan (2003) I concur with the shared key qualities of semantic technology and semantic web as "representing, acquiring and utilizing knowledge". Thus, the main focus would be how to define knowledge and keep focus on representation, acquisition and utilisation of it, which is not a self evident remark, because it implies a heavy focus on the purpose of initially developing such projects in the first place. In addition to this, there are several actors that place themselves in the middle of this scale, claiming that both sides of the development can contribute to solve key issues surrounding knowledge.

The intention in this paper is not to give focus to all technologies that deal with semantics or whether one belongs to the other category. I merely wish to look at the premises for semantics in which invites to look at semantic technologies and semantic web technologies in a more unified perspective. The landscape also laid the ground for such an assumption as the fragmented field overlaps on approaches and angles taken to solve specific problems. The level of transference between the different fields seems to decrease the necessity of completely separating them with or without the word web, but rather give focus to the underlying ideas and foundation of a semantic approach.

Due to the fact that the field is so young, I look at both as a whole, but I appreciate the outer most opposite contrasts of intent. I have this as my main focus when using it as an angle to view the Discovery Project, though I also discuss others aspects and projects that might fall under the category of semantic technology. The two fields have many intersecting ideas and several commonalities in which some general ideas of purpose and means to achieve the final

destination can take advice from each other. I do not think that they should therefore be completely separated in this discussion, but understanding that there is in fact a difference is crucial for the understanding of the field.

Why the idea of a semantic web?

From the previous section on the history of the web, we can read that there are some issues that have lead to a need for alterations of this original architecture. Large volumes, complex interoperations and insufficient algorithmic search tools are some of the reoccurring issues. Mostly they are results of an incompatible level of communication between humans and computers. The web is readable to humans, but not to computers. This is partly due to the few available computer applications to bridge the gap between heterogeneous data of everything stretching from computer games to hypertext literature and pure text documents.

To Berners-Lee et al the pending feature of universality is considered to be the main focal point of the web and a sustainable property for the future semantic web (Berners-Lee, J. Hendler, and Lassila 2001). Covering the variable data formats and programming language standards to macro level formats as applications and software standards, the semantic web will enable incorporation between all vertical layers as well as equal horizontal layers. But universality functions on more than just a computer interoperability level.

> Web technology, therefore, must not discriminate between the scribbled draft and the polished performance, between commercial and academic information, or among cultures, languages, media and so on.
>
> (Berners-Lee, J. Hendler, and Lassila 2001)

Including all verticals of information is a significant part of the semantic web, as interoperability and merging of standards becomes increasingly visible of its primary aim. As Berners-Lee et al describes there should not be any discrimination between social, cultural or ethnical and linguistic points of view. This can be interpreted as an idealistic way of understanding the role of the web, a web that interpret and distinct the granularity of human perception and societal dynamics or it can be seen as an extension of the strengths of the existing web that embraces the heterogeneity and natural variety of human nature. Opening up for similarity and understanding between computers does not mean deteriorating or level

nuances on the web. It is about accepting the differences and working with these differences to create an equal platform for interpretation on equal terms.

Through preliminary reading it became clear that there is a common understanding amongst web communities, businesses and scholars that the web is in need of structure or a retrieval system better than previous ones. Striving towards continuous improvement is a natural human quality and improving the web is no exception. Judging what defines as improvement is, however, what causes the problems.

Practically handling the entire web as one problem space is, mildly put, complicated due to the complexity and heterogeneous environment of the web. It consists of multimedia in form of texts, audio and video in different formats, languages, and applications. Heterogeneous data creates a problem when you wish to retrieve information on a uniform platform so the content can be easily processed. Connecting data semantically demands interoperability and openness between software. It also poses a computational problem if all these barriers need to be "translated" to gain access to the content in question. The general voice against the idea of a semantic web is the one that says that the semantic web is just that – an idea. It is too abstract and is not handling the practical issues directly.

There are discrepancies the web itself creates simply because of its nature. Future solutions depend on bridging these gaps and making use of the strategic advantages that has given the web its characteristic nature. More concretely, the hyperlinking structure of the web today is a powerful tool for disseminating information, yet the same fundamental technology also decreases efficiency of content retrieval without a significant effort from the users. Additional solutions are needed to create more beneficiary and efficient technologies that can handle the wealth of information that exist on the web - *not by humans, but by machines.* When I use the term machines I of course refer to the software or agents that internally perform tasks or services on behalf of people. Still, the process of enabling the semantic web vision discussed in this paper is based on the belief of "machine-accessible and machine-processable representations" (Antoniou 2004, 65).

Then machines could make logical inference between different structures and data that the existing web technologies cannot perform without significant help from humans. By inference and computer accessible I mean software that can logically deduct an interpretation of the

information it is being presented with. Dieter Fensel criticise the idea of inference because machines will never be able to understand any logical connection between strings or pieces of data the way people do.

Most of the work done on semantic web technology has mainly been concentrated on a research level and though the idea was presented a decade ago, the practical solutions are still in their infancies comparatively. Considering the short time span, the technology and idea has not had the time to mature. Still it has made tremendous progress in developing ideas, possibilities and envisioned solutions. Alongside theoretical research several test projects have also been developed, but usually on smaller amounts of data and mostly within restricted problem spaces as pointed out by a Dr.Miller in a podcast speech at University of Cambridge (Miller 2007). Semantic solutions on larger corpora are difficult to achieve because of the amount of time, money, effort and knowledge it takes to build, manage and maintain it. Unfortunately, how to deal with increasingly larger data sets in addition to heterogeneous data sets has shown to be a challenging engineering task. This has also been one of the most critical problems in the discussion on a semantic web, especially when there are issues of transferring semantics involved. Which will be further elaborated on later, some argue that semantics will be lost if technical data streams are not coherent and cooperative.

One can say that dealing with the semantic web is like abstract problem solving. On one end of the scale it attempts to categorise and theorise on issues that include *all the material on the web* – and that is a considerable size of material. A more practical approach has been able to provide solutions, but mostly for *smaller closed-off projects* for now. Creative minds with different visions and individual ideas on how to provide suitable or revolutionary solutions provide the fast evolution we are witnessing. The results create a solid foundation for future projects. Both capitalists and idealists are part of the debate and development and there are numerous projects being initialised constantly to contribute to finding future solutions. Because this development consists of both market vendors and enterprises putting the development at practice as well as researchers, the outcomes and solutions provided are naturally somewhat different. This causes *a variation of standards, approaches and ideas* and is causing a somewhat fragmented and scattered landscape of innovations.

Various/Non-profit web communities and organisations also work on spreading *the idea* of a semantic web across the web. They are also working on practical solutions for embedding

semantics in our web of today as an extension to it rather than a replacement of the existing web. The eldest and largest coordinating body for web development and semantic web development is the World Wide Web Consortium (W3C) who is working on setting standards for the further development. Standardisations are viewed by W3C to be an important, and also crucial, step in the direction towards interoperability and collaboration for the creation of a semantically structured web. W3C are viewed as a central actor and a great contributor of rules, guidelines and validation criteria for appropriate development and usage of web software and web platforms.

W3C is not the only contributor working towards standard software, platforms and open standards to add more sense making into the web. Other contributors do not work towards the semantic web standards envisioned by W3C, but are working on alternative solutions to reach somewhat the same goal. Among others there is rdfa.info, a blog community who are spreading information on standardisation of rdf as the semantic html language of the semantic web. Microformats.org is also working on metadata. DBpedia (http://dbpedia.org/About) combines the success of the wikis and approaches semantic technologies in a manner that is manageable to a large scale source. Search engine Firefox experiments with ontologies for personalisation with the help of add-ons. Yahoo incorporates some semantic elements in order to do the same and Swoogle (yes, from Google) (*Swoogle semantic web search* 2007) is a major project on semantic development to improve the accuracy of search engines. The above list represents not only a large variety of involved parties with different approaches in the pursuit of a semantic web, but they visualise the tendency towards the potential seen by more than just a small community of especially interested.

Even though the working community on standardisation is highly established, there is still no consensus on what the purpose of the semantic web is or should be or whether a semantic web is a solution to the current web's problems with large scale retrieval and the quality of the output from retrieval. There is *no consensus formed on technological issues* either which is why many actors use bits and pieces of the available technologies to enable partial semantics. Within the most active semantic web fields as offshore, high technological companies and libraries there are different standards in use. Even though some technological solutions are reoccurring in multiple projects or have a wider span than others it does not imply that these are given new standards for the future development. On the other hand there are bodies as W3C who are working on the semantic "stack" of standard technologies with an all inclusive

approach. In short, non profit and profit organisations pull in different directions. In spite of this it seems anyhow that the communities are working together to create more common practice as they come closer to solutions in each base camp. There is a pull to each direction, and whether complete standardisation of technologies or the approach to enable semantics independently of any specific technologies, is likely to be a topic of discussion onwards.

One certain fact is that the development will continue at such a rate that the solutions of today, most likely will be gone or improved, and that it is important to realise that there won't be any final answers and no ultimate universal solution. Another asserted opinion/notion/note as a result of the fragmented landscape, which is crucial to the understanding of true semantic web, is the *use of terminology*. Due to the perspectives and disciplines involved in the development, the idea of a semantic web has to a large degree been adapted to already existing technology, ideas or aims. This has altered the perception of the semantic web and created its various outcomes. Therefore it is necessary to outline what this can result in and what motions that stir in the background and has contributed to moulding different projects.

There are several different dimensions to the development of a semantic web and there are several pulls in different directions concerning this. To summarise the discussion I begin with underlying the strong pull in different directions of respectively semantic web and semantic technology which have both contributed to a variety of solutions, but also a confusion of terminology. This terminology confusion has lead to a broad use of single or partial technologies claiming to be semantic projects. However, this does not qualify of semantic web standards according to the W3C unless they keep to certain other standards. Features of universality, interoperability, heterogeneity, machine-processable data and scalability of projects are all reasons and preferred solutions and are answers and arguments from a W3C view to the question of why a semantic web is needed.

As briefly outlined in the section above, there are several forces that have been behind and still are behind the semantic web discussion and the self expanding notion of the ideas keep stretching in different directions. By viewing the web's current qualities, its birth story and driving force it shows why and how the semantic web developers are not always concurring on how to approach the construction of large, heterogeneous scales of machine accessible structures when the terminology is poor and confusing, and the technology has had relatively little time to mature.

How the semantic web idea originated and its current state frames the importance of including the context. I have introduced some of the challenges, levels and angles this view has presented. The next discussion will focus on perspectives and approaches to knowledge acquisition and social changes. Challenges presented by these views materialises themselves in opportunities and concerns with the semantics. Semantics have been responsible of the many levels that discuss the semantic web. It is important to establish what central discipline/s that are used to see the intended aims for theories that is preferred to lay the ground for semantic development and use on the web.

Semantics needs context – but what context?

In order to make the topic of semantics somewhat approachable it is wise to provide an overview of what it implies. Semantics is complicated in itself, but even more so in its practical application. Defining what a dog is depends on what philosophical point of view one has. Making a system for how to apply a certain limited set of criteria for defining a dog poses a larger problem.

The word *semantics* derives from Greek and the word "semaino" translates as "to mean" or "to signify". It is primarily a study of linguistics implying a study of meaning in words, sentences and their abstract or implied ideas or connotations (Online 2008). "Sema" is the core of which "semaino" derives from and can be translated to "sign" which in turn implies interpretation. This added value of meaning changes according to context. Semantics is a field of study on how meaning evolves and changes and how for example a population relate to meaning. This study is indirectly a study of human cognitive processes of understanding and connecting objects of the real world in form of our languages. It has also been adopted as a cross disciplinary because of its wide range, but how can this be transferable with regards to computers?

The level of context applied to the field of semantics itself will clarify from what perspective the word semantics is interpreted from. Several schools or disciplines are making use of the semantics in computer sciences or computer science related disciplines (Sheth, Ramakrishnan, and Thomas 2005). Making a clear statement on what context semantics is interpreted in, determines how semantics can be used properly. Malpas (2002) is of the opinion that meaning

cannot at all make sense without the context taken into account. He focuses on context as the primary binding connection of meaning.

> Indeed,more explicit attention to the notion of context as such may turn out to be crucial in any adequate account of meaning and understanding whether from within a narrower linguistic or a broader socio-cultural framework.
>
> (Malpas 2002, 405)

This sentence centralises the key aspect of the context discussion of this thesis. Independent of the framework, “a narrower linguistic or a broader socio-cultural” as such, it all fills the role of the context. Without the context, meaning or understanding does not take place according to Malpas. In the semantic web context, this naturally also has significant ramifications. Whether meaning is attempted extracted on a computational level, between data in applications and alike or exchanging meaning between humans and computers or humans and humans using computers, understanding or meaning will depend on the context of the recipient. This is a line of thought that transfers across all layers of the semantic web. This also implies that considering context on semantic web has to consider what context meaning is intended for. If a computer is to understand a message, it must be based on the conditions and constraints on a computational level. Likewise, producing content for an expert group which is received by a local African school serves no purpose, results in a situation where meaning falls out of context and looses its intended meaning.

> Similarly, semantics can be seen as providing an ‘account’ of meaning, but only as it arises within an already circumscribed frame, and in a way that already relies on a grasp of the concept that can never be completely elaborated within semantic theory itself.
>
> (Malpas 2002, 410)

Here he continues to elaborate that context is ubiquitous and not explicitly outlined and that semantics inescapably relies on a contextual frame to evaluate a term or concept. Any applied semantics will not sufficiently explain the meaning of something by it self. Meaning becomes a relative instance between the concept or term and humans perception of that term or concept. However, by applying his”principle of contextuality”, which is a response to the criticism of holism, he limits context to be specific otherwise one would need to have knowledge of the entire system to understand how one piece fits together with it. The principle of contextuality opens up to a more smooth and flexible notion of knowledge acquisition.

Influences that are focused on in this discussion are the contexts of linguistics, psychology and semiotics on the semantic web. Linguists constantly work on defining, refining, adapting and replacing words and providing a descriptive framework in which this can work in a computational setting proves to be a challenge of limitations and reasoning qualities possessed by humans and not computers. Within the field of psychology there is a constant ongoing discussion whether meaning or processes of knowledge acquisition is considered as cognitive or collective actions (cognitive versus constructivism). This affects the outline of the semantic web in usability and interaction with people and questions the range of changing the web to convey more meaningful information sets. Lastly, meaning implied in representation, symbols and greater contexts needs prerequisites of human association, knowledge and context and so the connotative meaning as a recognisable pattern on the semantic web is also a part of discussion.

These three might only be a few interpretations of semantics, but what is made clear here is that there is not just one view point on the semantics in a semantic web context. Even more confusing, they are rarely separated in the discussions and the mixture of disciplines has created a multiplicity of theories and angles of approach to the semantic web. It also reflects just as importantly, that the discussion concerning the semantic web began early on in each separate discipline, and merged in this particular context.

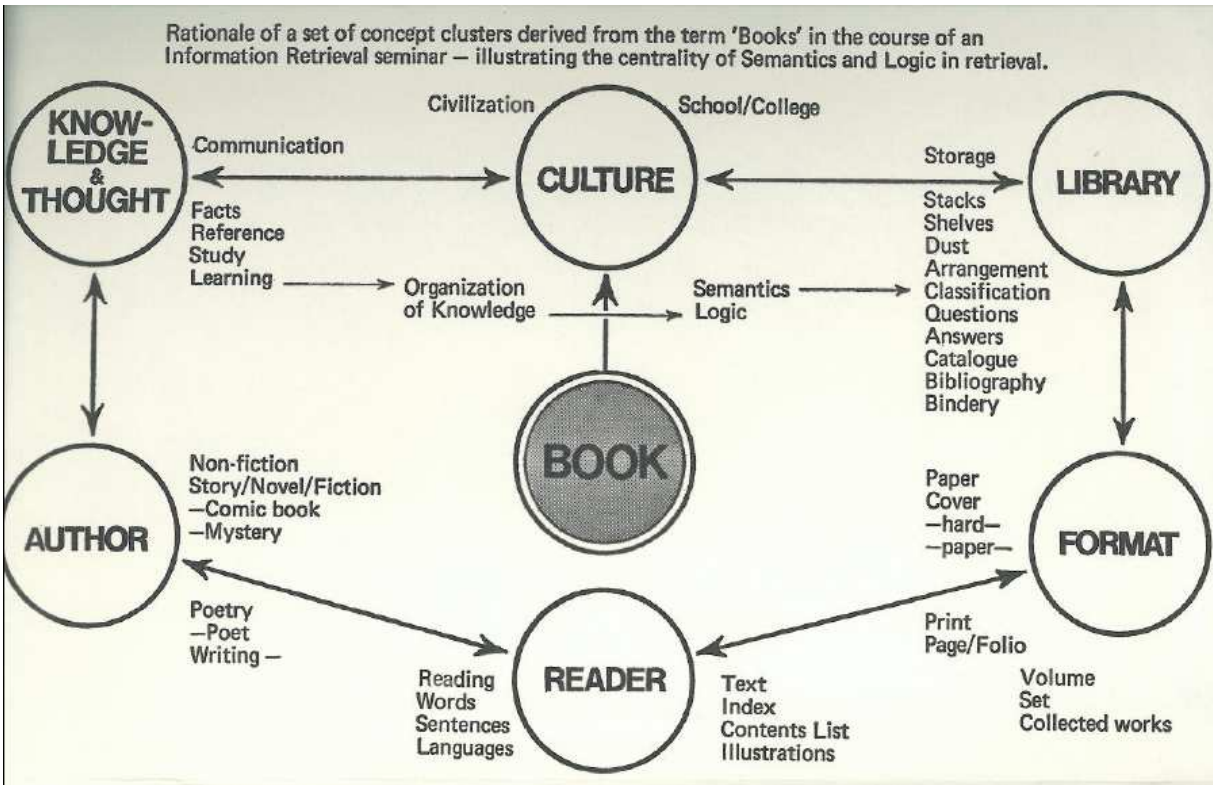
I choose not to separate them in this thesis either. As a continuing argument from Malpas, interpretations of a semantic web depend on the mixture of views to envision possible scenarios, its effects and impacts. More importantly, they affect different levels of the semantic web discussion – ranging from the micro levels of programming to the macro levels at the end user side.

As for a historical context of semantics, the field blossomed in the time of Neopositivists in the 1920s and 1930s where scientists and philosophers' early connections were seen as discussion arose on semantics as an extension to logic (Online 2008). Semantics as a field of research is of fairly recent origin. Already in early literature on library and information science there can further be drawn parallels to the definition and use of semantics. In McGarry and Burrell's book on semantics in an informational setting the authors draw up an interesting schema on relations between various elements and places and places semantics in a

context (McGarry 1972). The schema lines up a map in which the landscape of today's semantic web also can be trailed.

In more detail, "the book" is in this scenario replaced by data – including every form of content that can be placed in the web, any type of multimedia. "Culture" is represented by the World Wide Web which stores and holds all the content in spaces with markers such as URI and URL. "The formats" in which the contents are being presented consist of the various programming languages, presentational schemas, applications, platforms and research engines presentation. Interpretation of the content happens through the formats. Interaction between "the reader" and "the author" presents itself differently on the web than in a library and is hardly separated. A culture is still a culture of people in a community with common interests. The provision of knowledge and thought to the culture happens as a cooperative interaction between author, reader and culture rather than in a linear trail of action. With the help of the reader's interaction the author is presented with new thoughts and knowledge in which is added to a culture which in turn comes back to the reader. On the web this contact surface is much wider and opens up for the reader to be the author and directly interact with and influence the culture. How this process of providing knowledge is anyhow presented in McGarry and Burrell's schema as additions to the semantics and logics of the library (www). By refining the semantics, it refines the library and influences and is under the influence of the formats in which is being presented. Again, I wish to review the McGarry schema. In a semantic web environment, the semantics/concepts used by the culture can vary and therefore these are different in each set of cultures.

The schema is presented to show how important semantics is to the process of refinement of content and improvement of cultural understanding. It also illustrates what role semantics play in this process as a catalyst of progression. It shows how the force of semantics drives the steady circle of knowledge acquisition, inclusion of author and reader and refill of knowledge in the repositories of culture.



(McGarry 1972)

In a computational setting the interpretation of semantics is much more restricted. While the web opens up to multiple perspectives and complexity, it will also need restrictions concerning computational problem spaces. Data exists on the web, but not as autonomous entities without any connection to each other. To make useful connections between those entities and make strong semantic and logical connections the entities need ties that connect them together. Data can exist in a closed system which serves a certain purpose. The semantics that are needed in these scenarios are limited because they only need to make sense in that certain context. These closed world object relationships are defined according to the system. It is how they relate to each other that gives them their implied meaning. These relations and set of rules for how they should be related are provided by humans who define the semantic value of a word or a sentence or a text. Transferring this to an open world assumption as the World Wide Web, causes discussions on context and user purpose of the needed amount of semantics and how semantics is generated.

Lurias drawing of four elements may visualise what is meant by relative contextual meaning. In 1953, Lurias developed an experiment of placing respectively a saw, an axe, a hammer and a stack of lumber together in the best logical way. A person was then asked to group them

according to belonging. The test would reveal that those who grouped the tools together were strongly influenced from the Western line of thought while those who grouped the axe, the saw and the logs together reflected an Eastern line of thought.

Activity Theory can also contribute to the human perspective of semantic web according to the above schema. Activity Theory can provide to be an important tool to draw out the landscape of which human-computer interaction takes place. It can also remind researchers what context to relate their subject to. I believe a semantic web approach especially needs to consider human relations in order to understand and be aware of the intention and reach of "human actors".

> The basic principles of activity theory underwrite the emphasis in interaction design on the social, emotional, cultural, and creative dimensions of human actors in shared contexts.
>
> (Kaptelinin and Nardi 2006, 6)

I would like to emphasise "dimensions" and "shared contexts" as the focal points of this quotation. As an addition of theory I support that Activity Theory underlines the importance of the dimensions surrounding the technology. Human culture including the psychological, social, cultural and linguistic aspects jointly combines to make the landscape for the semantic web discussion in this thesis and Activity Theory binds them together.

Activity Theory is based on work by Vygotsky and was originally developed by Leontév. A subject is in activity with an object where the relation with the artefact is in itself a manifestation of the subject. Thus an activity is not what the subject does, it is what defines it. An activity is what manifests itself in the subject when an activity is performed. This changes depending on what object a subject is performing an activity with. Activity Theory offers an opposite perspective of context, history and culture instead of an autonomous view to the semantic web. On this point there is a major counterpoint to the development of a schema such as McGarry's, where semantics is the driving force of the circle. On the other side, what defines semantics as an external factor? Couldn't semantics be viewed as an activity that arises when a subject is in activity with an object? The semantic context would change depending on what object is included in the activity. The core issue of defining something is to contextualise it, and here comes the discussion of whether the definition is a commonly

understood definition or an adjusted definition accordingly to internal cognitive knowledge and experiences.

The Activity Theory opens up to complexity, but the semantic web limits this complexity in restricting feedback from the system. In this view, the systemic form of the semantic schema might appear to come up short, because of its imposed predictable and scheduled circle of motion. However, it can also be contradicted at first as it can be assumed that the closed premise of such a cycle could result in an endless spiral leading to stagnation. Merely represented as boxes and arrows it can appear to have a closed world assumption. The never ending circle where the book remains the book and the arguments and dynamics aren't providing new dimensions result in a systemic and locked environment. My counterargument in this regards is that the assumption of the World Wide Web to be of dynamic character and a heterogeneous environment provides the change and new directions needed for the above schema to compile and dynamically and naturally change. The added value of understanding and refinement will spur new directions and form others views of knowledge to aspire growth.

The most common criticism of semantics in a computational context is that the representations on the web are supposed to imitate natural languages, as the only way of realising our mental models is through language. The problem is that neither hardware nor software understands the natural dynamics of language. In spite of technological advances made on logical inference of words and contexts, computers cannot truly *understand* the meaning of a word by placing it in the right context, and relate it to logically linked concepts and so on, by themselves. Human beings on the other hand have the ability to rationalise and logically subtract relationships between elements – an ability computers have yet to attain. I will come back to this issue in the discussion on ontologies.

This also brings up the issue of computational power when a complete semantically enriched dataset would demand too much of any system. The amount of interlinked datasets required to provide a complete semantically enriched system, multiply relatively to the original dataset. In the computational sense it means that every original data set of entities has metadata relations to enable the semantic linking. For every single piece of data, there are several multiple linked pieces of data that follow it as semantic enrichment. Making the connections also take up human resources. Many semantic projects are therefore limited problem spaces.

The further discussion is very much based on to what degree we are able to develop a system that understands the meaning of the data that we provide for them. How can we provide a system that provides a contextual framework in which software can make logical closures?

Making sense of data

So how is any semantics enabled in a computational setting? Firstly, in order to understand the dynamics it is necessary to look at the atoms that hold the foundation together. I have mentioned data several times, but in the context of semantic web data becomes the essential building brick. As K.J. McGarry indicates, "The dominant and overriding attribute that emerges from all investigations on perception, cognition, learning and memory is the necessity of ***pattern***" (McGarry 1972). Data provides this pattern of which semantics can be extracted. Words and sentence need form in order to convey meaning, but are we searching for a form or a pattern to incorporate sense in machines? But replacing a pattern does that mean improving a pattern that was insufficient because it was a pattern, with another pattern?

Data are singular objects that together form information for people to interpret and understand. Simply put, one word could constitute *one* piece of data and a sentence constitutes information, but *only if* a person can read and understand the information value. Being able to connect one string of information to another piece of knowledge already in your possession, it can expand your knowledge base if you have acquired the necessary knowledge to logically interconnect them. If you do not know the meaning of one word, a sentence may appear as completely worthless. But, this sentence can become meaningful, if you connect the words to a context you are familiar with. *A semantic relationship realises relationship architecture.* Linking the documents together in a semantic web context implies linking them together in order to enable reasonable context between the *data* that constitute documents.

Thus to a machine, understanding data is being able to logically connect data to each other and place it in a context. By providing frameworks like ontologies computers have reference material to use as logical context and reasoning in order to "understand" how the data can be connected to other relevant data. *A conceptualisation* is a simplified worldview representation of objects; their relations and related objects that exist in or belong to the intended field of conceptualisation. This form for representation is required in every system dealing with

knowledge extraction (Gruber 1993). It ties objects logically together for a software system to be able to make use of the specific information about a subject, but also the background information related to the topic. Humans apply knowledge about relations and objects relative combination and add them together to form representations of realistic views. A conceptualisation can be said to consist of *knowledge representations* in an outlined system. Whilst keywords might appear more accurate than conceptualisations at first, keywords may contain ambiguities that have no context to correct or understand its purpose unlike a concept.

Having to standardise input to a form would help computers "understand" augmented parts of the World Wide Web. The thought of making a computer *understand (make it more clear)* something is probably the most scrutinized part in the realisation of a semantic web. Understanding the background of data exchange between agents to perform advanced tasks is as close as one gets. We can come closer to succeeding in doing so by providing a necessary reference network, conceptualisations and tying strings between objects of our world so that computers can learn from the sites that exist on the web. Today, most computational effort is in displaying findings, links and pages that might be of interest to a user. This is enabled from the HTML in a document, but search engines cannot fully utilise the search results from HTML documents without direct intervention from the user because HTML present the content without processing it. The HTML merely tags the document so it is retrieved because it contained the word you searched for, but it could be in a whole different meaning than the search was intended for.

When we get agents to understand that a user is in fact a person due to his/hers behaviour and identity number and so on, then it can link that person with interests and calendars and act on your behalf. *But we never taught computers what the data actually means which is why background data has become increasingly important to this type of research.* Instead of just a system that search for what people ask them to find, agents can help us actively by also finding related things that we are interested in individually because it understands what the content on a page is and what relevance it has to us. It will find this not only with the help of keywords, but with the help of the semantics in web pages and documents. It does not have to be either one or the other type of search, the semantic search can also be the combination of data that are relationally or logically or contextually linked together. Overall and more fundamentally than just the distinction between semantic or algorithmic searches, all searches are increasingly being based on the concrete data and not documents. In this way it has been

said that we are *moving from a web of documents to a web of data* (Miller 2007) *or a web of things* (Sporny 2008) which computers will understand and be able to interpret and interconnect with.

A suggested solution - a web with more meaning

The notion of a semantic web thus began when a few scientists and engineers saw opportunities for developing a differently structured web. Berners-Lee suggests a semantic web which takes advantage of the strengths of the web that brought its success so far. The fundamental idea of linking needs to stay in focus, though a more computer friendly approach is necessary. Today, it requires a lot from the users to retrieve the correct documents when doing a search. A search engine might find documents – thus recall is high, but the precision is not because the algorithms find words that are matches, not meanings that are relevant. A military enthusiast searching to find information about the tank called "panther" will end up with a result set for the MacOS operative system, a research tool: "Protein ANalysis THrough Evolutionary Relationships", the cat and so on. If thus software would be able to logically deduct information to connect relevant information, it would relieve the user from much work in sorting out the accurate answers. Trying to make the data further accessible is thus a way of making precision rates higher and recall higher – not only for a portion of the information, but ideally for all of the information available on the subject matter.

Several actors have thus realised that providing a foundation on which information can travel on can solve current issues. The consensus on the goal does not imply that the means to get there is the same. The result of an evolution of a better *machine accessible web* could be a semantic web (Antoniou 2004). The W3C works on a package of semantic web technologies to enable a seamless system to truly enable semantics. All in all, most communities seem to work commonly towards some sort of machine accessibility to gain control over large-scale information repositories.

It is also from a computer perspective we hear Berners-Lee with his own words explaining what he means by a semantic web: "The Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it." This statement, on the other hand, supposes words as "learning" and "making sense" of information which is

originally viewed as a human quality. The core of much discussion is on how to retract exactly this quality of meaning. Such a statement naturally leads to several difficult discussions on *the meaning of data and understanding.*

Berners-Lee formulated the idea that the web should exist on one platform with an overarching ontology which would make communication fast, exact, accurate and smooth. Unified systems with concurring ontologies, content and formats could convey the best results from a semantic structure. This would imply that all documents should be indexed in the same way and categorised in the same manner. But the web is a heterogeneous mass of documents and platforms. Rewriting the existing web is naturally not a possibility due to its considerable size. Then again, this major issue has lead to a more practical problem-solving approach where integration between several ontologies was a more appropriate approach to real world situations. you don`t need an over arching ontology or RDF etc to bridge (From a podcast of "from a web of documents to a web of data", 09.10.07)

Even if the technique of several integrated ontologies works, the web of different standards and systems pose another type of problems. Proprietary systems and applications close access of data between the different platforms. Integration between these standards is time consuming, complex and distorts the flow of data exchange. Open standards that everyone can use makes integration less painful. Open standards further minimize or remove issues concerning unified systems, platforms and languages that cause cooperation between different systems to be difficult.

Open standards: pre-requirement for enabling semantic standards

Open standards are easier to build upon and the idea is not new. The World Wide Web was itself built upon an open standard – the Internet. Without its open access, the web, instant messaging and email would not be compatible and work so unnoticeably easy together. Berners-Lee suggests to continue to use open standards to ensure continuous improvement (2007). With the use of open standards it is intended to improve interoperability between technologies and products to a large market. In order to build the flexibility that is needed to bridge the gap that exists now on the web and provide a completely free and open offer there

is a need for software that is provided for in the same manner. *Open Source* is often mentioned in this context.

Open Source is software that addresses the issue of flexibility and interoperability. Open Source is a fundamental change to software development, built by a small number of people, supported by a large number of people in for example a community. It can be downloaded for free from the web and function as building blocks to build your own structures on. Such cooperative functionality provides the foundation required to support interoperability which in turn enables knowledge sharing systems where smooth transition of data is a primary requirement. In addition it enhances equal opportunity for everyone because everyone will have access to it. This exact argument is being used by the schools in Geneva who are replacing all 9000 computers in their schools with open standard Linux (Schreurs 09.04.2008). Then the school children will have access to the same system for free at home as well as at school.

While large multinational enterprises integrate new inventions into their own systems for compatibility, the result is also locks customers to one specific system. Proprietary systems prevents from openness between applications. Licensing and intellectual property rights are of course a part of business strategy and capitalist forces drive most inventions. The distinction however, between vendors who take into use Open Source or those who are closed source is quite diffuse and many vendors have embedded some Open Source into their applications or platforms not according to their consumer's knowledge.

Like I mentioned earlier there are several actors working on standardisation. Open standards' obvious benefits have gained attention due to several problems with *not* having standards. However, firstly there is a distinction to take notice of concerning the practical implications of standardisation and Open Source (Hidas 11.04.2008). Open Source provides a slightly less formal and bureaucratic manner of standardizing and comes across as more flexible and present in the current development. While standardisations are proposed, debated and decided upon in an often long thorough process, Open Source is tested immediately by its users (often with beta versions) and continuously improved in a more iterative process.

The International Organization for Standardization (ISO) develops standards based on expressed needs on a national level which in turn should reflect a global need within a

specific sector or field in order to be revised as a standard (ISO home page). Unfortunately, standardisations is a slow process, meanwhile new applications appear as the need shows, some standardisations disappear before they are properly being taken into use as standardisations. Also, it is utterly important to point out that there is a difference in just standardisations and *open standards* which are in focus in this paper. Like the Open Source Initiative (OSI) I agree that the term is hard to define, but my definition complies with Open Standards Requirement (OSR) defined by OSI, which is that the standard must be accessible to all, compatible and interoperable with all technology - both hardware and software and intentionally support the work on Open Source.

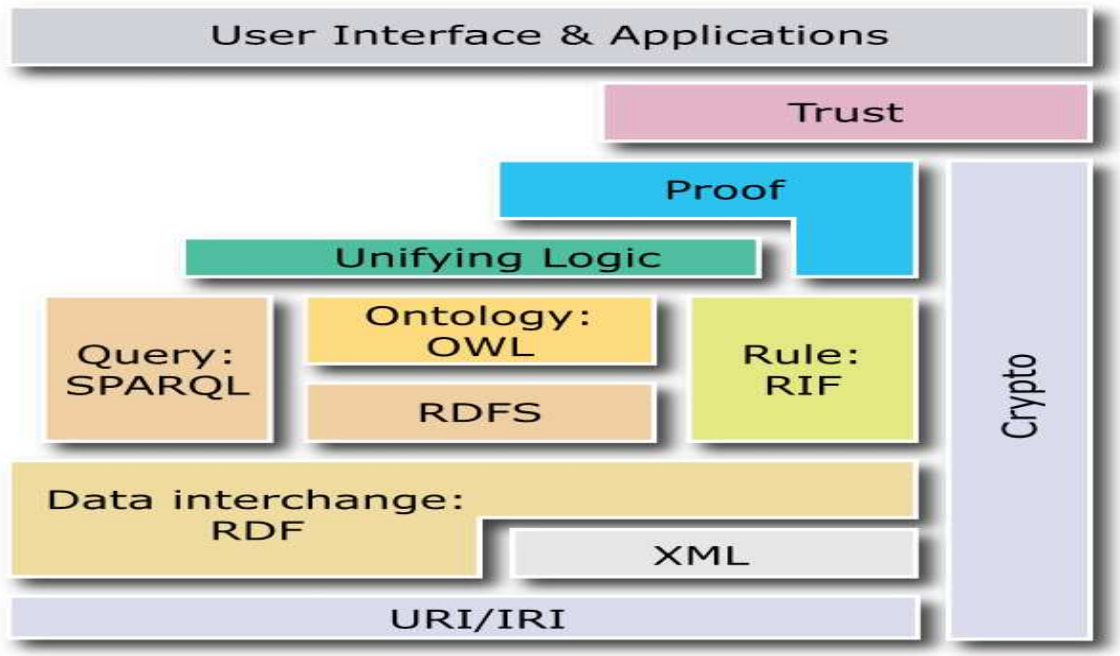
There has been a continuous war between large companies who dominate the market and the smaller-scale developers about locking systems so that they are not accessible or only accessible through certain systems as marketing strategy to prevail dominance in the market. The most recent standardisation of applications comes from Microsoft. The much debated OOXML has not been implemented without any controversy. In fact, many have raised the question to whether one can "force" a standard on users unwillingly, when the product has not been in circulation to be incorporated into systems. This shows the fundamentally different practice between standardisation and Open Source. Open Source is one force who wishes to work against this. The OSI is a non profit "community-recognized body" that support and inform about Open Source as a solution to "better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in" (OSI home page). An application is defined as Open Source if it complies with a set of criteria following along the lines of the latter description. Comparing these to ISO standardisation principles, they share the wish to have voluntary consensus driven solutions that be shared on a global scale. Even while ISO and OSI are working differently towards development support, both share partially the same goal of streamlining platforms, applications and hardware for communication. However, tool support is primary in Open Source when directly implemented and continuously altered, while standardisations often provide the documentation of a need before developing tool support.

Increasingly many organisations and enterprises have taken Open Source into use because of its provided transparency and support for existing applications in complicated structures. Considering the quick development, it is understandable that enterprises, communities and users scream for solutions that immediately respond to their acute problems in the present. W3C also believe in the openness of application to provide seamless communication. It never

releases any recommended standards before the tools are fully developed and have interoperable implementations, in line with Open Source guidelines. Their standardised Open Source applications are widely accepted within many industries and communities. All W3C software is Open Source and has standardised tools to support the semantic web advancements such as OWL (Ontology Web Language), RDF (Resource Description Framework) and XML (Extensible Markup Language).

How applications might provide understanding

This line of software tools, integration and need for interoperability is thus the first step towards a semantic web and it implies to standardise formats for data so that they can be integrated better and be made more sense of to computers (Berners-Lee 1999). He created together with the W3C a model of a semantic web layercake to illustrate how the layers support each other and to visualise the practical technical approach to a semantic web.



Figur 1: The semantic web layercake presented on the W3C web site on semantic web development.

As the discussion of Open Source, knowledge representation and finally usage of ontologies will unfold over the next pages, I regard it useful to show and describe this layer cake so that references to the interconnections of the semantic web will seem easier. Further it gives a helpful overview of the practical existing technologies and principles in question before reviewing the rest of the landscape. This layercake is a revised version of the first proposal,

but is equally based on the same principles. Basic languages make the foundation for more advanced languages to use and the top levels of the layer cake are abstractions that enable the foundation. A top-down structure provides functionality for all layers to reach downward in subordinate layers and access semantics vertically from all lower layers. It enables compatibility, interoperability and unconditional reading of subordinate layers. Since existing semantic web structures are yet incomplete and insufficient and do not coincide with all structures of the semantic web technologies, a partial bottom-up structure enables for less advanced languages to extract semantic content from a language in a higher layer which opens up to a less rigid structure. This adaptability and extension of the existing web opens up for a slow evolutionary development of the semantic web vision.

The first proposed layer of the semantic web is URI's since they are the atoms of the web. URI are the pointers or markers on the web to a certain place on the web and are essential for finding any data on the web. The next basic layer is XML which was originally created for the use on documents and complicates the exchange of data. However, it still has become a commonly used data interchange technology because of its strength of providing structure and customises tags, though not aimed specifically for the semantic web. Because XML has a tree structure, it proposes difficulties for the fusion of documents, but RDF enables this merge quite easily. RDF is made up from triples, meaning a subject, predicate and an object, which defines how one thing is linked to or defined by another. Relations between data and its respective background or context, are connected with the use of these triples.

<masterthesis> <author> <Pia>

Ontology Web Language OWL is one of the accepted framesets for knowledge representation and together with RDF and XML it facilitates the vocabulary and semantics *needed by computers to process content and not only displaying content*. Processing in this context will be elaborated on, but in short it implies extracting meaningful content that is connected or in context to a specific search. Logic must be unified otherwise the system will not concur with itself as double implementations and paradoxical elements might be paired. Logical reasoning techniques decide whether the system is usable in retrieving valuable searches. The trust layer should provide the trust needed from agents between software applications that the data downloaded is in fact correct otherwise the system will not be optimal for information extraction.

Applications or software interacting vertically between the different layers are called intelligent agents. In a simplified view, it substitutes human behaviour in a way that software agents will relieve users of certain activities. There are many visionary applications foreseen for agents many of which are “shopping bot scenarios” where a purchase scenario is adapted and layed-out according to specifications of the buyer. In the humanities, an intelligent agent installed with a language ontology can decide whether information is relevant and coherent with a search performed to find a context or culture specific result or a historian can decide whether information about a person’s identity is in fact valid and in existence according to specification given to an intelligent agent. There have been drawn several parallels between intelligent agents and Artificial Intelligence (AI), but a web that extract meaningful content does not necessarily operate on its own in the same manner as AI can do. Naturally, the semantic web should not be confused with this advanced development even if intelligent agent development has derived from the AI community. Intelligent agents are software programs or documents interacting between computers in a given environment, automating certain tasks and performing on behalf of other agents or people. Envisioning a web supported by agents who shop, deduct knowledge, book appointments or enable for you to line up your bank statements in your personal calendar, might appear as a futuristic mission to embark on. According to Berners-Lee et al. the true power of the semantic web will show when there exists enough semantic content on the web for agents to fully make use of it (Berners-Lee, J. Hendler, and Lassila 2001). On the other hand, as argued by Antoniou (2004) the web will still be of good use even if the intelligent agents work only partially extracting at least some of the information on the web improving the existing web anyhow.

Agents are often connected to the semantic web debate as one enabling factor to make it succeed and able to transfer knowledge between systems based on an agents ability to include context of the piece of information. Hendler (2001) assume the relation between intelligent agents and ontologies as the means of efficiently conveying knowledge transfer in some form of structure that will suite the heterogeneous environment on the web. I will come back to this in the section on ontology development. Passin (2004, 208) argues on the other hand, using Weiss (1999) classification of agent types, that a logic-based agent using ontologies, is not the only alternative in the development of the semantic web and that reactive, belief-desire-intention agents and layered architecture are all possibilities in terms of interaction method between all agents. However, we are arguably moving towards standardised tools that has given logic and deduction a solid head start due to its simplicity and applicability on the web

and Passin admits that the logic-based agents probably will suit the current development best. From the semantic web layer cake, layers of proof and logic prove consistency of the rules that are given in an ontology axiom and in interaction with other agents' rules and proof. The logic of the rule set decides whether a site is trustworthy and trust in the top layer is an authentication process of interacting agents and other sources on the web.

A solution such as the above scenario of the semantic web layer cake and intelligent agents is to some extent in use today, but it requires a certain level of standardisation through schemas and filters. According to (Decker et al. 2000) *the semantic web needs interoperability between syntactic layers for any semantic content to be accessible*. This is an important note – because it explains not only the importance of easily integrated layers and interoperability with all systems – but it indirectly implies that semantics are retrievable *only* on top of this context. Meaningful content or knowledge is retrieved on a foundation. It needs in addition a certain foundation on which to aspire and become visible for retrieval. On the other hand, to what sense can actually a new structure of tighter and new connections provide sense-making? Adding a new structure will logically not necessarily lead to better understanding. The current syntactic layers are structures, but semantic structuring is also a new form for just that – structuring. Will another type of structure provide any more meaning?

Concluding points of Part 1

To conclude up to this point, I have now discussed the semantic web evolution from its early seeds, issues of semantics in context, basic computational understanding and standardisations to provide a basis for understanding the semantic web's complicated structure and alternative implications. I have in the most recent chapter outlined the W3C web layercake to present more concretely the complexity of systems and interoperability of technologies that are visualised to solve the problematic issues of meaning, knowledge and understanding.

The semantic web's most significant ideas are as we have seen universality, extensibility and scalability, which should be respectively; including all forms of documents, formats, media, cultures and languages without discriminating between them; extend the system to be all inclusive to new formats, applications and systems and last but not least; to be able to treat enormously large scales of this heterogeneous and continuously expanding, material. Still,

data has shown its limitations when it comes to conveying meaning and in spite of an approach towards all-inclusiveness and the reduction of data stream resistance by utilising open standards, the debate seems to return to this issue of meaning and knowledge. These requirements of the W3C standardised semantic web vision depends on maximum integration of systems, cooperation and strict systems that streamline data to enable knowledge transfer on a computational level.

The machine-processable approach is actualised by simplifying data processes to the extent that applications can with less effort access valuable data, and its respectively linked knowledge or connections of metadata or triplets, in order to perform tasks. In a somewhat simplified way, it is an attempt to hinder that data is "lost in translation" between application layers and conversions between systems that cannot fully understand or properly access each other.

At the same time, the semantic web is an attempt to become more accessible to people in terms of production, cultivation and attainment of knowledge. After an examination of the constraints of data in comparison to linguistic semantics it is natural to conclude that in the latter semantic sense, understanding is primarily a human quality. Interpretation and contextual understanding are also established as necessities of meaning or semantics in this argumentative line. This complies with the user oriented view that also assumes a less autonomous way of understanding the semantic web, as this view's imperative is also some form of setting, environment or context.

In the next and second part of this thesis I still wish to focus on the pursuit of knowledge and sense-making of data, but from a theoretical perspective that addresses fundamental issues on knowledge acquisition, sharing and preservation. Till now I have focused on what a semantically structured web looks like in a web context. The notion of sharing knowledge and extracting meaning from information reaches further than that web context. Arguably the social knowledge perspective has had impact on a societal understanding and organisation in the second half of the century. This issue has flourished within the semantic web context as it invites to debate several issues. To what extent is a semantic web needed to extract meaning? What drives this development? What does such a view presuppose?

Part 2

Knowledge perspectives and semantic web

I left out the surrounding historical and societal dimensions of a semantic web till now on purpose. In the following section I wish to consider these dimensions before taking a closer look at some practical applications and developments of semantic web tools. Let us picture a semantic landscape that fulfils all requirements of a semantic web. If the notion of collective construction and social organisation is assumed on the web, then these ideas have significant implications. I argue that central theoretical aspects are lost in the technical specifications in production. This seems natural due to technological limitations, but if perspective is lost, a semantic web project might not resemble the desired result and the semantics opted for in the semantic web – becomes non-existing. Therefore larger perspectives should not be discarded too easily as impossible to reach.

I introduced the contrast of autonomy thinking and the contextual environmental perspective of sense making. In this introduction I also clarified my standing point from the latter functional perspective as the foundation for this thesis. A functional perspective implies a relational and relative view on an object accordingly to a context or environment and this is why I have chosen to include a social aspect and a historical angle to knowledge. As I also previously mentioned, a general discussion on knowledge will be of no use here.

The quote of Tim Berners-Lee saying that the web will be a “pool of knowledge” will be interpreted in light of the autonomy discussion to see whether this has some social implications. What type of knowledge can be represented and presented to a user with a semantic web structure? How will knowledge increase if the web gives more meaning? If meaning helps retrieval will knowledge be accumulated when consumed or given back to the web after being used in application? There are many questions and little clarity on how to address the concepts of meaning or knowledge in the semantic web landscape. I will not attempt to answer these questions, but merely shed some light on these intricate issues.

Another reason to include these aspects in this thesis is the question of whether Tim Berners-Lee´s semantic web depends on knowledge as a socially constructed asset on the web? First it is clarifying to present what I mean by social organising as a context to semantics.

Implied social organisation and distribution in the semantic web

As I argue that the semantic web cannot stand autonomously in this section I argue that *for a semantic structure of content to be useful, it operates under some social construction*. That this construction is not immediately visible does not equal its non-existence. On the web these social constructions are usually expressed as everything from virtual organisations, mental organisations, distributed communities, learning cooperation, working collaborations, peer networks, expert cultures, distributed organisations and so on. Even though they all differ from each other, they also share a common feature of some social organising around a field of interest. All based on the same principles of social organising of people with common interests or background gathered around the use of intelligent technology to overcome large amounts of information and solve common issues relevant to the group. Books have been written on the subjects in order to classify these terms and since I am not an organisational theorist I will only briefly identify a definition of how a social organisation is understood in this thesis.

Community derives from "communio" and is a form of social organisation that is based on an abstract level of shared ideals. An example of such a unity sharing a spiritual view is a group of monks. Few things are restricted physically as mental models and personal connections are the group's purpose. Another example is the Open Source community who is a group of people who believe in the use and power of Open Source development and solutions and have joined forces to support what they believe in. A peer-to-peer network has usually the commonality of interest and discipline. Peers review the work of other peers or discuss issues that concern the circle of peers. Within a peer-to-peer network the intention is for like-minded parties to come together to encourage, dispute or oppose the group of equal members who belong to the same field or discipline. An organisation is usually based on structural limitations like hierarchy, work tasks, rules and position. It is more rigid, has slower changing processes and can consist of a multitude of people with different interests or disciplines. As knowledge in the organisation presents as a product, the competition opts for the organisation to be flexible and dynamic to keep up with the pace of the market. Even though a static frame is hard to change quickly, the inside, the people within the organisation, is dynamic, which create a need for dynamic and flexible systems. The web's scalability and adaptability to expansion and heterogeneity of information and technology is quite suitable to the needs of an organisations search to satisfy those needs. Unlike the rest of these social forms; organisations

provide a firm structure that can refine knowledge in a specific direction, unlike the web that constantly opens up to new directions.

The social form is what I wish to keep focus on in this section, not the specific terms. In a specific environment an ontology provides conceptualisations that mostly only contribute to the knowledge base of those already familiar with the essence of the content. Otherwise one can assume that the ontology does not function as a contribution to any knowledge increase. Semantic web and its technologies such as ontologies and RDF that are designed to provide a foundation to make use of and connect data, are thus perhaps incarnations of various forms of these social constructions. Peer networks are the main focus here since the Discovery Project reaches out to the community of philosophers as peers.

The web is accessible from any given location, which presupposes a natural distribution of its users. Current phenomenon as web 2.0 and social networking has given people a space where they can come together to distribute, exchange, engage and participate in any desired activity. The trends of folksonomies, wikis and blogs have been adopted outside the web in enterprises as well. The attendance seen in these applications is so high, has shown new uses in alternative solutions that attend to knowledge management and workflow processes. Comparatively, the Open Source community is also based on this social aspect of sharing and contributing where a collective effort to spread personal knowledge and experience to interested parties is emphasised. No differently it is assumed that users of a semantic web will act within such a social construction where knowledge is produced, shared and consumed willingly across a distributed network. However, this does not in any way assume that everyone who is globally linked have equally distributed knowledge sharing among all the users of the web. On the contrary they are largely organised in smaller organisations, peer networks or communities where specialised knowledge is practiced.

Widespread implementation of and implications for social organisation and distributed cooperation is an assumption in the continuous discussion. It is also a natural inclusion due to its widespread adoption and influence on a global scale. It is also included as a discussion of the *e*Content*plus* programme which the Discovery project is part of, where the approach to improve accessibility, usability and exploitability is fronted to “tackle organisational barriers and promote take up of leading-edge technical solutions to improve accessibility and usability of digital material in a multilingual environment”. In the next paragraphs I shift focus from

treating the semantic web on its own, to the functional purpose of the semantic web, which includes the surrounding settings and implications of knowledge sharing under the assumption that people act as recipients, users, contributors and representatives of the semantic web.

As I would now like to embark on this discussion, I underline that this discussion concern knowledge on two levels. *In the broadest sense* the discussion is about the *idea* of clustering knowledge from people with the help of technology. What does a sharing assumption of knowledge imply to the practical development of a semantic web? I will use ideas from both academia and organisational contexts to shed a few different perspectives on this highly complicated theme. *On another micro level* we are concerned with capturing knowledge in technology in the most convenient, accurate and practical manner. This is a focus on knowledge as something formal and engineerable which is later in this paper discussed in the section on ontologies.

Complexities of a social approach – societal influences?

If we are to assume an idea of the knowledge society that encourages principles of sharing knowledge, what does that imply on society that this type of knowledge presents? To reformulate the question more accurately; what type of knowledge are we talking about? It can be more useful to take into use the more accurate term of transaction from the social anthropological discipline as introduced by Barth. It presupposes that people perform transactions across incorporated symbols in societies.

I turn to other disciplines that are perhaps not normally viewed as a part of humanities, but humanities cannot either escape the influence of for example economy as a setting and its impact on human culture. Transaction and interaction opens up to an economic perspective in a broad sense which is interesting when considering that the man who introduced the term knowledge society viewed economy as central to his theory. The business organisation theorist Peter Drucker coined the term *Knowledge Society* and simultaneously illuminated a path for future world economy (Drucker 1993). Since this increasingly began being seen as the new direction of the future, it seems to have been taken for granted that this is what we need technology to support us in – the quest for knowledge. Containing it in enterprises, sharing it to spread knowledge to prevent poverty and ignorance (Varis, Utsumi, and Klemm

2003; Matsuura 2005), or engineer technology for knowledge transfer in high-tech computer systems. The examples of knowledge application are derivations of a knowledge society, yet all of them are very different.

Drucker speaks of “knowledges” that are specialised and based on formal learning to begin with. Knowledge on the web is highly informal at most times, but the web functions as an extension or supplement to knowledge acquisition rather than a replacement of formal educational systems. Drucker seemingly further restricts the range of “knowledges” because “In the knowledge society knowledge for the most part exists only in application” (Drucker 1994). A philosopher does not know how to fix a computer or a car because he does not possess the knowledge to do so. “In application” means that knowledge in the possession of one person does not necessarily have any value to another person. An implication of this again is according to Drucker that knowledge workers need to work in teams and need to be a part of a unity in the form of an organisation. Only then can we use each other’s knowledge to solve issues in an efficient manner. On the other hand, connecting knowledge to application also opens up for the need of a context and this emulates situation dependent knowledge. Knowledge is not locked to one singular application so to speak. Application is therefore highly related to knowledge in context. It constitutes a variable factor that changes or limits the semantic interpretation of an object depending on the situation or perspective it is placed in. The required knowledge in the knowledge society is more specialised. As for the web, specialisation and engagement in discipline specific areas are privileges of the web as its nature embrace any content without restrictions (within legal boundaries). The web encourages for new directions to aspire and specialisations to take form. This specialisation sometimes forms smaller societies of experts or peers that share this specialised knowledge and help each other to increase the knowledge level further. Different expert communities come together to complement each other in solving tasks that require an interdisciplinary or multidisciplinary angle.

The knowledge society of Drucker will require the highest level of performance and productivity. He also lays a great deal of weight on the organisation as its function “is to make knowledges productive...” (Drucker 1994). Drucker sees the change as an inclusive to all parts of society, but not as a final solution. Scardamalia emphasize in the light of Stehr’s theory on knowledge societies (Scardamalia and Bereiter 2006) a larger collective effort to combine the advancement of knowledge levels in society. Scardamalia continues to say that this

knowledge advancement is “essential for social progress of all kinds” (2006). Considering knowledge as a means to achieve a better understanding of connections and consequently make progress in society is necessary for a society to move forward. Advancing forward and building new knowledge is considered generally a positive idea, but how to do it and access usable knowledge is left with several answers.

> The growth of the knowledge society depends on the production of new knowledge, on its transmission through education and training, its dissemination through information and communication technologies, and on its use through new industrial processes or services.
>
> (Reding 2003, 28)

In this framework, knowledge can be understood as an asset that gains competitive value and knowledge in application only for non-specific purposes. Productive knowledge can thus be understood as knowledge that is aimed to reach a specific goal. It does not necessarily lead exclusively to economic profit, but can lead to enlightenment as well. Does this mean that knowledge only grows within a structure? Agreeing with this comes naturally as an extension to the early statement of not treating the semantic web as autonomous. If everything has to have a context to be interpreted, then the production of knowledge should flourish within a setting where specific knowledge is meant to be explored. Does is mean that the more structure added, leads to more productive knowledge?

Productive knowledge grows within structure?

Previously I have said how heterogeneous, unpredictable and universal the web is. The lack of structure is something the semantic web hopes to change on a fundamental programming level. Still, the web’s form allows for searching for nearly anything in the same coincidental way the mind works. In fact, the web will be structured to enable it even better as it opts for contextual and relevant searches that better imitates human ways. As knowledge levels increase and thoughts of “life long learning” and “ubiquitous learning” are introduced as an extension to formal learning the knowledge acquisition setting pushes the construction and refinement of knowledge outside of educational institutions.

> In the knowledge society, clearly, more and more knowledge, and especially advanced knowledge, will be acquired well past the age of formal schooling and increasingly, perhaps, through educational processes that do not center on the traditional school.

Here, it is understood that knowledge is attained through education only to be applied in a work performance situation where it continues to grow. Thus emulating knowledge happens in practice and is a process that is in continuous transaction between a person and that person's surrounding. Connecting this to the position of the web - knowledge is accumulated and shared by people every day even if there is no strict frame to enable this process. However, there is always some contextual frame present but the level of or significance of the knowledge acquired is most likely subject to a different discussion.

Using the "unstructured" web to make people knowledgeable?

Scardamalia (2006) expresses the need for better incorporating people into society where the web is used as a knowledge provider:

> In this context, the Internet becomes more than a desktop library and a rapid mail-delivery system. It becomes the first realistic means for students to connect with civilization-wide knowledge building and to make their classroom work a part of it.

Here, the web is a realistic tool to provide a fitting framework for knowledge building as she writes, and this tool provides the connection to a civilisation in which a person can have a context to the acquisition process. This does not only apply to students. It is transferable to all parts of the society that exceed the frame of mere education. The web is a place where people can unfold, group in familiar interests, specialise, elaborate on subjects and it consists of a multitude in possibilities and variety of human culture, which coincides with the fragmented multitude of heterogeneous nature of the web. Even if there is chaos and seemingly anarchy on the web, the smaller groups of people constitute orderly proportions as well which knowledge can prosper within.

That organisations function to make "knowledge productive" is Drucker's opinion, but this makes sense when talking about social and distributed constructions as foundations to perform knowledge transactions across. Drucker speaks of organisations, but on the web and in the Discovery Project the terms distributed community and peer networks are most appropriate. Lewis' uses, as titled, the activity theory to explore distributed communities, but distinguishes firstly a possible knowledge creation model for this to seem feasible (Lewis 1997). A criterion for knowledge acquisition is that it is fundamentally constructed in interaction with others for it to be logically deductible ((Lewis 1997)).

Quality of content - knowledge

As the point to this section was to elaborate on the type of knowledge desired accessible in semantic web projects, there was one perspective repeatedly left unnoticed at the Semantic Days conference; there is not so much a dispute, as there is given only a minimal of notice to *the quality* of content provided in these knowledge systems. Content causes practical difficulties when it is reused as it is seldom reformatted to suit the purpose of this new system; instead it is kept in its original state, as it existed in the old system. Now, this presents issues as to whether the information is updated, wrong, gone or distributed or linked with the wrong content. Also, if larger dimensions to knowledge acquisition are avoided in this context too, the "pool of knowledge" will quickly seem like "another pool of information". Specifically important in a specialised expert culture where an increase of knowledge levels is desired.

Knowledge has become a product in today's society and has become a valuable asset in many businesses. Extracting and nurturing knowledgeable employees' understanding of the core business could be the primary competitive advantage to many businesses. As other costs and economic margins are minimal or eliminated the personnel becomes the new capital of which business can expand further with. Does the value or quality of the knowledge on the web have anything to say in the development of a semantic web project? What implications does it have on the users?

Knowledge sharing as a concept poses problems in its combination of words. Some phrase their beliefs in sharing as the means to become enlightened and that knowledge is what we share with each other. What I believe is a more accurate definition is to say that when information is shared, it needs applied context and knowledge in order to become increased knowledge in a collective effort. One can build knowledge by oneself with firsthand experience, but only in exchange with others can one reach levels of significant knowledge acquisition to efficiently move the knowledge on a subject forward in the speed it is required in an organisational context.

There are four angles of which an assumption of knowledge construction can be viewed from. One descends from the ideals of the enlightenment period where knowledge was viewed as a shared good of society. The second can be identified as a political outbreak from societal institutions. This counter culture to elitist structures in society sees knowledge as a liberating

value. Knowledge is viewed as an asset that should belong to all and not primarily be part of the knowledgably elite of society (Varis, Utsumi, and Klemm 2003). The third angle to knowledge sharing is a strict understanding of knowledge as created in construction such as the discipline of constructivism in pedagogy represents. Jonassen writes that "knowledge construction [is] context-specific" (Jonassen 1994, 37). The fourth and last angle is a new liberalistic view on knowledge. It treats knowledge in the new economy where immaterial values are higher recognised than material goods. Knowledge is viewed as a personal property in the extension and natural consequence of a gift economy where basic needs are covered. It reflects a knowledge society where transactions evolve around the intellect and not products and material goods to that same degree as previously.

In light of the presented views above, I accept the implications of constructivism, without specifically using the theory or elaborating on it.

Can the semantic web thrive without the assumption of socially constructed knowledge?

The semantic web of Tim Berners-Lee can fit into all four categories of knowledge construction approaches. It is based on a sharing culture that willingly wishes to add, share and refine information on the web for the better of the whole web. It also assumes that transactions across the web will contribute to an individuals knowledge levels. In interaction with others, that individual is thought to construct a concept and adapt it to the general framework of others as the web allows an individual to check how this new piece of information can fit into a personal world view as well as a general world view.

A user attains new pieces of information and adds them to their own knowledge to ensure it coincides with the general understanding of the general opinion. It implies that the learning process is personal and is attained on already acquired knowledge, and consequently constructed by the individual. Still, the information on the web is created in a social context where anyone and everyone are contributors. It does not imply that everyone have individual views of the world, which is a common criticism. Personal views are rather adjusted to what other people generally tend to think because some pieces of information are indisputable – such as gravity or the line of ruling presidents throughout history.

As I have previously outlined, there is an understanding of the semantic web as a provider of knowledge quality of the web: universality, flexibility, understood as simplified searches that provide result sets that are context related and more relevant than. By user annotation and ontology bulding, people odnt have to keep searching the same single pieces of information and set them together but can be presented with related strings of information that will provide knowledge about a subject. a scenario that is not possible today.

Sharing of meaning between experts does not lead any way in general, but in practice it has some implications. The social organising gives itself its context in its construction, thus in a specialised group the participants of the group gives itself the context – others outside it would not be able to understand it, because they are not a part of the context.

To shed light on whether Tim Berners-Lee's semantic web vision is in fact dependent on knowledge constructivism and social organising or not can be summed up in another question; could there be a semantic web, a tagging community, content providers of ontological definitions, if knowledge acquisition and consumption is a cognitive process?

Semantic exchange in a public sphere – producers as well as users are participators to enrich, criticise and refine the content

Social knowledge construction has to take place somewhere. As have been discussed in this section, even specialised knowledge has to conform to a general opinion of the world. Producing knowledge systems implies having some form of constructive way of criticising its content for it to be productive. On the semantic web, a result set should consist of relevantly connected texts. This purpose of linking meaning together is to refine the available content. Besides the producers of semantic enrichment, deciding what is meaningful or not in context with each other is often connected to user involvement. On the semantic web, users are performing transactions and exchanging information and are together in a constant redefining process about the content. Many users are also producers and the semantic web is a forum where conversations and construction of meanings are their placements in different contexts are taking form. In an organisational setting for example the content and knowledge exchange environment is created to e.g. collect, organise and retrieve the knowledge of the company's employees and perhaps improve efficiency and create meaningful knowledge exchange that

can contribute to a refinement of the knowledge one has and possibly change and improve the dynamics of the organisation.

Exchanging ideas and knowledge in a certain forum to obtain a conversational circle in which everyone contributes equally has been described before the arrival of the web. The "public sphere" of Habermas is based on the principle of democratic discussion to take place in a public sphere to enhance the cultural and political flora in a society. Media works as the catalyst of this process (Habermas 1960). The public sphere constitutes everything from a speaker at the market place to books and newspapers where a publicly accessible discussion takes place. This way, intellectual voices are displayed and debated to influence decision makers in a constructive and active manner. Knowledge exchange though is viewed as a societal act, performed with respect for the enabling factors of a functional democracy. These enabling factors for a productive conversation are among others approaching the debate with critical arguments and having respect for all voices as equal.

Knowledge exchange in the media has been altered as e.g. the Internet has phased out the distinctive lines between a public and a private sphere. In a recent speech on the Internet and the new public sphere, Habermas says (loosely translated and quoted) that the Internet has expanded the surface on which to perform public discussion, but has also become a clustering of communication patterns and networks. When the public sphere becomes less formalised, the roles of the communicators become increasingly similar, contributions to the discussion are unedited and access to the contributions are decentralised. It has lead to an intensification of opinion exchange where not just the expert or the classical intellectual shares their knowledge with the masses, but they are being held accountable to their knowledge and perspectives by the general public who are consisting of potential speakers with intellectual argumentation (Hansen 2006).

Linking it back to meaningful exchange and knowledge extraction from the web – this trend based on Habermas view is not so much an organisational or engineering push technology as access to the public debate has become a public phenomenon, privilege and a must. The lack of formalised, centralised and organised channels for interaction though, has intervened with constructive and efficient communication between different parties. It has perhaps lead to a larger amount of clustered networks and not opened up so much for "everyone" to participate. The "intellectual" is perhaps the creator of a semantic structure. This way the intellectual has

risen above the mass to organise the knowledge for the mass to digest. Who is he (or them) who choose what is of importance and what is not? This is an important issue to transfer to the semantic web and the idea of conceptualisation. It arises as central to the discussion and has important implications and notions that one should be aware of.

Knowledge/power and user trust

There are some issues concerning the qualities of knowledge if it is to be distributed or shared amongst people. What will make people contribute to a common knowledge pool? In a knowledge society where specialised knowledge means competitiveness, knowledge provides power. Foucault's term knowledge/power represents the co-dependency between the two terms. "It is not possible for power to be exercised without knowledge, it is impossible for knowledge not to engender power" (Mills 2003, 69). A pursuit of knowledge is also a pursuit of power, Foucault continues. Increasing the knowledge levels by sharing knowledge on the web, could seem like an unlikely scenario in this view. Sharing knowledge would lead to a loss of power if others were to be in possession of your knowledge. It could also be interpreted as a reason to social unbalance discarding an egalitarian view to the benefit of an elitist view. The powerful possess knowledge that in turn elicits, if any, what knowledge should be distributed back to others. However, Foucault reasons for a different approach. He expresses that *knowledge production is a collective effort*. Those who possess knowledge will seek to produce knowledge about marginalised groups that possess no knowledge. This implies that marginalised groups will first attain knowledge about themselves, *then* begin to produce knowledge themselves which in turn will adjust the knowledge balance. Sharing will raise the general knowledge level and has therefore an indirect value to those who share something because they will get something back.

Equally important, knowledge/power raises the issue of who people represent when they participate in a meaningful exchange on the web? Will a web user contribute to the debate or share their personal knowledge of self-interest or for the benefit of the public well being? Whether this knowledge exchange on the web today occurs with people acting as public or private persons on the web I cannot answer. But it raises a question about people's will to contribute – can one address either the private or public interests on the web?

What we can do with limited knowledge – measures taken to approach real world requirements

As I have demonstrated discussing autonomy, the connection between knowledge and the semantic web appears to be opposed by a practical pragmatism. Practicality is in need of some form of further explanation. On the web, as I outlined in the above sections, there is a process of moving, building, collecting and retrieving information between both people and technologies. Approaches taken to these processes have to a variable degree been focused on practical application value in real world settings. Practical application of theory means managing solvable issues using technology at hand. Often there are posed questions to matters that require, what is till now, non–existing technology. This creates an impossible situation for those in need of immediate and usable solutions, and there is a need for usable and practically available solutions within semantic web development.

Knowledge representation, knowledge transfer, knowledge sharing- dealing with the problems of knowledge promises

In the envisioned discussion of knowledge and the semantic web, a general categorisation of knowledge does not lead to a constructive debate. Dealing with the problems of knowledge settings has challenging notions. There is potential in the knowledge debate, but the approaches taken to the subject matter varies to such an extent it reflects some of the fundamental problematic issues of the semantic web in itself. I refer to the issue of appropriate perspective and setting in the approaches taken to knowledge on the semantic web. Subsequently, I do not wish to present deep philosophical truths on what knowledge is, but I do think it is necessary to reflect on how knowledge on the semantic web will serve us.

> The Semantic Web is what we will get if we perform the same globalization process to Knowledge Representation that the Web initially did to Hypertext. We remove the centralized concepts of absolute truth, total knowledge, and total provability, and see what we can do with limited knowledge.
>
> (Berners-Lee 1998)

This quote is a pragmatic and technical perspective on what we are able to extract from the web. With this quote it is also obvious that the semantic web is not meant to be a utopia and that Tim Berners-Lee's intention is not to construct a knowledge representation system that reaches for "total knowledge" as he says. However, even limited knowledge is still knowledge.

From this view, what you can do with formal and engineerable knowledge is the starting point of a web that can be called semantic.

Variants of knowledge transfer and sharing seems to be shared commonality between the concept of the semantic web and its practical approaches, though knowledge is represented in significantly different manners and manifesting itself on significantly different levels of perspective. Knowledge promises on several levels are expressed on a technical level as knowledge representation, in organisation as knowledge management and idealistically as knowledge sharing whereas all levels are part of this widely debated issue. The discussion separates in a distinction between engineerable and formal knowledge and a notion of knowledge as a quality restricted to human beings. The sharing concept might result from a paradigm shift in society; however, what are the precise components that constitute this change depends on what one opts for. As knowledge is an interdisciplinary field that intersects with most parts of human culture and nature, there are several perspectives that can influence its position on the semantic web as well. Knowledge on the semantic web can be seen as an extended cultural, sociological, organisational, economical or technological phenomenon to mention some. I will not explore all of them in depth as they are not expressively relevant to this context, and more importantly, as disciplines in depth they lay outside of what is my discipline of humanistic informatics. I still wish for the reader to consider these disciplines as aspects, which a humanities perspective on technological development requires and is concerned with to gain a higher level insight than that of pure programming and technical specifications.

As I introduced in a much earlier section, on page 24, conceptualisations are used to express a limited worldview. Ontologies are such realisations of conceptualisations and are only representations of knowledge. Making representations of human knowledge and imitating human behaviour in computers seem to face the same obstacle of misconceptions of understanding, reasoning and logically relating something. Also, it can lead to unwanted trust and elitist issues if producers are to decide how objects are related, thus providing an already organised worldview in behalf of the users. *It seems therefore reasonable to demand a limitation to the idea of what a semantic web can actually do.*

Firstly, one angle of approach can be to identify the knowledge that is desired to have represented and then look at how it can be represented. It may sound simple enough, but if not

carefully taken into consideration the result can end up as a regular information database in which the application value is lost. We make representations for computers to better "understand" us, but ultimately it is the users who have to interpret the "understood" or processed material. Again, who is this system created for? Like Drucker seems to say that knowledge is visible only in its application so is also the most approachable view to presume from an engineering point of view. When knowledge is applied to projects it becomes visible in its use, but stored on the web its value is less representative. Unfortunately it leaves the engineer with questions such as – are we truly capturing the knowledge? Will this construction continuously contribute to build the knowledge base of the organisation? Sharing experiences and information for applied use builds the knowledge base of the respective community and can partake in the heightened knowledge of the participants. From the discussion on semantic web as an enabler of processing information and not only gathering it, and the general trend of interacting in social identifying groups I was pondered by two questions:

- Are these trends of open source, knowledge sharing and semantically processable software a reflection of societal movements about distributed cooperation?
- Have these trends been enabled by the web or been made visible by the web?

Before heading into the more concrete, practical and technical part of representing knowledge with the use of ontologies, I think these two questions are important not to answer, but to use as a summary of some of the reflections made about the knowledge topic discussed in this latter section. Both intersect with multidisciplinary fields on many levels adding to the complexity of an attempt to answer them. I think they are significant questions that reflect upon a humanities perspective in order to attempt to understand the purpose of a semantic web and consequently to understand its usage value.

Conceptualising the world on behalf of others?

Now, this section of the thesis is not meant to outline technical designs and structures of ontologies. Instead I wish to focus on the issues the idea of an ontology creates. As an introduction to ontologies is required, I will briefly outline its basic form, and some interpretations of its use before I attempt to look at a few issues in connection with the construction of knowledge. Ontologies are utilised as a part of the technologies on the W3C's semantic web as a knowledge layer, which insufficiently is supported by the web architecture

that is in use today. Ontologies gather information and searches interconnections between pieces of information with similar texts holding the same, similar or related information. The methods from which the ontologies are created are partially conceived from earlier database management systems construction and Artificial Intelligence, but it also consists of methods derived from philosophy on how to build theories (Smith 2004). Ontologies on the semantic web are based on conceptualisation that I mentioned earlier (on page 35). Conceptualisations are encoded as ontologies using vocabularies that show connections between the concepts. More familiar is Gruber's definition of an ontology as an "explicit specification of a conceptualization" (Gruber 1993). This construction however, implies that the concepts must be somewhat agreed upon for it to be useful in a community of experts for example. If those concepts can't be agreed upon, how will it be possible to build a network of those concepts?

To paraphrase Mädche, ontologies provide the information extraction tool needed to collect various sources together in one knowledge source so that the user is presented with a clear selection of relevant sosrces that coincide with what they envisioned (Mädche, Staab, and Studer 1999). The diagram illustrates how different sources of material can be merged in an ontology and then be processed before passed on as structured and coherent information. More technically, ontologies are definitions of how objects are related to each other and what classifies one objects relation to another (Gruber 1993) and how these rules given to the relationships "enforce a well-defined semantics on the conceptualisation" (Mädche, Staab, and Studer 1999). These rules are defined as axioms and these axioms define one objects relation to another based on correlating relationship descriptions.

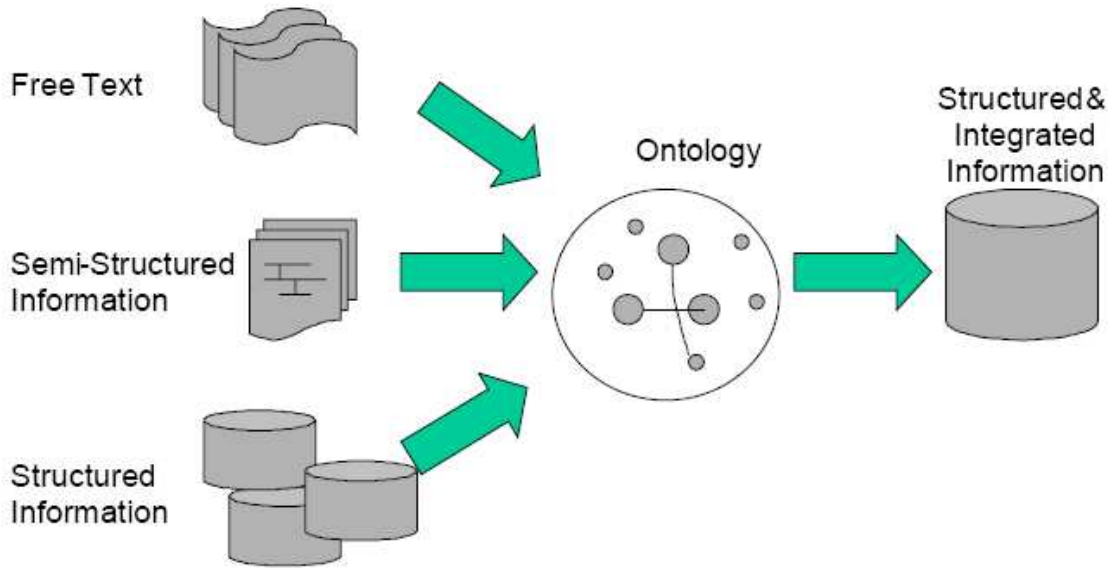


Figure from (Mädche, Staab, and Studer 1999)

Domain ontologies provide the information needed for the domain in question while formal ontologies solve integration issues when communicating between domains that deal with different information sources. Like illustrated in the above diagram, and in accordance to the semantic web vision as mentioned, an ontology should process information and not just merely present it. It is a knowledge representation system, where data is connected to enable a further "understanding" of information in context.

More on the definitions and ideas of ontologies need to be elaborated on though in order to fully understand its complexity and the opposite communities of supporters. Decker says ontologies generally are defined as a "representation of a shared conceptualisation of a particular domain" (Decker et al. 2000). Ontologies can be viewed as large conceptualisations that require them all to be collaborating, yet there are mostly being built minimal ontologies covering small and closed information clusters. The reason for this is that ambitious ontologies are costly and require a long time commitment and with the technology changing so fast as well as opinions on best practice change continuously, these mega projects belong in a minority, but are increasingly visible in the landscape (Cyc, CIDOC-CRM, Europeana).

Others such as Hendler (2001, 30-31) envision a web consisting of many ontologies built of the *same fragmented and coincidental structures as the web we have today*. He sees several small ontologies pointing to each other. Such merging of ontologies produces another set of

problems with ontologies. The ontology community has in general moved away from the notion of an over arching ontology, and is mostly dealing with smaller sets of ontologies like Hendler envisions. To make this suggestion work properly these ontologies need to be able to merge in order to cooperate functionally. This could however result in chaos as the understanding of ontologies as well as the interpretation of idioms, words and other definitions might vary to the extent where the same concept might be categorised completely different in different ontologies.

The term ontology is multi faceted and was initially derived from philosophy to later be adopted as a term within Artificial Intelligence. From there several other fields within computer science have taken the term into use, but with slight variations in purpose and meaning. It is important though to remark that ontologies are not artificial intelligence and do not operate on their own in such a manner. In the opposite end are those who doubt the use of ontologies in the first place, dismissing them as anything else but refined databases. This has been the largest and still is a vivid debate of current research on ontologies.

The latter distinction between databases and ontologies was clarified during an oral presentation during "Semantic Days 08" by Ian Horrocks of the Oxford University. By comparing the two he presented clear distinctive features of ontologies: an open-world assumption, no unique identity assumption and the ontology axioms behaving as supplementary descriptions of relationships. In an ontology the result sets are based on both the axioms to provide logical theorem as well as the content, contrary to the schema in a database. This database schema only provides constraints and has no role in the query process as such. it means that ontologies can process incomplete information and present result sets including both intended as well as extended answers. Since two objects can have identical names, an ontology that has no unique identity assumption processes this information and presents it as two options to which the user has to decide whether they wanted to know more about; "Frank Peter" or "Frank C. Peter" – who lived in the same city around the same time. Due to the open world assumption an answer can express that "there are at least two of something, but there might be more".

Smith points out that "Different databases may use identical labels but with different meanings; this is polysemy and ambiguity alternatively the same meaning may be expressed via different names, this is polynymy" (Smith 2004, 6). Ontologies on the other hand are not

bound to any single created identifier and so “a meaning” is not tied to an “artefact” that is needed to create a model for information, but instead an ontology process one meaning in connection to the domain without restricting it to its encoded place in a model (Gruber 2008). “Nonetheless, the definitions and formal constraints of the ontology do put restrictions on what can be *meaningfully* stated in this language (…) [A statement has to] be *logically consistent* with the definitions and constraints of the ontology” (Gruber 2008). Still, ontologies that express relationships can run into more severe issues than polynymy. This next example illustrates one semantic (and moral?) problem with conceptualising entities and their relationships. A logically functioning ontology on history might contain a domain view of race, e.g. [race->superior->ARYANS, race->inferior->JEWS, genetic predisposition->race]). Now this is an extreme example (that is possible to avoid) of a problem concerning conceptualisation. The need to continuous revision conceptualisations before implementing them as ontologies are exemplified above and to have loosely connected links in order to not end up categorising a too strict (and wrong) worldview on behalf of the users. This is also a more severe issue in the humanities since language, history, art and so on are based on perspectives and the openness to variable answers. Within the field of philosophy the definition issue is reoccurring if a user’s perspective influences the meaning of a term.

In extension to this, web ontology developers can be seen as managers of the ontologies by implementing hierarchy classifiers in applications. Then the users are given the task of developing these classifications further. This presents two issues relevant to the development of a semantic web. When ontologies are viewed as *classifiers* instead of *concepts* it raises questions to whether claims made to express knowledge are acceptable. As the vision of a semantic web is to enable and support improved ways of expressing *meaning* that people can make useful knowledge of, classifying things only ascertain a statement, it doesn’t allow reflection or exchange of information. The second issue of this view is that when the user is set to develop the classifiers, *is thus the knowledge building tied to the engineered conceptualisation “the knowledge structure” produced by application developers, or is it the social interaction, the “knowledge sharing” that applies the meaning to the web?* In the development of such a solution the user becomes a central actor and is a provider of adjustments and improvements to achieve optimal results. This side to the ontology development in semantic web projects should be thoroughly considered before implementation and has been a major part of the discussion concerning building ontologies.

We will encounter this ambivalence with where the knowledge takes place when dealing with the Discovery project in the last section.

This issue has also been addressed by Tim Berners-Lee. In *practice*, as he says, technology gives other restrictions than the idea itself does. I believe that there is not enough awareness of this and that this separates properly outlined knowledge projects from inventive database systems. The difference is in the application and usage.

> In practice, it is useful for the system to be aware of the generic types of the links between items (dependences, for example), and the types of nodes (people, things, documents..) without imposing any limitations.
>
> (Berners-Lee 1989)

This quote is a simplified description of the functionality of ontologies today. The last sentence "without imposing any limitations" poses problems in this statement. Any object put in a system would have certain limitations in any case. This leads me to the next part of the thesis that opens up to a view on knowledge as an elusive "element" that cannot be captured in a system. Limitations will be imposed on knowledge the moment it is placed in a system.

How knowledge can be viewed as incomputable

Deleuze (1987) compared knowledge acquisition to a *construction site*. In his view concepts and terms were to be built and constructed in a constantly evolving motion of coinciding fields and influences. Not to be confused with constructivism, Deleuze did not explain the construction of an idea or a concept as merely a combination of words for identification that is added to a knowledge base. He understood it more as a fluid tendency where in the same moment as something is constructed, it is immediately altered and from there continuous to change accordingly to the environment, also implying that this motion is situation dependent.

One central idea was what he called *multiplicity* (Deleuze and Guattari 1987), not as in simple plurality, but the plurality of differences that exists. He stresses the importance of this quality as something that should be cultivated and not restricted within limits. Multiplicity does not withhold a bad quality and what it offers is possibilities and exploration of concepts. This is somewhat in tune with what the web is thought to possess. Embracing the multiplicity of the web is however, not as easily done in ontologies or classification at all. Ontologies call for

restraining concepts, organising them in classification schemes, disabling the quality of multiplicity.

Deleuze's third and most important idea is what he called *rhizome*. It can be compared to polyurethane plastic that expands unpredictably and fills all the cracks in the wood as it finds it random path. Spontaneous, it finds its direction as it expands and paves its way through new territory. "It is composed not of units but of dimensions, or rather directions in motion. It has neither beginning nor end, but always a middle (milieu) from which it grows and which it overspills. (…) [The rhizome] is a short-term memory" and it "operates by variation, expansion, conquest, capture, offshoots" (Deleuze and Guattari 1987, 23). This means that it cannot be predicted and thus not pointed at to say – there it is or there it will be. The process never stands still and fills the voids that it meets on its exploration. As pointed out to the introduction to "A Thousand Plateaus" in *The New Media Reader* (Wardrip-Fruin and Montfort 2003), it should be noticed that they pointed to texts and explored concepts as being rhizomatic and not the tools taken into use for producing these texts. Thus, in the semantic web context, it is the concept itself that is considered rhizomatic and not the form in which is it put on a semantic web.

Knowledge creation or conceptualisations in Deleuze and Guattari's perspective alter the perspective and removes the foundation on which technology based knowledge representation relies upon. It does not provide a true or even close to realistic recapture of what knowledge projects are about. To limit the frame of constructing concepts would directly and fundamentally collide with its nature. This leads to a question of whether a semantic web deals with knowledge at all? Even if the use of only limited knowledge is argued for, it does not stand as a valid argument in this context.

Equally challenging to the idea of "capturing knowledge" in the semantic web is the thought of "liquidity" – a concept professor Baumann explores in his book "Liquid Modernity" (Bauman 2005). Applied on the semantic web and its conceptualisation and knowledge the idea that everything solid is liquefied, transfers to an understanding of concepts as flowing or running over and turn to new blends and alloys. He also says that the individuals are the first to change. If thus the individuals change first, then the concepts will be obsolete after the change. This will lead to fragmentation says Baumann, as the concepts will randomly blend with each other. On the other hand, this is not considered a revolutionary change and the slow

alteration of individuals will slowly change the concepts to belong in contexts, just like the natural process of language evolution. However, liquidity removes all constant unit which creates subtleties and never clear lines between different units. For the heterogeneous quality of the web – the fragmented and random blend of ontologies is anticipated and is sought to be similar to the way that concepts are natural and dynamic. However, if concepts are liquid, how do you capture them in the first place? Also, it requires a massive turn over to remove and change the concepts that do become obsolete. The need to express several subtle differences on the web becomes even more important.

A reoccurring subject when dealing with terms as *conceptualisation* and *meaning* is of course people, as these traits usually are considered to be human traits. Capturing and producing concepts or meaning is usually mentioned in context of human involvement. Just like it is argued against ontologies due to their static classification of words, those in favour of ontologies claim that the user involvement will provide the seemingly needed natural evolvement of the concepts. Naturally, this view, a user oriented view, has spurred several theories on social web, human interaction and user involvement in this view. I will focus on involving people in the making of ontologies and I will elaborate on some of the issues with it when I write about the Discovery Project in the last part of the thesis. What function does human involvement play in this context and how is it influencing the construction of knowledge systems?

User involvement – the result of an egalitarian trend?

> Democratization of information is now becoming...democratization of application development...so the line continues to blur.
>
> VP & Gartner Fellow David Mitchell Smith (Gartner 8 June 2007)

First, I will present the idea of user involvement and perhaps why this approach has been suggested as a possibility. Involving users in development or as content providers has become increasingly common. This trend is, among others, described by Thomas Vander Wal, accredited as the one who coined the appropriate term for a development beginning in the very late 1990's further evolving into services like Flickr. and Del.icio.ous in 2003 and 2004 respectively. Flickr. is a social networking site where photos are being uploaded, tagged and

shared within the community. By adding meaning to the photos with words and descriptions, photos are connected in serendipitous ways or by concrete themes. Del.icio.ous tags web pages and order them after content and popularity. The term *folksonomy* was coined and defined by Vander Wal (Wal, 2007):

> Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.

He says that three elements are needed to create a folksonomy and not a mere tagging service. In addition to the *object itself* and *its tag* he makes an argument that *identity* is the most important element of these three. Seen from historical perspectives, tagging with metadata is not the revolutionary part of this trend. Identity instead connects the tag with the "tagger" and allows other users to see the interlinking between previous tags from that person and the present tags making it possible to logically reason, make assumptions and possibly produce new knowledge about the photo or the person.

Vander Wal emphasises that the tagging of documents is "a means to connect items" for the purpose of users to easier retrieve and find clusters of relevant information. Users provide links to the content that they themselves use by tagging metadata to documents and thus adding "meaning in their own understanding"(Wal 2007). Since information is put together differently to all of us, making loose connections that bundle up relevant items is meaningful when done by regular users and not on forehand categorised links by the producers of applications. Vander Wal also states that this type of tagging is not categorisation, but merely, as stated above, a connection between information.

This implies that a folksonomy is a dynamic process and depending on number of links connecting the various topics together, the closer or more distant the interconnection they will have after a while. Concepts, ideas and views slightly alter in a continuous flow and the views of a public opinion are as multiple as there are people on earth. As one of the major counterpoints to content provision by application developers, the evolving feature of our natural language and meaning as previously discussed has been claimed to stagnate or evaporate when restricted to name tagging. Therefore it would be a natural inclusion to

support these dynamics by deploying people in the development of tags, classifiers and relations.

The widely popular term Web 2.0, or the "social web", has also become a controversial term. There exist a variety of interpretations of what Web 2.0 is depending on how general one wants to be. At the core I believe there is a common agreement that the Web 2.0 includes the increased ubiquitous access to the Internet and that it consist of applications that support some level of social correspondence between users, often in a community. Blogs (personal), wikis (folksonomies), Facebook (social interaction) and LinkedIn (networking) all have in common that anyone can participate on the web. The most common disagreement concerning Web 2.0 is whether it requires some sort of interactivity. Many social networking sites on the web are largely built upon applications where users themselves share, build and alter the content and the application itself. Users are producers and consumers at the same time. "User-generated content" is continuing to grow in popularity. A blog does not need to be interactive, and a blogger does not need to seek other blogs to exist. It is still considered a part of the social web, because it provides public display of a private person.

However, Web 2.0 is not to be confused with semantic web because they're purpose is not the same. By sharing content or tag them does not give the web a more semantically structured character. On the contrary this "personalised" web has contributed to the wealth of information that exists – a lot of it uninteresting or not of a knowledge building character. However, user involvement has showed to be a positive addition to the web and should therefore in my opinion, in a moderated and adjusted form for its purpose, be included in the process of further improving the web. Several voices are speaking of the combination of these two notions. Céline Van Damme, Martin Hepp, and Katharina Siorpaes (2007) wrote a paper on Folksontology where they argue that the popular feature of folksonomies and the structure of ontologies should be combined. Simultaneously exploiting the popularity of tagging, annotating and involving users on the web with existing ontologies and freely available semantic structures from the semantic web (Van Damme, Hepp, and Siorpaes 2007). This bridge between the existing web and the envisioned W3C semantic web uses currently available resources. It also provides a solution to some of the central problems with ontologies. However, this is not miles away from the idea of the semantic web.

Berners-Lee always wanted the web to be write-accessible which is rooted in his belief that people need to be creative and concretises their thoughts. The Web 2.0 is an enabler of this feature. Berners-Lee argues that RDF coupling removes the need of structure. As structure has been criticised, in this thesis too, this coincides well with an attempt to avoid too strict structures to allow more natural dynamics of people to centre the development. RDF already exists on the web and is therefore available to take into use.

The question I posed previously in this paper (on page 63) separated the knowledge discussion on two levels. One level dealt with the content itself and the other dealt with knowledge among the users. Gruber argues that everything (read: knowledge, information, data) can be captured and that the technology or the ontology is not the problem. The most acute issue to address is that of what people know and what people do with technology. He also accepts the collective effort of users that is manifesting itself on the web through social web, grassroot journalism and other collaborative tools. How to achieve knowledge collections from people on the web and how make it a shared asset is his conclusive statement of this argument (Gruber 2008). I concur with this view and consider it to be a valuable approach to the semantic web. However, Gruber says that if a system is being used, it will build itself. I consider this to be partially true. There has to be taken some consideration into a semantic web system, about what type of content that is being built into it. To phrase it simply, I believe that if the input is made up of any data regardless of its quality, the output is not necessarily knowledge.

Annotation, users and trust

Before there is any confusion – in this section I am speaking of annotation from a user perspective and trust issues that occur in sources that are user annotated, unlike the semantic web layer cake where trust is a technical aspect between agents. There should be a general concern that the users of annotation systems actually wish the best for the community. There will always occur, like there are viruses on the web, certain people with dishonest intentions. The likelihood of people having the same goal on the web is unrealistic. One concern, which is a pertinent in this discussion, is the need for trust to exist within a community or group of experts or peers. Two angles of approach are of importance here. The first deals with the trust held to the creators of the system which in the Discovery Project poses an interesting dilemma as the producers are also the target user group. In addition, as concepts and content is

annotated by users – what poses to say that other people share that view and will use the system if the view does not coincide with their own? Will they annotate themselves or abandon the site?

Making use of user annotation systems take the natural evolution of language and concepts into account and therefore stretches far to conform to the ideas of amongst others Deleuze. Even though it does not fully comply with the dynamic and uncontainable idea of a rhizome, it is approaching this notion as far as the frames of the technology permits. Users are not passive recipients of information. According to Kaptelinin and Nardi, "human beings develop their own meanings and values not by processing sensory input, but by appropriating the meaning and values objectively existing in the world" (Kaptelinin and Nardi 2006, 50). In light of the original notion of activity theory, a cultural-historical psychology view emphasizes a problem of user acceptance and trust when applied to conceptualisation. Hence, there are severe implications of assuming that a user will accept an "input" of a certain perspective of a concept instead of contributing to its perspectives actively.

Coinciding with the view of (Braun, Schmidt, and Walter 2007) the success of Web 2.0's low involvement barrier in social networks should be made use of to have success with ontology annotation. The informal and simplistic way of engaging users ought to be further incorporated in the user interface for ontologies. The issues of; matter-of-perspective real world concepts; communities of shared knowledge and practice; extensibility and rhizomatic: ever changing evolutionary vocabularies; all can, if not be solved, then at least be approached by involving users at a high end of implementation. They should be minimally presented with the direct interface of ontologies and be increasingly semi automatically involved in a natural mark up process. This view of invisible user interface applications is also shared by Hendler (2001):

> Semantic markup should be a by-product of normal computer use. Much like current Web content, a small number of tool creators and Web ontology designers will need to know the details, but most users will not even know ontologies exist.

From the conference of "Semantic Days 08" several presenters uttered that the semantic web should not be visible to the user. The semantic web works when the structures work so smoothly that the user is unaware of the technology behind it. Equally important is it to provide a user annotation system that is embedded in the system somehow to ease the weight

off the user. Before engaging in the analysis of the Discovery Project, I wish to briefly view some aspects of finalised projects that have been reviewed in conferences and workshops. Some of the issues that are current on the web are perhaps not as unfamiliar to semantic web development in enterprises as one might think.

Semantic web technology – similarities in different projects

There is a number of projects and research done on semantic web technologies, the modelling of technologies differing a little in each field they are applied to.
There are common obstacles to overcome in deployment and practical solutions that almost every field encounter in some form and to some degree. Though Discovery is walking new ground when applying semantic web technology to philosophical texts, it could be interesting to make a quick review of what reflection professionals on the field have made on the basis of other projects and conferences held on the subject. This can be useful when exploring possible obstacles in Discovery, to see whether the obstacles occur in other areas than or are similar to those in previous projects.

Large-scale projects are a reoccurring objection of realising the W3C semantic web vision. Dealing with large size semantic web projects requires undoubtedly good engineering skills, but size has become a less problematic engineering challenge than before. Enterprises are in possession of mountains of documents in a variety of systems and also have to share them with other enterprises with other systems. They have complex tasks to solve and on a global scale also deal with multilingual and multicultural issues. They have relied on the system to make it work and their hands-on approach can be of use in the researcher community. They are also handling large scale environments that no longer can be called close world any longer. This has become a problem that also the enterprises have to deal with, just as the web research community. Dealing with heterogeneous material is one of the most precarious issues of the web today and to work with interoperable standards are increasingly important to overcome this problem.

In addition to my own observations, the observations mentioned here were also pointed out at among other places a workshop that brought both practitioners and academics together to evaluate and analyse the current situation of semantic technologies and what value and return of investment its holds to current as well as potential investors (STAB 2007). In this

workshop the broad variety of opportunities was emphasised covering concrete solutions with the use of specific semantic technology to approach more general issues and reflecting on visible overall themes. *Ontologies pose a problem to enterprise adoption as well as in Discovery, yet it is according to the discussion at STAB that the refined quality is a high requirement in humanities and on the web, but is not viewed as critical in enterprises.* The disciplines within humanities are challenged with the semantic web system that requires categorisation, alignment and hierarchy and it is faced with questioning concepts and ideas central to its practice. Thus, constraining an idea as fixed data leads understandably to some trouble.

Part 3

Discovering humanities in technology, making sense and sense-makers at WAB

I have restricted my research within the project group at WAB because there were sufficient examples to attend to there that could illuminate important issues that have come up during the discussion about the semantic web in this thesis. However, every aspect of this theoretical discussion is not directly related to the work of the project group at WAB. Naturally, this led me to concentrate on a few central points which reflect the findings that concurred with or were opposed to the basic outlined discussion of this thesis. Additional natural limitations were outlined in the introductory chapter "Explorative research of theory and method". In this chapter, I emphasised that the research style I adopted ambitioned to exploit the potential tension between theoretical material and as the preliminary findings of the practical pilot research I would carry out at WAB. As a consequence, my approach has been essentially explorative. I have sought to identify possible tensions between theoretical insights and my practical findings to enable this analysis. My primary focus was to map the degree of understanding about the semantic web as reflected by the project group. Most of the comments apply to the various contributions to the Discovery's main content repository I had to limit my coverage to Philosource, as; the main public repository derived from Philospace, had not yet reached a practical production level in the project, and, as a consequence, was still at a conceptual stage.

The roles in the project group at WAB can be divided in the roles of the engineers and the philosophers respectively. The first group, the philosophers professional activities at WAB with tasks explicitly related to the production of PhiloSource appears to extend well beyond the traditional core activities for an academic philosopher. Large portions of their work is justified by their expertise as domain specialist, e.g., Wittenstein's philosophy and text genetics, that is experienced as crucial for producing additional knowledge about the sources which will be part of the PhiloSource repository. The second group, the engineers work primarily with programming or conversion of text formats and the implementations of ontologies and annotated texts. They identify themselves as "developers" in the project, a self-description more acceptable to them, but I refer to them as "engineers" here in order to

not create confusion about what "development" covers, since, in the project both roles, philosopher and engineers, are "developing" something, either annotated source, nomenclature (philosophers), or ontologies and programming code (engineers). There are assumptions and perceptions made by these two groups about each other, which are not necessary concordant or symmetrical. To what extent self-perceptions and perceptions of each other's role, correspond to the findings and interpretation of my field work will be discussed below.

WAB – working with several meta-levels

WAB's official role in the Discovery Project is primarily to be a content provider to the Discovery Project. This work consists of both applying semantic enrichment as well as structural enrichment to the Wittgenstein manuscripts before uploading them to the hyper federation (I will comment the use of this term later) of Discovery Project. These activities involve the philosophers annotating Wittgenstein's texts, facsimiles and notes using a suitable mark-up system, combining classical text mark-up (following the TEI standard) with semantic enrichment. Annotating these texts and producing ontologies of conceptualisations requires, in my and the project team's opinion, a deep understanding of the texts and of linguistics – resources that are both accessible at WAB and deeply rooted in their practice. It also requires a certain understanding of ontologies in this context. I will nevertheless, in this chapter, attempt to shed some critical perspectives on the self-understanding of what I have called "roles". Annotation means marking describing, expanding, deepen or reformulate the text with terms that describe a specific paragraph or section of a text. The semantic annotations are mappings of data that are later linked to formal ontologies. The textual and thematic data is, by the help of annotation, linked to other available and "relevant" pieces of data because the conceptualisation operated at a human level (by the philosophers) and algorithmically (by the engineers) connects these and situates these in their hopefully relevant context. This content is then further processed by several concepts that internally collaborate to establish its relevant context (the multi-level notion of relevancy is not discussed in depth here). Relating the semantically annotated data and creating the dynamic relation links is what is considered to offer structural enrichment. However this requires the philosophers, some understanding of the computational requirements of doing so. As consequence, a full separation of roles between philosophers and engineers appears to be impracticable. However, as I have discussed in the section "Semantics needs context – but what context?" above, it is

complicated but the most challenging fact that there exists no singular agreed explicit definition of semantics.

In the WAB team, pure philosophers and experienced text encoders are working together to reach the best achievable annotation result (in relation to the "engineers" this subgroup are the "philosophers"). This interlocking of the two related but distinct expertise, the domain knowledge on Wittgenstein's philosophy with text encoding constitutes a necessary practical and potentially authoritative (this last issue will be dealt with below) starting point for providing a potentially better understanding of the sources and their authors exploiting new knowledge connections between relevant concepts in the source texts and commentaries. The philosophers need, ideally and practically, to interact, not only with the text encoders, but also with the knowledge engineers to ensure that their annotation structures were computational as ontologies. These engineers' role is to convert and adjust the philosophers' annotations, emerging from the collaboration between the pure philosophers and the text encoders, so that the resulting semantic structure may be embeddable in the universal platform created and interoperable with other material within for the entire federation without losing its "meaning" in the process. The term used to characterize the architecture is "federation" (directly derived from the French expression "*fédérer des textes* "). This term seeks to express ambitions to express the interoperability of all the included sources of the Discovery and the policy adopted by the project consortium to achieve this. It functions like an overarching interoperable architecture and content provision. The providers within the federation are meant to collaborate even if the different communities and institutions involved have produced specific and proprietary solutions to cover individual needs. Federating these contents introduces new requirements and standards, e.g., respecting multilingual environments and cultural context. The Groupware used within the project is meant to be frequently used for this collaboration purpose. I will however, not pay particular attention to this as the detailed dynamics of interpersonal collaboration within the Discovery team is outside of the scope of this work.

The philosophers that work on annotation could mainly concentrate on the texts and they should, ideally, not need to consider the technical aspects. However, the pure philosophers' annotations have a direct implications for both, the formal decisions and design choices of of text encoders and the possibilities, constraints and choices imposed on the knowledge engineers. However, the philosophers and engineers appeared in the course of my

conversation not to share the above formulated definition of a fundament interdependency of all design stages and roles. In reality, a stricter separation was achieved, not mediated by some kind of procedures for reciprocal validation, as the general opinions and decisions of philosophers and the design choices made by the engineers, seemed to be automatically assumed and accepted. The engineers were not put in the formal or informal position to produce opinions on the annotation of the philosophers, and, most often on the overall thematic architecture of the text encoders, because their domain of expertise was, in their own view and in the view of the philosophers, restricted to engineering. This applies too in the reverse direction. I will come back to these issues when discussing semantic enrichment later in this chapter.

The specific texts of Wittgenstein add an extra twist to the annotation issue because Wittgenstein himself discusses the problematic issues of knowledge acquisition and meaning in texts. The fact that the philosophers are chosen to annotate the texts seems like a natural choice as they are already familiar with Wittgenstein's ideas and are likely to be better suited to capture and communicate the essence of his texts better than the engineers. However, this places large assumptions on the capabilities of the philosophers in terms of understanding ontologies. However, I found that there was generally incomplete understanding of what an ontology really is and how it relates to a semantic web construction. Both parties, respectively the philosophers and engineers expressed that there was little contact between them. Here, there is a perspective lost of the overall goal of the annotation, in which a closer cooperation between the engineers instead of an almost exclusive work applied by the philosophers in annotating the texts. The programming aspect is not the primary part of the annotation process either, and so an engineer alone would most likely not be suitable for this work. This division of the roles and the assessment of semantic value in the texts pose several interesting tensions when it is interpreted in light of the semantic web discussion of this paper.

First of all, it is important to understand the many meta-levels of semantics that can be identified in this project. This is *the* cause to most major complexities in this particular project. Some distinctions have been mentioned and project members are aware of these interpretations and their influence, but have perhaps not explicitly been identified. I will clarify three levels here, though there might be more. Annotating the Wittgenstein texts can be regarded as semantics in the traditional linguistic manner. The textual content is identified to have a certain meaning in which semantic marking is applied to the data itself. I regard this

as the first level of semantics. This level is relatively visible to both the users and the producers as it describes the texts themselves. The second level of semantics is the meta-data. This is semantic data about the descriptions of the texts. These are the comments and added value to the texts at WAB. The semantics at this level is harder to identify and categorise, but plays a central role in connecting sections of the texts together. Finally, semantics are regarded from a computational way in the discussion of semantic technology itself. From this point of view, semantics are the connective pointers that put a set of data in a setting with other related data. Again and in other words, these three levels can be identified as (1) a pious respectful lexical paraphrastic expansion of terms and concepts that (2) with precision and carefully handpicked cross references add up as an expositio which is visible as (3) an interpretatio in the use of semantic technology. The very first interpretation can also be expanded to include all scaling. On all of these levels, as outlined in this thesis, the complexities and implications of interpreting semantic value from different perspectives and contexts can invite to unpredictable results if not taken into consideration. The entire project at WAB illustrates key issues concerning semantics and meaning discussed in this paper as well as it illustrates the key concept of the project. In other words, the project group of philosophers at WAB deal with the same principle ideas that the entire project itself tries to capture. Dealing with the many levels of semantics and attempting to transfer knowledge to others by using correct terms are among the central concerns at WAB. I assumed at first that there would be problems with ongoing discussions on terminology within the texts. However, I did not anticipate that the problem would lie within the project on so many levels as with the semantic discussion.

An example of this came up during a discussion on ontologies in the groupware used by all project members of the Discovery Project. On several occasions during this discussion there seemed to be slightly overlapping comments which did not connect because there was uncertainty about whether the comments were of a technical or a philosophical kind. All parties at WAB expressed that they were aware of the differences between the discipline and the aspects of philosophers and engineers, but they tried to make the communication as smooth as possible. A classic view on the correlation between the two sides is that philosophers feel that engineers do not understand the content of the documents because they view them as documents only. An engineer considered philosophers to be more pragmatic in their approach to the texts than necessary because they were concerned with making the texts suitable for the technical structure. As unusual as this may sound at first, it is not an unlikely

assumption. From a philosophical viewpoint, and referring to the previous discussion in this paper, a knowledge system does not necessarily contain the knowledge and insight as promised by many technologists. The value of the information in itself as it exists in a computer format does not give any insight to a user. If the information string is connected to another string of information that by time could be regarded as relevant and thus clustered closely together with annotations, *then* there might be serendipitous insight to retrieve from a system. Anything that is included into a system is not always usable for knowledge purposes.

Returning to the philosopher's role in the project, it can be added that they can be seen as both producers as well as potential users. This duality was sought by the philosophers to come to the benefit of the users. The philosophers know the texts well and they apply their best knowledge and insight when working on the annotation of these parts. In this dual role they are connoisseurs of Wittgenstein, and viewed by other philosophers as peers, but also function as mediators in which an interpretive role is taken. The Wittgenstein corpus may contain many interpretations and perspectives. From a "working" perspective in which the philosophers are annotators of the texts, they have different premises to evaluate the texts within. However, an observation from my point of view as well as an expressed concern of the philosophers I spoke with was that they thought it was important to be as objective as possible and not to apply too specific meanings to the specific passages. If this happened, there was a danger of ending up with restricting a passage to one meaning whereas other philosophers might not agree on this view at all. I'll come back to this shortly when discussing which user perspective that was taken into account during the annotation work.

In some regards, this dual role is not much different from how the social web is constructed. Users are often content providers acting in the role of a producer, but equally acting as the consumer and user.

Several sources I spoke with at WAB seemed to quote the project description regarding the user segment being young scholars and academics. When attempting to differentiate and identify a specific user, slightly different types of users were mentioned in the descriptions. This varied from "anyone interested in philosophy", "those wanting to find introduction to the texts and connections between the philosophers" and "those wanting to gain access to specific passages that were unattainable as easily as the Discovery Project would enable". The range describes anyone from those who are new to the discipline of philosophy to those who know

what to look for in the different texts. This does not completely coincide with the idea of allowing for discussions and comments on the texts through a peer review board in order to contribute to the levels of knowledge about the texts as this would require a user to have a certain level of insight into the texts. Of course, they do not contradict each other either. However, involving the user perspective earlier and as a part of the philosophers work at the very "atom" level, could perhaps have relieved the annotation process when knowing from the beginning who the targeted group is. The idea of the user seemed unclear at this stage of the project, but always present and with a certain point at young scholars. Undoubtedly, this is related to the fact that the project had not yet reached the particular user stage in the development at the point I was conducting my earliest interviews. All attempts to focus on any concrete user perspectives in this thesis yield only preliminary and a possibly evolving picture.

Who among the different project actors at WAB are thinking of the user? All sources at WAB mentioned that the user was central to their development. As there is no immediately clear definition of the user at this stage, several participants of the projects, both engineers and philosophers expressed some concern about the reaction to this entire system – as a system to the specific discipline of philosophy. Philosophers may claim their right to individually interpret and dynamically reappraise and evaluate the structure and dynamics of philosophical texts. A collaborative approach to the texts of philosophical texts could, according to one of the philosophers, create some difficulty in making the community of philosophers actually use the PhiloSource for their own purpose. This actualizes in a new context the old issue of textual authority. This is not stated as a fact, but only as the observation of a dilemma and the awareness by the philosophers in the project of a possible obstacle to reception that need to be dealt with in an explicit manner, Interestingly enough, one source commented on the user perspectives of the philosophers as not being optimal because some of the semantics that was being opted for was lost in the shortcomings of understanding the technology. The same source further elaborated on this by saying that perhaps the philosophers were even were more pragmatic in their approach to the texts than the engineers, which contradicted my posed assumptions about the producers. From my viewpoint the user view was very much present in the annotation process and during conversations with project members. I found that the emphasis was on how the project appeared to the user and if the philosophers' annotations were not sufficient and logically constructed for a user then the user would not have any benefits of the project. However, I was also surprised by the focus of the content providers of

having texts that were computable and interoperable to be incorporated to the federation .An engineer commented that this was of more concern to the philosophers than to the engineers. This too could reflect the difference in understanding the content and what adjustments that were needed for the philosophers to feel that meaning was enabled in the structure they provided for it.

Trust between producer/user

One of the philosophers expressed concern about the "expected" users. This philosopher had shared his concerns with another philosophical peer from outside the project who expressed Even if the annotator is an expert within a field – the users do not necessarily know this person. If a philosophers in the project then attempts to keep comments and annotations as objective as possible, there is a chance that the annotations will be met with disbelief. Especially within a discipline that is so sensitive to multiple aspects and conflicting definitions. While the markings of the texts are meant as merely pointers that suggest a belonging, preferably, without posing too many limitations on the content, their reception by new academic users may build upon a qualitatively different presupposition than the producers. The notes may not be taken as 'suggestions' and 'references' but as disputable statements. This is a well known and reoccurring problem when it comes to ontologies and semantic annotation and was seemingly also a concern of the philosophers and engineers.

What I found the most interesting concerning the user perspective was that a periodic report about the entire Discovery Project released 1st of May includes the survey results about the user segment. The specific answers provided by the respondents cannot be outlined here as the report is classified as confidential. However, what is significant about this is that the exact results from the report seemed to be unknown to the participants at WAB.

Trust between peers

Between the peers there need also be a level of trust which builds upon the mutual recognition of some authority. The Discovery Project aims to be a source for academic and scholarly use. On a micro programming level this is solved by adding unique markers URI's to all the documents that are put on the site. In addition there is a complex archiving system that systemises the versions of Wittgenstein's paragraphs, sections and other alterations to his

texts. Comments and added annotations is envisioned to, with use over time, mature and be semantically clustered and incorporated as part of the semantics added in the original texts. Experienced users as well as new researchers to the field will apply their personal knowledge and insights to the texts as a supplement to the knowledge base for others to learn from it, comment on it or provide feedback to the originator of the comment. However, philosophers are not necessarily seeking answers in other's experiences *online* to find applicable solutions in the same way a project-centered oil drilling company might do. Support to writers on philosophical work of writings such as reviews, comparisons and new texts are possibly intended by the Discovery project as the main service. This author work is eminently *personal* to a large degree. The community of practice involving annotators and future academic users are served with a collection of work of significant philosophers and provided with tools to find certain passages and interlinked passages with less effort.

The Discovery project plans to appoint a peer review board to evaluate standards, formats and review the quality of input from users. The review board's presence might contribute to the opted quality insurance. It can on the contrary also limit the purpose of a freely editing community of philosophical peers. The Discovery Project will balance on a fine line between removing the egalitarian social perspective of free subjective perspectives and being a reliable academic source. This resembles the same problematic issues most social sites on the web face today. However, as the semantic web vision is for the knowledge base to grow as a consequence of user involvement by annotation, putting too much restriction on this function could lead to an intellectual decline. In addition, there is no time or resources for all the ontologies to be reviewed before they are published to begin with. At WAB there is a closely connected community of researchers and practitioners of philosophy that help each other.

Conceptualisation and knowledge perspectives dependent on many levels

Within the field of philosophy, conceptualisation is not an easy topic in itself, again emphasising the meta-levels of the project. I will return to the issues of agreeing on concepts in a few paragraphs later on, as this can also create trust issues from a users perspective. Philosophy requires timely consideration to the accuracy of a single term. On the contrary philosophy also requires an understanding of evaluative perspectives and embraces the relativity in terms and concepts. Like humanities in general, a major part of the work of the philosophical discipline consist of defining and redefining the field. This view can be

extended and related back to the discussions of Deleuze to consider the discipline of philosophy as a construction site that is rhizomatic as well as situational. Within the field of humanities, philosophy is probably the discipline that poses exceptional challenges of definition to semantic enrichment. In this regard, I appraise the Discovery Project for attempting, what I would consider to be the hardest task of them all; to apply semantic web technology to induce more meaningful connections between the works of a discipline that for centuries has always been puzzled with the complexities of human culture as well as language.

Philosophers tend to translate their expertise by highlighting and weighing subtle, often implicit differences in rather simple textual statements, while knowledge engineers tend to favour explicit, uniquely defined, efficient terms and clear relations between these terms, often classifying them hierarchically or in tree structures. Conceptualisation of philosophical texts needs special non-explicit knowledge of what are underlying notions as they appear in the specific texts as subject to annotation. The philosophers at WAB, and on the Discovery Project act on the explicit or implicit assumption that they possess this specific knowledge.

As many of the project members are used to working with digitalised texts and text-encoding, they are a part of what can be labelled a precision culture. They already have a culture of precision incorporated into their general academic. It can be regarded as an advantage to have a precision culture which already existed amongst the project groups members at WAB from the beginning. It presupposes a culture for detail work in the scale that the works of Wittgenstein requires of WAB. They have significant experience of text and language analysis and comprehensive understanding of texts in digital environments. Taking on the role of “user as producer” can increase the respect and understanding for the precision required of their field. As the production of specific terms and concepts for the text corpora in Wittgenstein require this precision knowledge, WAB manifests as a culture which already has the tools and routines that a tightly connected community possesses, and this should greatly contribute to the quality and the processes of semantically enriching the texts of Wittgenstein. The capacity and will to deliver knowledge is more likely to occur in a community of peers that share commonalities and personal relations. It is the same principle that drives a tightly connected community of practice to share their knowledge and expertise with each other, which otherwise would have remained implicit. That being said, a precision culture is consequently not as different from an “engineering culture” as one may think. It demands efficiency, exact computing and

The core discussion about terminology and semantics has treaded into the background at this point to benefit a focus on structural enrichment. This is seen and mentioned by two of the philosophers as an enabler of a functional solution. As this becomes the focus of discussion, it too gave room for a more pragmatic interpretation of what semantics are. An engineer said that the minimal involvement of engineers in the enrichment process at WAB, perhaps lead to an unnecessary underestimation of the capabilities of the semantic web technologies.

Practical approaches

Initially there were two approaches to annotating the texts in the project. One of them was a bottom-up approach and the other was a top-down approach. In practice, top-down means drawing up a list, just like a dictionary, and using the terms in that list to appropriately mark the texts. A top-down approach requires the annotator to produce a list of appropriate terms during the ongoing work. A nomenclature (an overarching thematic structure) is agreed upon so that these annotations can comply with under some kind of structure. Finally, these structures are connected to a large superstructure combining all the projects of the federation together in the Discovery Project. Each provider of content to the federation of Discovery is dealing with different structures of data (multimedia, video, text, audio) as well as different vocabularies to suit the themes of the philosophical texts. To that end such a structure was built. In practice this meant that this bottom-up list might consist of several more terms than what the top down approach contains. In addition, it will contain many individual terms, instead of a set of given terms, that have to be incorporated into an ontology. On one hand this may cause inconsistencies in semantics and pose difficulties for the computation, but the nomenclature can contain them to a large degree. On the other hand, it can provide an accuracy of meaning that cannot be provided by a pre-chosen vocabulary. At WAB, there was a trial period in order to find the right solution for the Wittgenstein texts, but finally the bottom-up approach was chosen.

As the Wittgenstein corpus has to be converted from mexvit to XML-TELi, WAB has to carefully remove mistakes from the new text corpus. A problem with reusing old information repositories is the overflow of documents, tags, mistakes, errors and multiple copies and versions. This leads to two issues. Firstly the purpose of producing a new system is (hopefully) not to immediately create the same problem that the new system was supposed to attend to.

This again will corrupt a semantic enrichment of a text. Secondly, at WAB they have to deal with digital facsimile, encoded texts and different multimedia formats. Generally converting formats to semantic formats introduces a risk and require extra attention to the details in the information. This issue was also mentioned on several occasions at the Semantic Days 08 conference. If all the material is just exported from one system to another, the content as such is not necessarily even appropriate to ensure a true knowledge base that is of value to a user. In a closed project where users are involved in the continuous expansion and alteration of the content this should be possible to take into consideration. However, as was remarked by an engineer, there is a huge risk of and it is also a fact that information will be lost when there are several conversions that need to be seen through in order to incorporate it into a superstructure. The semantic web layer cake has a structure that ensures a structure for streamlining data transfer, but at the same time contributes to this current issue of converting formats. The same source also commented on the risks of losing information in the Discovery project. As the philosophers are marking the texts, adding information value to them in form of metadata and supplementary commentary, this information is an interesting and potentially significant value to keep and convey knowledge that would otherwise be lost in the conversion of data from one system to another. Then the ontology does not only consist of information pieces, but also added descriptive metadata about the content which can contribute to the overall semantic understanding of the content even without the data that was lost in the transference.

"Reflections on the Discoveries made"

Assuming that the Discovery Project is of stronger ideologically motivated character than say, enterprise projects are – it has nevertheless the opportunity to exploit and re-contextualize practical experiences taken from enterprises in order to demonstrate how the semantic web may be made more efficient, and most importantly and meaning (cp. making sense) to a community. To an extent I believe that the enterprises have pressured needs for improving technical solutions and thus quickly come up with usable practical solutions, but, one may ask - are research projects without this kind of pressure?

I have given some treatment of the producer/users approach earlier on page 17. The Discovery Project can be viewed as a user-oriented project that aims to create a support tool for researchers and scholars within philosophers. Nonetheless, as it is also a funded EU-project, it

is contractually bound to exemplify as a successful humanities project for others to follow. The project needs to produce a product that can "sell" or have a demonstrable impact on a larger community. This places the project under a certain amount of pressure of performing within budgets and time frames and the project itself is visibly affected by this in both positive and negative directions. On the positive side the project can show to efficiency and results, however on the negative side this can affect the quality of the product as smarter and quick decisions are being made to meet time limitations.

Surprisingly few of the sources at WAB mentioned the purpose of the project as creating any "sense making" to the users. When the computational aspect to semantics was discussed during the interviews, the implications of semantics were quickly dismissed in that sense. I trace this to a certain degree back to the terminology confusion in the theoretical discussion of the thesis, but from what I could sense, there were few who spoke of or had been thoroughly introduced to the idea of a semantic web at all. In general, the explanations of their work was focused on providing access to the specific parts of the Wittgenstein corpus and being able to set single sets of texts about a specific theme in comparison to other philosophers who might discuss the same theme. It seemed to me as a simplified interpretation of the semantic agenda, but at the same time a reasonable and pragmatic way of getting results and not only discussing the alternative solutions and possible implications of various decisions. The general complexity of the Discovery Project has been greatly taken into consideration at WAB and even though the philosophers have a solid understanding for marking the texts, the technological understanding of a semantic web was minimal. If there have not been clarifications of the projects purpose in the beginning as to what the terms imply it can have major implications when modelling the project. WAB show a willingness to stretch far to create good communication between the members. The atmosphere between the philosophers is good and they are active in helping each other with clarification issues of the content – a culture they are benefiting of, as it was pre-existing to the project. An effort is made not to misunderstand the concepts; however, if the concepts are not put in the context of semantic web, as the Discovery Project attempts to make use of semantic technologies, then the purpose of adding semantics can be very different with the different involved parties.

Building semantic web applications is a time- and knowledge consuming task and very demanding of all involved parties, both to the engineers and what was mostly the focus of this thesis, the philosophers as content and structural providers.

The semantically enriched environment that is opted for in the final results might depend on all the details for the work to prosper. Tradition in cultures and tight communities within a specific field of work enables easier structuring, but imposes a more rigid authority and add formal and well as informal constraints on what can be done. Even if both the project group at WAB shows strongly connective relations as well as the user group show community belonging, there are no guarantees of the success of the project. This largely echoes the notion of Gruber (2008) of "if they use it, it will build itself". There can be detected a anticipation to the Discovery Project from the project description that the system will provide new and more meaningful and useful information that is based on human contributions and only augmented by the semantic web technology that is used, in the lines of how Gruber describes "collective knowledge systems" in his presentation (Gruber 2008). I do not have that same impression of this same anticipation from the participants at WAB. There seems to me that a more pragmatic approach is taken to the semantic enrichment of the Wittgenstein corpus. As I have illustrated in the above section about WAB, the focus of the philosophers was to label and create relations between selected themes and entities in a logical system. What is important to note though, is that I do not think that the semantic perspectives are eluded in the project development, only that the sources that were interviewed from the project group did not reflect upon the larger semantic web implications of creating the complex system they were participating in.

Both the views of the philosophers as well as the engineers are of course necessary in the project. As I have concluded with earlier on, semantic technologies require high capacity, integration, interoperability and processing power, which demand a lot from the producers of the technological solutions. In the Discovery Project there is also a demand for highly skilled producers and managers of the philosophical content. Philosophy poses great challenges to the semantic web idea in terms of semantics, perspectives, to the users and in production. This and I would assume similar issues in other humanities in relation to semantic web projects; become highly *knowledge intensive* projects as well as *production intensive* projects.

A part of the project description on the web site goes as follows: "(…) explore how Semantic Web technology can help to create a state-of-the-art research and publishing environment for philosophy". While the end result should be primary purpose and focus to the project, the production process of the development requires an enormous amount of detail work. These

constructions and pathways to solutions were created along the way. What puzzled me during the research period was the repeated emphasis on “exploring through practice”. Naturally since the project is paving new ways, the philosophers had a major task of attempting to provide sufficient solutions for annotation and structure of the Wittgenstein corpus. Referred to as a bleeding edge project the sheer size of the project implied not only a need for tight cooperation between the different roles, but also a unified understanding of the purpose of such a large sized semantic web project. Without a fundamental understanding of the semantic web vision, how are they to reach for it?

Context and continuum

Being a part of the EU

Even though it is a project in practice – a reflection of a semantic web technology development in its larger context can be quick and useful. The Discovery Project is a research project in construction that belongs under the umbrella of EU projects, which opts to research on and illustrate the possibilities of new technology. It is not an autonomous project completely without surroundings, and this has naturally influenced the decision making and choices of technology.

Standardised technologies that can be regarded as part of the semantic web layer cake are being used such as OWL, RDF, XML and so on. The web layer cake diagram is also pictured on the webpage of the Discovery Project, which is a clear statement of what standards that are being used for this semantic web project. Open source is also a requirement from the EU headquarters illustrating the support given to standardisation of technological data streams in Europe. The EU is a central actor when it comes to developing Open Source applications and applying them to multiple projects. The EU has written in its publicly available reports about the project that the EU wish to seek solutions to meet the demands of the knowledge economies facing the European Union in an increasing scale. Preserving the culture and history of Europe is just one of the multiple ongoing projects in the EU.

As a funded project within the EU at large, the Discovery project works as a binding project of several high valued ideas. Firstly I have discussed the notion of a knowledge society and the actual need of improved performance developed from an increase of knowledge sharing. I

write sharing here as it leads me to the next point of cooperation. Typical knowledge economies are mostly seen in the European countries, as they are among the leading actors of the industrial activity towards knowledge directed areas. Special competence within highly advanced and specialised knowledge engineering is for example one type of knowledge provided. Another is the typical knowledge work consisting mainly of decision work and strategic moves within a business area. Both constitute the era of a knowledge economy and a notion of the knowledge society. The EU relies on cooperation and is perhaps in itself the best example of distributed cooperation to provide solutions through connections and education between nations. Sharing knowledge between nations is seen as the best enabler for future sustainability of growth for Europe as a whole. This is similar to the same shifts as seen in for example the educational sector.

However idealistically the brief first impression of the Discovery Project may seem, the project is funded by and deeply invested in by the EU as well. All investments in the future that are sought to forward improve solutions can be viewed as research investments to the EU. Return of investment is important, but in a social anthropological perspective, a return of investment can be understood as any valuable asset, also knowledge value. Discovery Project the aim is perhaps not to return investment, but provide a solution and experience that in turn can be used in other future investments. It is a long-term investment that might contribute to improve societal issues as well as it is a tradition of investing heavily in the development of possible future scenarios. All speculations aside, because the Discovery Project is funded, it is subject to limitations – in time, money and resources. If it doesn't show results the project will be terminated. It is clear however that on the contrary to purely economically driven solutions in enterprises, projects in the humanities are additionally more idealistically motivated. Economy is not to be undermined in the humanities, but profit is not the golden goal of creating a work space for researchers within philosophy as in Discovery.

Concluding this Discovery

Taking an explorative approach to the historical development of the semantic web, I have uncovered a field that is significantly fragmented. Some of these fragments are causing a variety of possible interpretations which seem to have been dominating the field of semantic web development causing even more sporadic constellations of understanding the semantic web. Semantics and knowledge are such fragments and sources of interpreted semantic web constructions.

As I have demonstrated in the theoretical discussions of this thesis the semantic web seems almost like a fragmented piece in its own puzzle. There are so many different models at work, all assuming different premises, which are all interacting with each other. The prevailed confusion about what terminology to use within the field of semantic web is not surprising. There is a need to come closer between high knowledge level and low computational level of semantics to close the gap between terminology and application.

In the same way as the semantics could be identified on three different levels at WAB, these three angles can also be detected in the general semantic web landscape of theoretical perspectives. The different manifestations of semantics have been outlined in this thesis. The first approach considers the *structure* to be the empowering element in semantics – this can be exemplified with the interoperable and streamlined technologies that are sought to optimise semantic transfer in the semantic web layer cake. Another view opts for a content approach in which the power of semantics is invested in the pure choice of specific content to convey semantics. This states that the quality of the *language* is what carries with it knowledge. Others, such as Gruber, states that there is too little emphasis on the users and, indifferent of how a system is made and ontologies are thought through, the *user* sits with the power to use it or choose not to use it, which means that there is no semantic transference in which there is no knowledge system.

The aspect of technological autonomy is also a guiding idea in this paper and throughout this thesis the arguable difficulties for a system to act autonomously, is supported by several arguments about contextual dependencies in both linguistics, cultural, social, psychological and practically any human contexts. The idea of automation and logical deduction systems falls through the theoretical discussion of this thesis, and the implications of non-automated

systems should be able to shift the focus of the semantic web, increasingly towards the user centred approach. The original vision of Berners-Lee was to enable a construction that would improve the communication between people and computers

The question still remains whether the Discovery Project will succeed in its exciting semantic construction of the philosophical content. Even though there are still certain obstacles regarding the construction of a semantic web, I would argue that the semantic web has a strong potential provided considerable amount of attention is given to understanding the overall purpose of such a knowledge and production intensive project. As I have argued in this thesis, a contextual semantic aspect does not necessarily impose restrictions upon the field of humanities. The definition issues posed by the semantic web should be approached with due awareness; even so, such a view can also offer the flexibility of interpretation as is incorporated in most humanistic disciplines, a flexibility that had been less available in the traditional taxonomian form.

Bibliography

Bibliography

Antoniou, G. G. 2004. *A Semantic Web Primer*. Cambridge, Massachusetts
London, England: MIT Press.

Bauman, Zygmunt. 2005. *Liquid Modernity*. Cambridge: Polity Press.

Berners-Lee, T. *What the Semantic Web can represent* 1998 [cited. Available from http://www.w3.org/DesignIssues/RDFnot.html.

———. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*: Harper San Francisco.

———. 2007. *Hearing on the "Digital Future of the United States: Part I -- The Future of the World Wide Web"*, 03/01 2007 [cited 06 november 2007]. Available from http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html.

Berners-Lee, Tim. *Information Management: A Proposal*. ©Tim Berners-Lee 1989, 1990, 1996, 1998 1989 [cited. Available from http://www.w3.org/History/1989/proposal.html.

Berners-Lee, Tim. , J. Hendler, and O. Lassila. 2001. The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* May 2001.

Braun, S., A. Schmidt, and A. Walter. 2007. Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. In *Proceedings of the Workshop on Collaborative Construction of Structured Knowledge (CKC) at the 16th International World Wide Web Conference (WWW 07)*. Banff, Canada.

CERN. 2008. *The website of the world's first-ever web server* 2008 [cited 20.03 2008]. Available from http://info.cern.ch/.

Decker, S., D. Fensel, F. van Harmelen, I. Horrocks, S. Melnik, M. Klein, and J. Broekstra. 2000. Knowledge Representation on the Web. *Proc. of the 2000 Description Logic Workshop (DL 2000)*:89–98.

Deleuze, Gilles, and Felix Guattari. 1987. *A thousand Plateaus* London, New York: Continuum.

Drucker, P. F. 1994. The age of social transformation. *The Atlantic Monthly* 274 (5):53-80.

Drucker, Peter F. 1993. *Post-capitalist society*. Oxford: Butterworth-Heinemann.

EU. 2007. *Europe's information society, thematic portal* [website] online [cited 15.11 2007]. Available from http://ec.europa.eu/information_society/activities/econtentplus/projects/cult/discovery/index_en.htm.

Gartner. 8 June 2007. *Trends in Web Innovation*: Gartner. podcast.

Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. In *Knowledge Acquisition,*: Academic Press.

Gruber, Thomas. R. 1993. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43 (5/6):907-928.

Gruber, Tom. 2008. Despite our Best Efforts, Ontologies are not the Hard Part. Presentation at AAAI Spring Symposium: tomgruber.org.

———. 2008. Ontology. In *Encyclopedia of Database Systems*, edited by L. Liu and M. T. Özsu: Springer-Verlag, .

Habermas, Jürgen. 1960. Borgerlig Offentlighet Bergen: Published in "kopisamling for seksjon for medievitenskap til MEVI 101", 2003.

Hansen, Lars Holm. 2006. Internett og den nye offentligheten. *Morgenbladet*.

Hendler, James. 2001. Agents and the Semantic Web. *Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications]* 16 (2):30-37.

Hidas, Peter. 11.04.2008. Standard og standard, fru statsråd. *Computerworld*

Jonassen, D. H. 1994. Toward a constructivist design model. *Educational Technology* 34 (4):34–37.

Kaptelinin, Victor, and Bonnie A. Nardi. 2006. *Acting with technology: activity theory and interaction design*. Cambridge, Mass.: MIT Press.

Lambe, Patrick. 2007. *Organizing Knowledge: Taxonomies, Knowledge and Organizational Effectiveness*: Chandos Publishing Oxford Limited.

Lewis, R. 1997. An Activity Theory framework to explore distributed communities. *Journal of Computer Assisted Learning* Volume 13 (Issue 4):Page 210-218.

Malpas, J. 2002. The weave of meaning: holism and contextuality. *Language & communication* 22 (4):403.

Matsuura, Koïchiro 2006. *Towards Knowledge Society.* 2005 [cited 23.02 2006]. Available from

http://www.eLearningeuropa.info/index.php?page=doc&doc_id=6825&doclng=6&menuzone=1.

McGarry, K.J, and Burrell, T.W. 1972. Centrality of semantics and logic in retrieval. London: Clive Bingley.

———. 1972. *Semantics in the organisation of knowledge*. Edited by C. D. B. B. FLA, *Programmed texts in library and information science*. London: Clive Bingley.

Miller, Dr.Paul. 2007. *Moving from a web of documents to a web of data*: University of Cambridge. Podcast.

Mills, Sara. 2003. *Michel Foucault*. Oxon and New York: Routledge.

Mädche, A., S. Staab, and R. Studer. 1999. Ontology-Oriented Information Extraction and Integration. *Workshop on Language Technologies in Information and Knowledge Management. 7th Conference on Computational Linguistics of the German Society for Language Technologies, Saarbrucken, Germany*:7–8.

Noah Wardrip-Fruin, Nick Montfort, ed. 2003. *The New Media Reader*. Cambridge, MA: The MIT Press.

Online, Encyclopædia Britannica. 2008. semantics In *Encyclopædia Britannica Online*.

Passin, Thomas B. 2004. *Explorer's guide to the semantic web*. Greenwich, Conn.: Manning.

Patton, Michael Quinn. 2002. *Qualitative research and evaluation methods*. 3rd ed: Sage publications.

Reding, Viviane. 2003. The Global University system and the European Union. In *Global Peace through a global university system*, edited by T. Varis, T. Utsumi and W. Klemm. Hämenlinna, Finalnd: Saarijärven Offset Oy.

Scardamalia, M., and C. Bereiter. 2006. Knowledge building: Theory, pedagogy, and technology. *The Cambridge handbook of the learning sciences*:97–115.

Schreurs, Nard. 09.04.2008 Dropper Windows for Linux. *Computerworld*

Semantic Web Search 2004-2007 [cited. Available from http://www.semanticwebsearch.com/query/.

Sheth, Amit, Cartic Ramakrishnan, and Christopher Thomas. 2005. Semantics for the Semantic Web: The Implicit, the Formal and the Powerful. *Int'l Journal on Semantic Web & Information Systems* 1(1), 1-18, (Jan-March 2005).

Sheth, Amit., and Cartic Ramakrishnan. 2003. Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis. *IEEE Data Engineering Bulletin,* .

Sheth, Amit., and Meersman Robert. 2002. Amicalola report: database and information systems research challenges and opportunities in semantic web and enterprises. *SIGMOD Rec.* 31 (4):98-106.

Smith, Barry. 2004. Ontology. In *Blackwell Guide to the Philosophy of Computing and Information,* , edited by L. F. (ed.).

Sporny, Manu. 2008. *Introduction to the semantic web*: Open source at Digital Bazaar. podcast.

STAB. 2007. The first international workshop on semantic technology adoption in business, from concepts & prototypes to practical solutions. Paper read at STAB 07, 1 June 2007, at Vienna, Austria.

Swoogle 2007 [cited. Available from http://swoogle.umbc.edu/index.php?option=com_swoogle_stats.

Swoogle semantic web search 2007 [cited. Available from http://swoogle.umbc.edu/.

SWSE 2008 [cited. Available from http://swse.deri.org/.

The Discovery Project. *Philosophy in the digital era? Making sense of stuff with semantic web technology* 2008 [cited. Available from http://discovery-project.eu/making-sense.html.

Van Damme, Céline., Martin Hepp, and Katharina Siorpaes. 2007. FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies.57-70.

Varis, Tapio, Takeshi Utsumi, and William Klemm, eds. 2003. *Global peace throuhg a global university system*. first edition ed. Hämenlinna: Saarijärven Offset Oy.

Wal, Thomas Vander. 2008. *Folksonomy - coinage and definition*. Thomas Vander Wal, 2007, 2 February 2007 [cited 21.03 2008]. Available from http://vanderwal.net/folksonomy.html.

Wall, Aaron. 2008. *Search engine history*. Aaron Wall 2008 [cited 25.02 2008]. Available from http://www.searchenginehistory.com/.

Wardrip-Fruin, Noah. , and Nick Montfort, eds. 2003. *The New Media Reader*. Cambridge, MA: The MIT Press.

Weiss, G. 1999. Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence.