

2009-07-01

## The genesis and emergence of Web 3.0: a study in the integration of artificial intelligence and the semantic web in knowledge creation

David Mulpeter

Technological University Dublin, [brendan.tierney@tudublin.ie](mailto:brendan.tierney@tudublin.ie)

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomdis>

 Part of the [Computer Engineering Commons](#)

---

### Recommended Citation

Mulpeter, David, "The genesis and emergence of Web 3.0: a study in the integration of artificial intelligence and the semantic web in knowledge creation" (2009). *Dissertations*. 23.

<https://arrow.tudublin.ie/scschcomdis/23>

This Dissertation is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin.

For more information, please contact

[yvonne.desmond@tudublin.ie](mailto:yvonne.desmond@tudublin.ie), [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie),

[brian.widdis@tudublin.ie](mailto:brian.widdis@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)

# **The Genesis and Emergence of Web 3.0**

## **– A study in the integration of Artificial Intelligence and the Semantic Web in Knowledge Creation**

**David Mulpeter**

A dissertation submitted in partial fulfilment of the requirements of  
Dublin Institute of Technology for the degree of  
M.Sc. in Computing (Knowledge Management)

**July 2009**

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Knowledge Management), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

***Signed:*** \_\_\_\_\_

***Date:***                      ***29 July 2009***

# 1 ABSTRACT

The web as we know it has evolved rapidly over the last decade. We have gone from a phase of rapid growth as seen with the dot.com boom where business was king to the current web 2.0 phase where social networking, Wiki's, Blogs and other related tools flood the bandwidth of the world wide web.

The empowerment of the web user with web 2.0 technologies has led to the exponential growth of data, information and knowledge on the web. With this rapid change, there is a need to logically categorise this information and knowledge so it can be fully utilised by all. It can be argued that the power of the knowledge held on the web is not fully exposed under its current structure and to improve this we need to explore the foundations of the web.

This dissertation will explore the evolution of the web from its early days to the present day. It will examine the way web content is stored and discuss the new semantic technologies now available to represent this content. The research aims to demonstrate the possibilities of efficient knowledge extraction from a knowledge portal such as a Wiki or SharePoint portal using these semantic technologies. This generation of dynamic knowledge content within a limited domain will attempt to demonstrate the benefits of semantic web to the knowledge age.

**Key words:** *Semantic, Web 3.0, Artificial Intelligence, Knowledge Creation, MetaKnowledge*

## **ACKNOWLEDGEMENTS**

I would like to express my sincere thanks to my supervisor Deirdre Lawless for her advice, support and guidance throughout the project. The input and insights she gave at various stages of the project were invaluable and key to the successful completion of the project.

I would also like to thank those colleagues who took part in the pilot, surveys and interviews that were central to the research. Among these colleagues, I want to give particular mention to Padraig Lennon, a fellow KM practitioner who provided valuable feedback on various aspects of the project.

Thanks must also be given to members of the open source community who helped and assisted with the semantic technologies.

To my family and friends; thank you for the continued encouragement and support throughout the project, it was very much appreciated.

Finally, I would like to thank my girlfriend Eimear for her love and patience that she has shown over the last number of months, it made everything that little bit easier.

## TABLE OF CONTENTS

<b>1</b>	<b>ABSTRACT.....</b>	<b>II</b>
	<b>TABLE OF FIGURES .....</b>	<b>VII</b>
<b>1.</b>	<b>INTRODUCTION .....</b>	<b>1</b>
1.1	BACKGROUND .....	1
1.2	RESEARCH PROBLEM .....	2
1.3	RESEARCH OBJECTIVES .....	2
1.4	RESEARCH METHODOLOGY .....	3
1.5	SCOPE AND LIMITATIONS.....	4
1.6	ORGANISATION OF THE DISSERTATION .....	5
<b>2</b>	<b>GENESIS – THE CREATION OF THE WEB.....</b>	<b>6</b>
2.1	INTRODUCTION.....	6
2.2	PRECURSOR TO THE WORLD WIDE WEB - THE INTERNET .....	6
2.3	THE BIRTH OF THE WORLD WIDE WEB .....	7
2.4	CONCLUSION .....	11
<b>3</b>	<b>THE EVOLUTION OF THE WEB .....</b>	<b>12</b>
3.1	INTRODUCTION.....	12
3.2	WEB 1.0 - THE INFORMATION PORTAL.....	12
3.2.1	<i>Web 1.0 Features .....</i>	<i>13</i>
3.2.2	<i>Web 1.0 Structure .....</i>	<i>14</i>
3.3	WEB 2.0 – THE WEB AS A PLATFORM .....	15
3.4	WEB 2.0 TECHNOLOGY .....	16
3.4.1	<i>Wiki's.....</i>	<i>17</i>
3.4.2	<i>Blogs.....</i>	<i>18</i>
3.4.3	<i>RSS.....</i>	<i>18</i>
3.4.4	<i>Tagging.....</i>	<i>19</i>
3.4.5	<i>MashUps.....</i>	<i>19</i>
3.5	WEB 2.0 COMMUNITY .....	20

3.6	WEB 2.0 WITHIN BUSINESS.....	22
3.7	CONCLUSION .....	24
<b>4</b>	<b>THE CONTEMPORARY WEB – GOOGLE AS THE GATEKEEPER.....</b>	<b>25</b>
4.1	INTRODUCTION.....	25
4.2	GOOGLE’S DOMINANCE.....	25
4.3	GOOGLE’S PAGERANK LIMITATION .....	28
4.3.1	<i>Search Engine Optimization.....</i>	<i>29</i>
4.3.2	<i>The Deep Web.....</i>	<i>31</i>
4.4	GOOGLE CHALLENGE .....	32
4.5	CONCLUSION .....	33
<b>5</b>	<b>WEB 3.0 – THE INTELLIGENT WEB .....</b>	<b>34</b>
5.1	INTRODUCTION.....	34
5.2	THE SEMANTIC WEB .....	34
5.3	SEMANTIC WEB TECHNOLOGIES .....	37
5.3.1	<i>The Data Model (RDF).....</i>	<i>38</i>
5.3.2	<i>Structure &amp; Semantics (RDFS, OWL, RIF).....</i>	<i>41</i>
5.3.3	<i>The Query Language (SPARQL).....</i>	<i>44</i>
5.3.4	<i>Linked Data .....</i>	<i>45</i>
5.4	SEMANTIC SEARCH.....	47
5.5	SEMANTIC WEB OUTSTANDING ISSUES .....	48
5.6	CONCLUSION .....	50
<b>6</b>	<b>META KNOWLEDGE FRAMEWORK.....</b>	<b>51</b>
6.1	INTRODUCTION.....	51
6.2	THE KNOWLEDGE MANAGEMENT VIEWPOINT .....	51
6.3	COMMUNITY METADATA .....	52
6.4	COGNITIVE METADATA .....	53
6.5	THE TAXONOMY COMPROMISE .....	54
6.6	CONCLUSION .....	56
<b>7</b>	<b>SEMANTIC WIKI PILOT .....</b>	<b>57</b>
7.1	INTRODUCTION.....	57
7.2	THE ORGANISATION .....	57

7.3	EXISTING TECHNOLOGY .....	57
7.3.1	<i>ITECH (SharePoint)</i> .....	57
7.3.2	<i>K-Portal (TWiki)</i> .....	58
7.4	SEMANTIC WIKI .....	62
7.4.1	<i>Product Choice</i> .....	62
7.4.2	<i>Semantic MediaWiki Design</i> .....	63
7.4.3	<i>Semantic MediaWiki MetaKnowledge Design</i> .....	65
7.5	EXPERIMENT .....	68
7.5.1	<i>Experiment User Base</i> .....	69
7.6	EVALUATION .....	70
7.7	CONCLUSION .....	87
<b>8</b>	<b>CONCLUSIONS AND FUTURE WORK .....</b>	<b>88</b>
8.1	INTRODUCTION .....	88
8.2	RESEARCH DEFINITION & RESEARCH OVERVIEW .....	88
8.3	CONTRIBUTIONS TO THE BODY OF KNOWLEDGE .....	90
8.4	EXPERIMENTATION, EVALUATION AND LIMITATION .....	91
8.5	FUTURE WORK .....	92
8.6	CONCLUSION .....	93
	<b>BIBLIOGRAPHY .....</b>	<b>95</b>
	<b>APPENDIX A .....</b>	<b>107</b>



## TABLE OF FIGURES

Figure 2.1 Tim Berners-Lee First Web Browser (W3C).....	11
Figure 3.1 Web Structure (IBM Arasu <i>et al.</i> , 2002) .....	16
Figure 3.2 Web 1.0 vs Web 2.0 Technologies (O'Reilly 2005) .....	28
Figure 3.3 Reshaping Your Business with Web 2.0 ().....	21
Figure 3.4 Social Networking Survey .....	22
Figure 3.5 The Information Lifecycle (Hirsch <i>et al.</i> , 2009) .....	23
Figure 4.1 US Search Share (Hitwise 2007) .....	27
Figure 5.1 The Semantic Web (Olsson 2007) .....	35
Figure 5.2 The Semantic Stack (W3C) .....	39
Figure 5.3 RDF Representation .....	40
Figure 5.4 RDF Graph .....	41
Figure 5.5 RIF Interaction (Kiefer 2007) .....	44
Figure 5.6 Linked Open Data Cloud 2007 .....	47
Figure 5.7 Linked Open Data Cloud 2008 .....	48
Figure 6.1 General Application Scenario (Schueler <i>et al.</i> , 2008) .....	55
Figure 7.1 K-Portal Wiki Design .....	59
Figure 7.2 Tagging Within Organisation Knowledge Portal .....	60
Figure 7.3 RSS Feeds linked to Organisation Knowledge Portal .....	61
Figure 7.4 Semantic MediaWiki Development .....	63
Figure 7.5 Semantic MediaWiki Architecture (Volkel <i>et al.</i> , 2006).....	64
Figure 7.6 Semantic MediaWiki Categories .....	66
Figure 7.6 Semantic MediaWiki Search Help Page.....	81

## TABLE OF TABLES

<b>Table 4.1 - Table 4.1 – Search Engine Launch Dates.....</b>	<b>27</b>
<b>Table 7.1 - Table 7.1 - SMW: a variation of RDF.....</b>	<b>65</b>
<b>Table 7.2 - Ontology Construct in Semantic MediaWiki .....</b>	<b>65</b>

# 1. INTRODUCTION

## *1.1 Background*

This project is an evaluation of the web from its inception through to the present day analysing it through the eyes of a Knowledge Management practitioner. Its formation began with the creation of the ARPNET network in 1969 by the US Department of Defence. This network enabled computers to communicate with each other using modems connected to leased telephone lines. ARPNET opened the door for the creation of similar networks around the world, and the connection of all these networks led to the creation of the Internet. It was on this platform that Tim Berners-Lee developed his World Wide Web project at the CERN institute in 1990, a project based on linking documents on a network using the hypertext protocol HTML. This project grew within the academic community over the following years as its potential became apparent. In 1993 CERN officially opened the World Wide Web to everyone free for use, and Guilli and Signorni (2005) estimated that as of 2005, 11.5 billion pages existed on the web, all of which are electronically accessible by the users of the web.

The World Wide Web has irrevocably changed the way we all share and access information and has helped create a culture of participation and collaboration among people. This has been achieved through continuous evolution which has been driven by the demands of the web user. This has brought its development through various iterations over the last two decades. From its origins as an information portal, it has progressed to a stage where the web is now regarded as a platform. This platform, commonly referred to as Web 2.0, operates on the architecture of participation where web users are empowered with new web technologies that enable them to create and share content more effectively.

However, the continued growth in the content on the web and changes in the way we use it, is making it increasingly difficult to find, access and maintain the information of use for the collection of different web users. In the pursuit of a richer user experience for the web user, the underlying vision for the World Wide Web has suffered. A gap has developed between the presentation of the web content in natural language and

technology's ability to interpret this content (Klein *et al.*, 2002). This gap presents a bottleneck for knowledge management as web searches for information content on the Web face potential issues with search engine algorithms struggling to deal with the large volume of natural language content that exists.

The emergence of the Semantic Web presents a possible solution to these issues through a re-engineering of the web. Tim Berners-Lee, founder of the Web, presents his vision of the Semantic Web as an extension of the current web where content is machine-readable and well defined enabling new functionality that will extend beyond the current capabilities of the Web (Berners-Lee *et al.*, 2001). It is this potential solution that this research seeks to assess and evaluate using knowledge management tools within an organisation.

## ***1.2 Research problem***

This primary objective of the research is to determine if the semantic web technologies aid in the creation and retrieval of web content in the context of Knowledge Management. Secondly, the research will look at the field of Meta Knowledge and evaluate how a framework can be developed to intelligently tag content with metadata.

The research will look in depth at evaluating current web technologies used within the organisation and how content is managed on these systems. It will investigate the possibility of generating a Semantic Wiki with AI tools to generate the relevant knowledge content for the end user.

## ***1.3 Research objectives***

The following objectives have been achieved throughout the dissertation and contributed to the final outcome:

- Performance of a literature review to examine the evolution of the Web from its beginnings to the present day.
- Performance of a literature review of current web search technologies and evaluated their strength and weaknesses.

- Consideration and evaluation existing Web 3.0 tools in the context of Knowledge Management.
- Development of a framework to represent Knowledge about the Knowledge – Meta Knowledge.
- Investigation of the feasibility of generating a Semantic Wiki with AI tools to generate the relevant knowledge content for the end user.

#### ***1.4 Research methodology***

A number of methodologies were used to carry out the research undertaken as part of the project. Both primary and ancillary data were collected for this purpose. The ancillary data consisted of data from the literature reviewed from books, journals and the Internet while the primary data took the form of information/results collected from informal interviews and a survey conducted within the organisation where the experiment was held.

The initial phase of the research consisted of collecting data from the literature review. The literature review concentrated on all aspects of the web from its creation through its various iterations and its future development. This phase helped establish and generate the research ideas for the project. It helped in developing an understanding and insight into the previous research carried out on the subject of the Semantic Web and the results of these studies. Following the literature review, an analysis was undertaken on the semantic technology available for conducting the research project. This analysis was required to determine which software would best fit the defined research problem. Testing of different semantic applications was carried out with comparisons made on key features. This testing augmented the literature review by providing a better understanding of the underlying semantic technology.

This software analysis review was followed by the final phase of the methodology, the creation of an experiment with a corresponding survey to collect both quantitative and qualitative data. The experiment involved the creation of a Semantic Wiki for

consideration by a number of key user groups within the organisation who currently utilise a standard Wiki as their Knowledge Portal. The Semantic Wiki pilot ran for three months after which time the user groups were asked to complete a survey which was embedded in the application. Qualitative feedback on the experiment was obtained through a number of informal interviews. The qualitative and quantitative data gathered from the survey and informal interviews was then carefully analyzed in the context of the research problem.

### ***1.5 Scope and limitations***

The scope of the research was limited to an organisation and to a specific department within the organisation. The IT department has introduced a Wiki technology previously which has implications for the new semantic technology introduced as part of the survey. Preconceived opinion on the current Wiki technology could affect the autonomy of any comparison with the semantic technology. The experiment was extended to business users who do not utilise the incumbent Wiki technology but these users chose not to participate in either the experiment or the survey. Possible reasons for this decision are discussed in chapter 7.

The content created on the new semantic technology was also limited to trading system content which was migrated from the existing Wiki technology. The structure and layout of the content on a site level mirrored the existing Wiki to provide a fair comparison. The possibility of organising the content differently was considered but it was felt this could weaken the research.

The research is based on a survey research method which has some limitations particularly in relation to the user base selection bias. The relationship of the researcher to the user base may alter the results as per the respondent's interpretation of what the researcher is trying to achieve.

## ***1.6 Organisation of the dissertation***

Chapter 2 discusses the history of the web beginning with the origins of the Internet through to the invention of the World Wide Web. It also explores the underlying technologies and concepts that helped in its creation.

Chapter 3 examines the evolution of the web from its early academic background through its various iterations over the last two decades. Key milestones in the evolution are outlined and discussed.

Chapter 4 describes the contemporary web and specifically Google's role within it. The architecture and structure of the web are discussed in terms of the web search. The strengths and weaknesses of Google are argued and alternatives are discussed.

Chapter 5 is concerned with the next iteration of the web, the Semantic Web, and how it will change the web. Semantic Technologies are discussed with a particular emphasis on how they aid in the creation and retrieval of content. Outstanding issues with the Semantic Web are also outlined.

Chapter 6 introduces the concept of the meta-knowledge framework and discusses it in terms of semantic technologies

Chapter 7 describes the Enterprise Semantic Wiki pilot which took place within the IT department of a financial services organisation as part of this research. The initiation, scope, aims and challenges which presented themselves in the running of the pilot are discussed throughout the chapter.

Chapter 8 discusses the conclusions of the project and presents suggestions for future areas of research which could build on the findings of this research.

## **2 GENESIS – THE CREATION OF THE WEB**

### ***2.1 Introduction***

Twenty years ago, Tim Berners-Lee submitted a document to his supervisor Mike Sendall entitled “Information Management: a Proposal”. This document formed the basis for the creation of the World Wide Web which was developed over the following months at the CERN institute in Switzerland.

“The WorldWideWeb (WWW) project aims to allow links to be made to any information anywhere” (Berners-Lee 1990).

This chapter discusses the history of the Internet and how this lead to the creation of the World Wide Web. It provides the starting point for this research by assessing the drivers for the creation of the World Wide Web and how technology helped achieve this. This assessment is relevant to the research as it examines how the World Wide Web came into existence and outlines how its original aim mirrors the ambitions of the Semantic Web.

### ***2.2 Precursor to the World Wide Web - The Internet***

The Internet began life as a network called ARPANET (Advanced Research Projects Agency Network), a project commissioned by the United States Department of Defence to study country-wide data communication. The project officially began in 1968, and by the end of 1969 four computers (called hosts) located in 4 different cities across mainland America were connected on the ARPANET network using the 1822 network protocol (Heflin 2001). From these initial small steps, ARPANET grew in size and new functionality such as email and FTP began to appear on the network, email itself accounting for 75% of the ARPNET traffic by 1973.

While ARPNET proved successful in connecting machines, the underlying 1822 protocol did not prove to be adequate in dealing with multiple connections that were



now appearing on the network (Cerf and Kahn 1974). To overcome these issues, new network protocols were developed leading to the subsequent development of Transmission Control Protocol (TCP) and Internet Protocol (IP), known commonly as the TCP/IP network protocol in 1975. Testing was carried out with the protocol for a number of years after its development and on the 1st January 1983, the ARPNET network was successfully migrated onto the TCP/IP network protocol (Ruthfield 1995).

TCP/IP remains the protocol of choice for almost all Internet traffic. High level protocols including File Transport Protocol (FTP), the Simple Mail Transfer Protocol (SMTP), and the Hypertext Transfer Protocol (HTTP), all depend on the underlying TCP/IP protocol in order to transfer files, perform remote logins, transfer electronic mail, and exchange web documents using the Internet (Heflin 2001).

ARPNET, which remained part of the US Department of Defence network, was superseded by new networks developed by academic institutes and other government agencies worldwide. These new networks fuelled the growth of the Internet throughout the 80's supporting a broad community of researchers and developers across the academic spectrum. Towards the end of the decade, countries around the world began setting up national organisations to manage the distribution of IP addresses as networks continued to grow, and it was during this time that CERN, a research laboratory based in Switzerland received its first router from Cisco and opened its first external TCP/IP connections in 1989, a milestone that enabled Tim Berners Lee, a researcher based at CERN to develop his idea of the World Wide Web on this architecture (Berners-Lee and Cailliau 2004).

### ***2.3 The Birth of the World Wide Web***

As mentioned in the introduction, a document entitled "Information Management: a Proposal" which proposed "a large hypertext database with typed links" for managing information was put forward by Tim Berners-Lee for consideration at the CERN research institute in March 1989. It was to provide the basis for the development of what would become the World Wide Web. Below are the main objectives of the proposal (Relihan, Cahill and Hinchey 1994).

- *the provision of a simple protocol for requesting human readable information stored in remote systems accessible using networks.*
- *to provide a protocol by which information could automatically be exchanged in a format common to the information supplier and the information consumer*
- *the provision of some method of reading text (and possibly graphics) using a large proportion of the display technology in use at CERN at that time*
- *the provision and maintenance of collections of documents, into which users could place documents of their own.*
- *to allow documents or collections of documents managed by individuals to be linked by hyperlinks to other documents or collections of documents.*
- *the provision of a search option, to allow information to be automatically searched for by keywords, in addition to being navigated to by the following of hyperlinks*
- *to use public domain software wherever possible and to interface to existing proprietary systems.*
- *to provide the necessary software free of charge.*

The fundamental concept behind the proposal was to use hypertext as a means of organising a distributed document system. Hypertext, a technology invented in the 60's, refers to text with cross-references (also known as hyperlinks) to other text that enables the reader to follow the linked text in a non-sequential manner.

With hypertext providing the means for the document distribution, Berners-Lee then had to develop a mechanism for addressing documents on different machines (Web Browser), a separate protocol that allowed computers to request documents (HTTP) and finally a intelligible language that could be used to describe the documents (HTML).

To help achieve this Tim Berners-Lee set up the World Wide Web Project with this initial post on the on the alt.hypertext newsgroup.

*“The WorldWideWeb (WWW) project aims to allow all links to be made to any information anywhere. [...] The WWW project was started to allow high energy*

*physicists to share data, news, and documentation. We are very interested in spreading the web to other areas, and having gateway servers for other data. Collaborators welcome!" —from Tim Berners-Lee's first message (Tim Berners Lee 1990).*

Following on from this rally call, the technologies needed to make the World Wide Web a reality were developed by Tim Berners-Lee along with a number of key collaborators. These key technologies are outlined below and while other technologies played a part in the growth of the web, these technologies are regarded as the core technologies of the World Wide Web.

- **HTML (HyperText Markup Language)** - HTML is the mark-up language in which the World Wide Web hypertext documents are written. It is the language that brings the web together through hypertext links and clickable images. A web browser is used to render the HTML for presentation on the screen (Raggett and Jacobs 1999).
- **HTTP (HyperText Transfer Protocol)** – This is the protocol that facilitates the transfer of information on the World Wide Web. The protocol transmits requests and responses between clients and servers. The client submit the requests through a URL (Uniform Resource Locator) on a given web browser whose web servers then interpret to resource to determine which content to deliver to the client. The URL is the mechanism for addressing objects on the Web. It specifies the global address of web documents and web resources. More commonly known as the web address, it consists of the protocol identifier, typically HTTP and the resource name which usually points to a domain name where the resource is located (Fielding *et al.*, 1999).
- **Web Browsers** – These are the tools used to browse the World Wide Web. Web Browser client software programs such as Firefox or Internet Explorer receive and interpret data from Web Servers and display results. A fundamental feature of web browsers is that they are consistent across all types of computer platforms so users have the ability to access information from many different

types of computers thus fulfilling one of the aims of Berners-Lee original proposal.

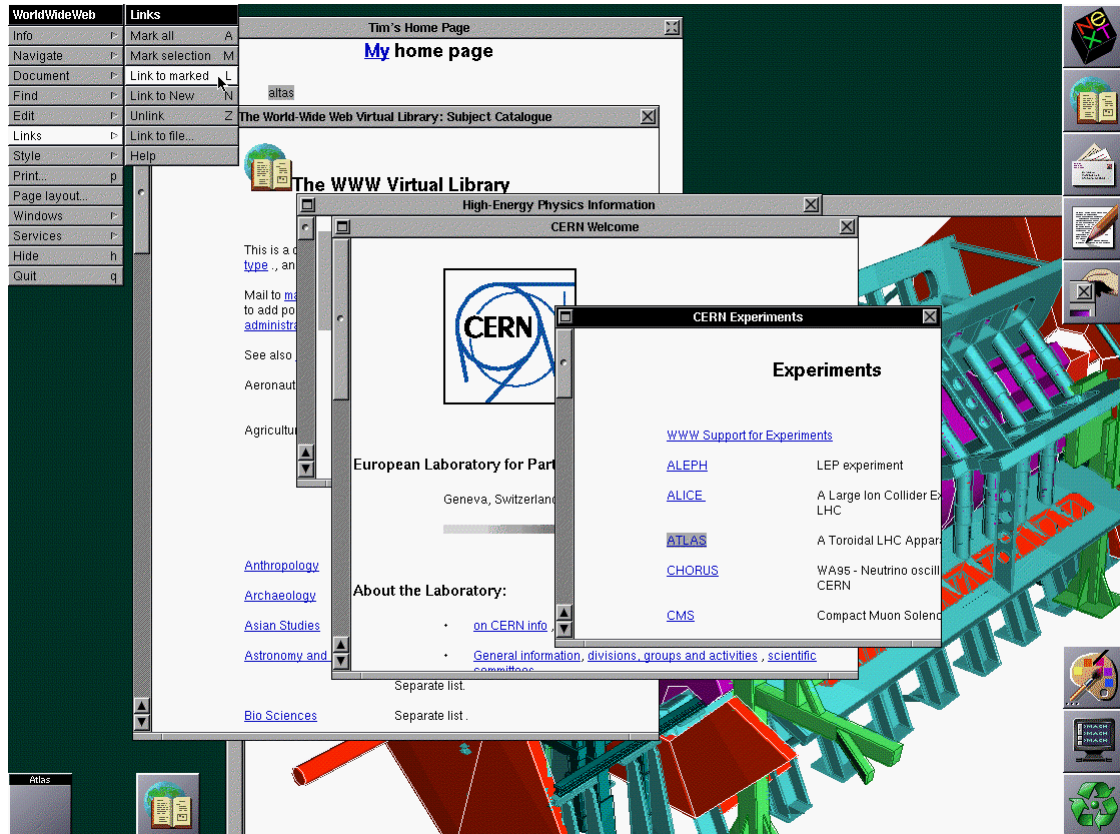


Figure 2.1 Tim Berners-Lee First Web Browser (World Wide Web Consortium W3C)

With the Web now firmly established in the academic community, it was the development of the graphical web browser in 1993 that led to the exponential growth of the web user base. PC owners were now able to browse the web from their home computers with these new web browsers coupled with the new dial-up Internet access now on offer from the telecommunication providers (Vetter, Spell and Ward 1994).

Though the World Wide Web was created over a short period of time in the early 90's, its transport mechanism, the Internet, was still limited to research, education, and government use. However, the commercial potential of the web was identified when the corporate world identified it as a vehicle for which to reach the consumer (Hoffman *et al.*, 1995). This stimulated the rapid growth of the web leading to the

creation of a new industry within the field of technology and the first version of the web - Web 1.0.

## ***2.4 Conclusion***

This chapter presented the events and technologies that lead to the creation of the World Wide Web. It provides a context on how the Web reached its current standing analysing how it evolved from its academic background. The underlying technologies that helped created the World Wide Web are still present on the web of today. An understanding of these technologies is required in order to assess the strengths of the new semantic technologies which form part of the research project.

## **3 THE EVOLUTION OF THE WEB**

### ***3.1 Introduction***

This chapter examines the evolution of the web, from its early academic background through its various iterations over the last two decades. Key features and technologies that evolved over time are discussed in order to determine how the Web progressed to its current state. Features such as the structure are evaluated and particular attention is given to the core technologies of Web 2.0, which helped bring about a significant change in the way we use the web. The chapter illustrates how the structure and technologies of the web relate to the way content is represented. The chapter concludes with an examination of the web from this social context. The impact of the social web is discussed in terms of the web community and the enterprise. In terms of the research problem, it provides detail on why the Web has evolved in the manner it has while also describing in detail some of the key technologies and features used as part of the research experiment.

### ***3.2 Web 1.0 - The Information Portal***

Web 1.0 is the term used to refer to the web as it existing before O'Reilly (2005) introduced the term Web 2.0 at the Media Web Conference in 2004. Web 1.0 can be used to define the time frame (1990-2000) where the web was used primarily as an Information Portal. It allowed for the sharing of files, information and software over the Internet. The web was divided into working directories and in theory everyone had their own space in cyberspace (Cormode and Krishnamurthy 2008). It was a top down model with the control of web content in the hands of webmasters who designed and built the websites, most of which were static and non-interactive. The pioneers during this period were mostly individuals who created their own websites for posting their thoughts and ideas to a wider audience. College enterprises such as Yahoo also formed around this time setting up websites to catalogue the growing number of new pages and websites appearing (Etzioni 1996). Corporations later joined these early adopters and started to utilize the web to distribute information to potential customers, which in turn sparked the dot com boom.

### **3.2.1 Web 1.0 Features**

Once established, the World Wide Web underwent almost constant evolution, which was driven by the continuous innovation of new technologies. Building on the core technology of HTML, dynamic technologies such as Shockwave, Flash and Dynamic HTML were quickly developed to improve the experience for the web surfer taking advantage of the new web browsers now available. Marketing and the user experience gained importance as the corporate enterprises built presence on the web and visual design was seen as the primary way to differentiate content and draw attention to the website. Cascading Style Sheets (CSS) were developed and utilised to define the style of the websites by decoupling the presentation tags of the webpage from within the HTML. This enabled the webmaster to separate the web content from the web presentation style and thus create a branded style template for all pages for application to both internal and external websites (Mace *et al.*, 1998). While attention was given to the ongoing development of these new style sheets and other design improvements, the underlying content retained its original HTML structure and remained unchanged.

The focus of Web 1.0 features on the superficial aspects of web sites was to the detriment of other areas of potential for the web such as knowledge management. In terms of KM, Web 1.0 provided the technology platform on which to share knowledge content but the limitations on content creation limited the potential. There was also no context in which to connect knowledge workers, as the portals such as corporate Intranets did not emerge until near the end of the decade. The techno-centric nature of the early web may not have best suited a KM initiative due to the lack of focus on the people and process components. It is for these reasons that Web 1.0 is commonly referred to as the Information Portal, providing the vehicle for the representation and creation of content but not for the distribution and identification of this content (Wilson 2002).

### ***3.2.2 Web 1.0 Structure***

The webmasters who created the pioneering websites did so with a free rein. The new content they created was not well structured and silos of information began to occur on the World Wide Web.

A collaborative study carried out by researchers at AltaVista, Compaq, and IBM to analyse the structure of the web in the year 2000 yielded some interesting results in this regard on the inter connectivity of Web 1.0. The study analysed the connectivity of more than 1.5 billion links on over 200 million Web pages (Arasu, Novak, Tomkins and Tomlin 2002).

The resulting topography revealed a web structure that resembled a bow tie. It consisted of three regions, a core and two peripheral regions. The core region represents pages that are strongly connected by extensive cross-linking to other pages within the core. Links on core pages enable Web users to move relatively easily from one page to another within the core. They are also the links most likely to be caught by a search engine web crawler. The left region represented pages not linked from other pages but which themselves linked to pages in the core region. These pages cannot be reached from the core as no backward links exist to these pages from the core. These Web pages were either new or obscure pages that had yet to attract interest from the larger Web community. The right region represented pages that can be accessed via links from pages within the core region but these pages do not link back into the core region. Examples of pages that exist in the right bow are commercial pages consisting entirely of internal links and thus serve as destinations in themselves. As Figure 3.1 illustrates, an additional 20% of the web pages remain disconnected where no links exist to the three cores of the web.



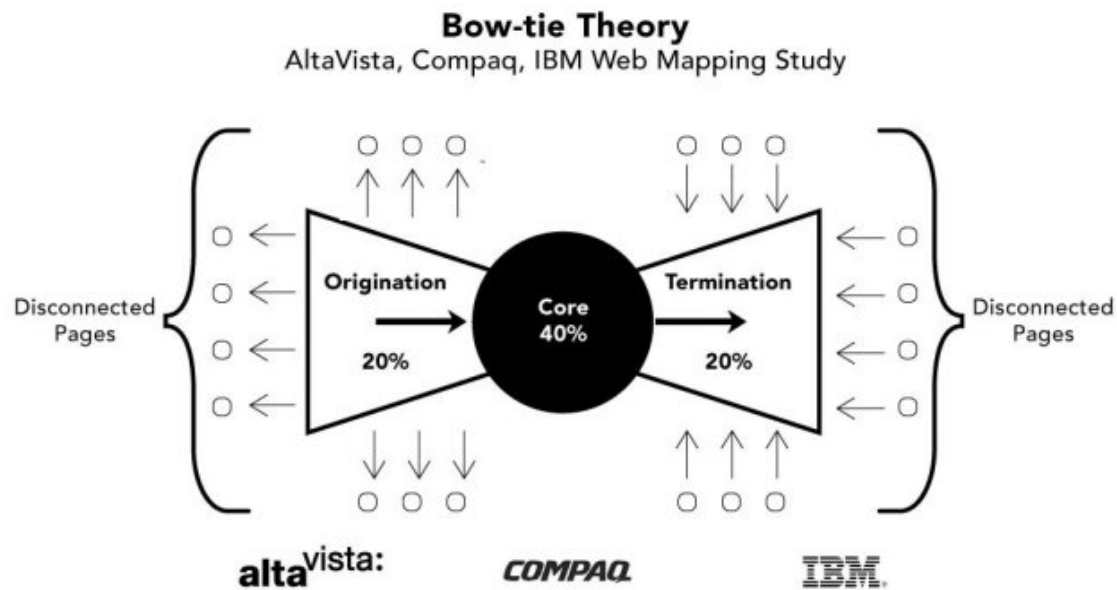


Figure 3.1 Web Structure – IBM (Arasu, Novak, Tomkins and Tomlin 2002)

Kinsella *et al.*, (2008) state that people implicitly annotated web resources by creating links to interesting Web sites onto their personal homepages in the early days of the Web. This early form of tagging content meant HTML anchor text linking pages was spread across the entire Web with no central repository existing. This sporadic spread of HTML anchor text contributed to the Bow-Tie structure and was the catalyst for the creation of the search engine to put some form of order on the Web. Google founders Sergey Brin and Larry Page recognised the importance of anchor text as a rich source of Web page annotations to improve the search quality (Brin and Page 1998). This anchor text would form the basis for the PageRank algorithm which Google invented. While this “*web 1.0 tagging*” was a genuine attempt at linking up the web, it was still limited to those few users that create their own Web pages, which considerably restricted the input of the entire web community.

### 3.3 Web 2.0 – The Web as a Platform

*“Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them.”* (Musser and O’Reilly, 2006).

This is how commentators who first coined the phrase Web 2.0 described the new media that was emerging on the Web. Commentators, such as Singel (2005) quoted Mayfield, a CEO of a software Wiki Solutions Company who states that "*Web 1.0 was commerce. Web 2.0 is people*". Re-enforcing the change taking place on the Web, McLean in his paper on the new Web, suggests that: "*WEB 2.0 is the catch-all descriptor for what is essentially much more dynamic Internet computing*" (McLean, 2007).

Cormode and Krishnamurthy (2008) argue that the core distinction between Web 1.0 and Web 2.0 is the role of the content creator. There are a minimal number of content creators in Web 1.0 with the mainstream users acting as consumers of content. Contrast this with Web 2.0 where any participant can be a content creator. As part of this evolution, numerous technologies were created to maximize the potential for content creation. This was initially driven by dot com consumers who required improved web services for carrying out on-line transactions which quickly developed into a social platform as peoples empowered themselves to become participants on the web. Web content creation no longer lay solely in the hands of the webmaster. Now any web participant could create content with new sandbox technologies, which enabled the masses to contribute in a very real way.

### **3.4 Web 2.0 Technology**

Analyzing Web 2.0 there are several popular applications, or application classes, that are well known and in some aspects define the functional features of Web 2.0. What has been proven with these Web 2.0 technologies is that simplicity wins and the potential for application extendibility has a bearing on its success. All these application are network-centric i.e. they are not tied to a specific Operating System or device and as such as globally linked with no obvious boundaries. The below diagram illustrates some of the main applications of Web 2.0 that have evolved from earlier incarnations of technology that existed during the Web 1.0 era.

<b>Business Area</b>	<b>Web 1.0</b>		<b>Web 2.0</b>
<i>Advertising</i>	DoubleClick	➡	Google AdSense
<i>Photo Sharing</i>	Ofoto	➡	Flickr
<i>Content Distribution</i>	Akamai	➡	BitTorrent
<i>Music File Distribution</i>	Mp3.com	➡	Napster
<i>Encyclopaedia</i>	Britannica Online	➡	Wikipedia
<i>Personal Presence</i>	Personal websites	➡	Blogging
<i>Traffic Metrics</i>	Page view	➡	Cost Per Click
<i>Content Management</i>	Content Management Systems	➡	Wikis
<i>Content Organisation</i>	Directories(taxonomy)	➡	Tagging (folksonomy)
<i>Content “Push”</i>	Stickiness	➡	Syndication

Figure 3.2 Web 1.0 vs Web 2.0 Technologies (O'Reilly 2005)

The core technologies of Web 2.0 that provide the sandbox functionality are outlined below. This suite of tools put into practise the principals of Web 2.0 and helps to create a more user-centric, participative model for the web.

### 3.4.1 Wiki's

A Wiki is a website that allows users to create, edit and link web pages easily. The best-known example is the public Wikipedia website, which in its own words defines a

Wiki as “*collaborative technology for organizing information on Web sites*” (source Wikipedia). The term itself, Wiki originates from the Hawaiian word meaning fast, which is a key concept behind the technology. Wikis is a tool suitable for collaborative knowledge sharing both locally and across multiple locations. They do, however, rely on cooperation, checks and balances of its members, and a belief in the sharing of ideas (Macaskill and Owen 2006). Semantic Wikis, the next step in Wiki technology will be discussed later in this dissertation.

### **3.4.2 Blogs**

A Blog is a return to the days of Web 1.0 where people maintained their own web space. The term Blog is the abbreviation of the term WEB-log, otherwise known as the new personal diary. The diaries contain entries, which can include social commentary, descriptions of events, and other items such as graphics or video. Entries are subject oriented or personal in nature and are displayed in reverse-chronological order. The Blog is a form of citizen journalism and allows those so inclined, to share their views and opinions with those who have an interest (Nardi *et al.*, 2004). Other types of Blogs such as Corporate Blogs and Video Blogs also exist and form part of the community of blogs along with private citizens. The term Blogosphere has been coined to describe this community of Blogs that exist on the web and Wordpress, the tool of choice for most bloggers has 5.2 million Blogs registered online as of the beginning of 2009 (source Wordpress).

### **3.4.3 RSS.**

RSS (Really Simple Syndication) is a relatively new technology for pushing edited content to the end user. Content creators, be they large corporate websites or individual bloggers, syndicate their content as an RSS Feed. This then allows people to tune into information sources that interest them by subscribing to the relevant feeds. The content source for RSS feeds could be any form of Web content, from simply Blogs through to videocasts, on any web site that provides RSS feeds. RSS tools provide an interface to the content ensuring that users are automatically kept up to date with the latest content changes on their feeds, without the need to check each website for updates (Wusteman 2004).

#### **3.4.4 Tagging.**

A tag is a form of metadata used to describe web content. Content creators and viewers can tag all web content using tags that they think best describes the content. These tags, whether public or private, are the basis for new connections, links between different pieces of content, sharing something in common, via the tags. Web 2.0 websites such as Delicious provide tagging and other related services for end users, so they can easily categorise their websites of interest. In Web 1.0, tagging was used as a form of taxonomy with content categorised into directories as a form of classification. However, with Web 2.0, a looser approach is encouraged with no structure defined on the tags. This new personal approach to tagging has led to the practise being re-branded as folksonomy (Chopin 2008). Ontologies, the method for tagging content on the semantic web are described in later chapters of the dissertation.

#### **3.4.5 MashUps.**

Together with these new technologies there was a change in the way software was developed and rolled out on this new platform. O'Reilly (2005) and his collaborators advocated an “*end of the software release cycle*”, and instead placed on emphasis on operations as a core competency within a Web 2.0 enterprise. This emphasis on operations is promoted in order to support the release of small, frequent and rapid updates of data and software on a continual basis - a concept termed “*perpetual beta*”, as opposed to the traditional release cycles of shrink-wrapped software, which can take months or years.

While this model is different from the predominant tightly coupled corporate-sponsored web services, the loose coupling is seen as a key enabler of application growth on the open web platform. In this regard the Web 2.0 mindset is decidedly different from the traditional IT mindset, to such an extent, that some applications are designed for “*hackability*” and “*remixability*” with the end user treated as potential co-developers. An example which illustrates this mindset is Paul Rademacher's HousingMaps.com which combined Google Maps with the USA-based Craigslist of flats available for rent to map out rental locations on the on-line map (Yu *et al.*, 2008). This mash-up of two different applications would not have been possible without the

lightweight programming models that exposed the Application Programming Interfaces (API) of the underlying software.

### 3.5 Web 2.0 Community

Berners-Lee (1999, p. 123) in his book titled *Weaving the Web*, emphasises that the Web “*is more a social creation than a technical one*” and nowhere is this more evident than in the new generation of technologies that have been developed on the Web 2.0 platform. These technologies were developed using “*an architecture of participation*” and collaboration where end users were treated as co-developers in a transparent environment (O’Reilly 2005). Commentators such as Jack Maness see the key elements of Web 2.0 as user-centred participation in the creation of content and services where there is strong social communication between users embodying an innovative community spirit (Maness, 2006). Figure 3.3 illustrates how participation and contributions evolve through a contributory feedback loop where user contributions adds value to observers, who then participate based on this value in turn adding to the original contributor's value (Casarez *et al.*, 2008).

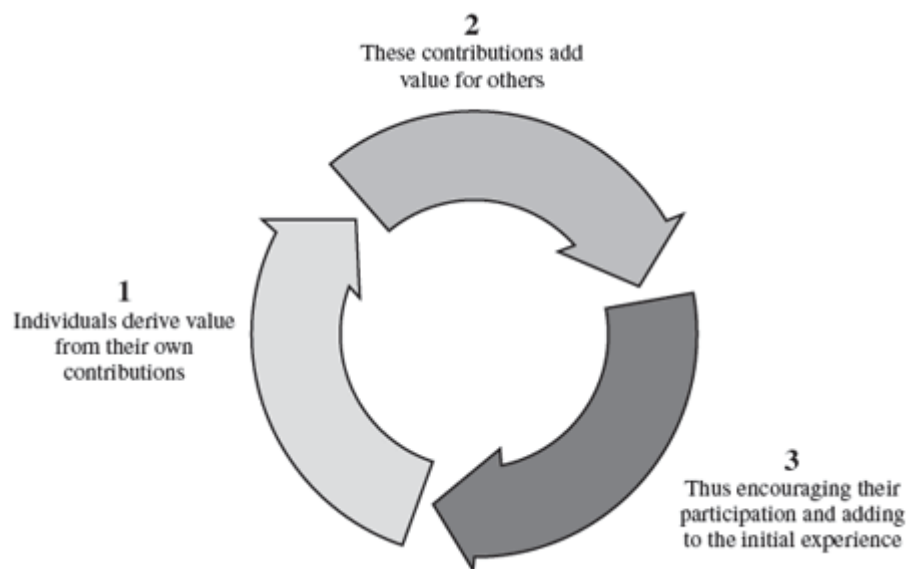


Figure 3.3 Contributory feedback loop (Casarez *et al.*, 2008)

Casarez *et al.*, (2008) views the contributory loop occurring through websites, taking users explicit and tacit activity into account. This activity delivers content, which

grows organically with every new user activity. It is this empowerment of web users through the structure of websites that enables the behaviour of systems to emerge naturally. Examples of this contribution cycle range across all applications on the Web 2.0 platform, be it a user contributing content on Wikis and Blog, or acting as a content reviewer on auction sites (eBay reputation ratings). In all these examples the user participating is involved in the continuous building of the Web 2.0 platform and the power of the “*wisdom of crowds*” comes into play (Surowiecki 2004).

The SPIRE project led by the University of Oxford carried out a detailed survey on Web 2.0 services in 2007 and found some interesting results in end user contributions in the new web community. Figure 3.4 illustrates end user behaviour across a range of Web 2.0 technologies with 20% of those who use Video Sharing sites such as youTube contributing in some form. Results for Social Networking sites such as mySpace are also positive with around a third of active users contributing. The most interesting results can be seen with blogging, and while the results do not show how successful Blogs are in terms of dissemination and collaboration, it does highlight that the writing and reading of Blogs is no longer a niche activity (White 2005).

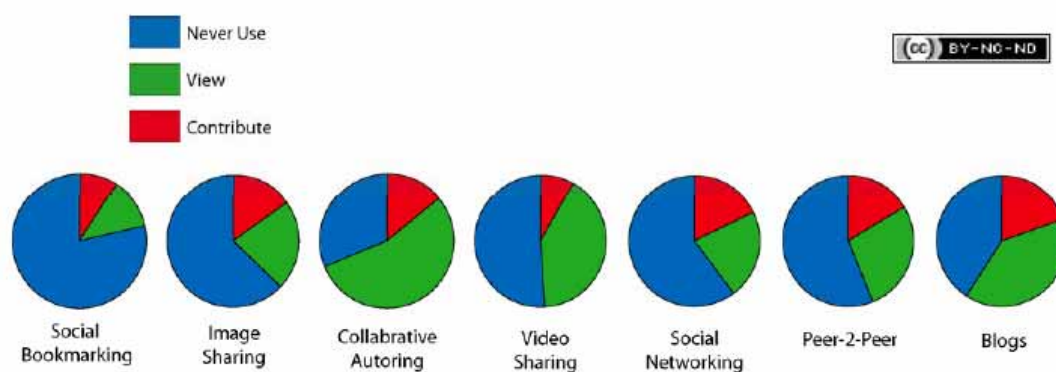


Figure 3.4 - Ratios of contribution to viewing for groups of services (White 2005)

The new focus on community within web technologies has strengthened the original concepts of the Web that Tim Berners-Lee outlined nearly 20 years ago. This change in emphasis from the personal-focus to the community-focus as part of the Web 2.0

movement has also had an effect within the workplace and this dissertation will now discuss these changes in further detail highlighting how it is changing work practices.

### 3.6 Web 2.0 within Business

Companies have recognised the potential of Web 2.0 solutions in solving business challenges and have begun to explore and apply the technologies and concepts of web 2.0 to business intranets. These applications and solutions are often referred to as enterprise 2.0, a term coined by Andrew McAfee (McAfee 2006). Tredinnick (2006) states that these technologies provide strong potential for the development and management of business information and knowledge assets by offering a more democratic way to exploit the information within organisations. This leads to a knock on effect on different areas within the business, such as project management, innovation processes, and knowledge management (Tapscott and Williams 2006). It is for these reasons that management consultancy firms have got on board highlighting the potential of these technologies in the business world. McKinsey, a leading management consultancy firm, has reported that the popularity and importance of these enterprise 2.0 applications is constantly growing (McKinsey, 2007).

The core concept of enterprise 2.0 can be summarised as a gradual progressive change in how the business content is processed. These changes range from the ways information is created and organized via its distribution and search to its application (Hirsch *et al.*, 2009). Within the corporate context, this has lead to a shift from a top-down to a bottom-up creation of content.

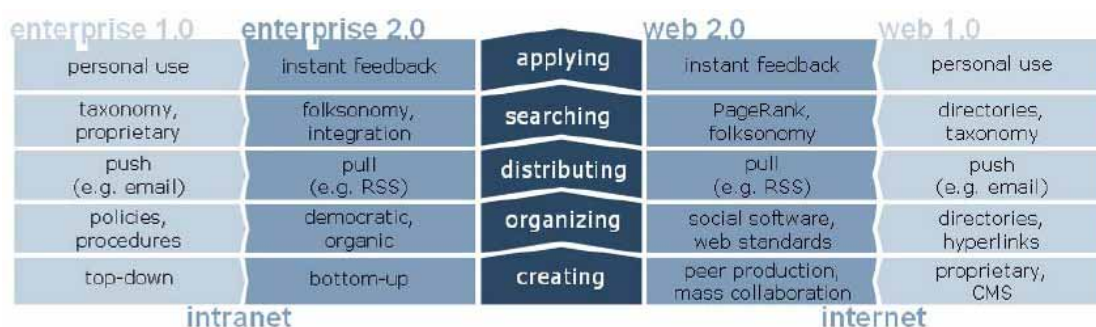


Figure 3.5 - The information lifecycle in the internet and intranet (Hirsch *et al.*, 2009)



Figure 3.5 illustrates the other changes that corporate content is subject to under the enterprise 2.0 intranet. As well as bottom-up content creation, the other major change is in the way content is now distributed. Where previously content was pushed to the employee through e-mail, it is now the case that content can be pulled to the employees PC using tools like RSS feeds, complementing the traditional push model. The content is also now being structured as per the viewpoint of the employees in the form of folksonomies, moving away from the hierarchical model previously employed. This democratic approach to content structure greatly aides the content discovery process for the employee while also aiding the spread of knowledge throughout the organisation. Finally, the application of the content is changing from personal one way publication to a more inclusive collaborative publication model where two-way communication is the norm and instant feedback is received on published content.

This enterprise 2.0 model fits into what Malhotra (2005) terms the evolution of the Real Time Enterprise (RTE), an enterprise based upon getting the right information to the right people at the right time (Gartner Inc 2002). Malhotra (2005) sees a move from the traditional push model of technology roll-out to a pull-model that treats business performance as the prime driver for knowledge utilisation as well as providing a better fit to the RTE model where knowledge is made available as requested as opposed to an overload of knowledge that can exist by having multiple technologies available providing similar functionality. The RTE model focuses on technology as an accelerator within business and removes the focus from specific technologies. Enterprise 2.0 and its supporting Web 2.0 technologies are a stepping stone to the realisation of the Real Time Enterprise where every employee is empowered with the required knowledge.

For enterprise 2.0 to be successful within an organisation, the suite of web 2.0 technologies needs to be aligned with the business processes to provide maximum benefit to the enterprise and the technologies need to be utilised based on what it offers and not on what it promises. For example, technologies such as video conferencing, on-line communities of practice and the new emergence of unified communication tools, provide support for the transfer of tacit knowledge overcoming traditional geographic barriers thus making it possible to share tacit knowledge with people in

other locations throughout the organisation. It is as a result of business benefits such as this that Forrester Research predict the Enterprise 2.0 solutions market will be \$4.6 billion by 2013 (Business Wire 2008).

### ***3.7 Conclusion***

This chapter presented an overview of the key technical and social developments since the inception of the web. It has shown how the web evolved into a collaborative environment overcoming the issues of the early web. Early facade technologies such as DHTML and Flash which helped hide the underlying content representation have now disappeared replaced by sandbox technologies that empower the user to participate on the web.

These sandbox technologies led to a change in the culture of the web which in turn is helping to change the cultures of organisations within the corporate world. It is this culture change which has helped drive knowledge sharing within the enterprise and fuel the growth of Wiki's and related technologies in corporate networks.

## **4 THE CONTEMPORARY WEB – GOOGLE AS THE GATEKEEPER**

### ***4.1 Introduction***

The web evolution culminating in Web 2.0 has led to the exponential growth of data, information and knowledge on the web. However, the foundations on which this occurred have remained largely unchanged. The World Wide Web today is still relatively simple technology containing distributed hypertext accessed via link navigation and search engines (Horrocks and Bechhofer 2007).

One company has developed the technology to put order on the domain of distributed hypertext to enable us to best traverse its content in its current state. This chapter will now examine how Google's dominance has led to the company holding this position on the Web. It will also discuss some of the problems associated with the current linked web architecture and present some shortcomings with the hypertext paradigm. Finally it will consider the future direction of the web search in light of developments and the semantic web and the continued growth of the contemporary web.

### ***4.2 Google's Dominance***

It can be argued that the name Google is synonymous with the modern day Web with the company's search engine and email websites dominating the landscape of the web. Newer applications such as Google Earth and Google Docs are also beginning to gain new market share as the company goes from strength to strength. In its core competency, the search engine, it has been reported that 64% of searches carried out on the web are done on the Google platform with the nearest competitor Yahoo only responsible for 20% (Hitwise 2007). Figure 4.1 supports this claim showing Google with a 70% share on the US Market. One of these tools, the Google search engine, has become predominant, to the extent that “to Google” had become de facto a verb in the English language by mid-2003, despite the objections of the company (Quint, 2002; Duffy, BBC, 2003)

## U.S. Search Share, June 2008

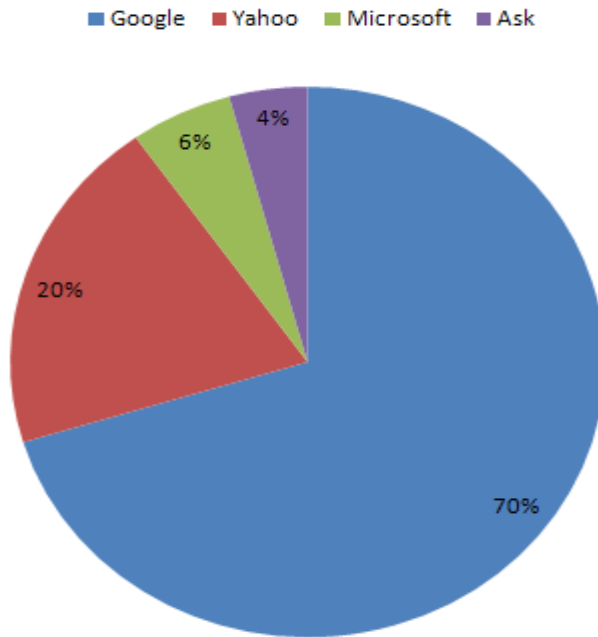


Figure 4.1 Google Search Share - Source HitWise 2007

John Batelle, one of the founders of Wired Magazine is quoted as stating “*The web has an interface and I would argue that it is Google right now*” (Chapman 2007).

Google has grown into this position not by being first on the market but from enhancing the technology used by the first generation search engines such as AltaVista (Broder 2002). Initially, these search engines functioned in one of two ways. Some provided the option of openly searching the Web's content (e.g. WebCrawler and Lycos) while others organized information into Web directories and people could access content by clicking on categorized links (e.g. Yahoo). The former relied on computer programs whereas the latter was manually compiled (Chu and Rosenthal 1996). It can also be noted that most of the search engines began life on the college campus but these ventures soon left the academic settings and became profit-seeking commercial enterprises as illustrated in Table 4.1. The move into the commercial domain can affect the perceived independence of the search software as discussed later within this chapter.

Search Engine	Launch year	Original Affiliation
Lycos	1994	Carnegie Mellon University
WebCrawler	1994	University of Washington
Yahoo!	1994	Stanford University
Altavista	1995	Digital Equipment Corporation
Excite	1995	Excite, Inc.
Infoseek	1995	Private
HotBot	1996	Wired Ventures
Google	1998	Stanford University (Google, Inc. by the time of launch)

Table 4.1 - The launch date of some major search engines and their original institutional affiliations (Bar-Ilan 2007).

Google took this content searching feature a step further by taking this technology and enhancing it with a ranking system on the results where a web page would be ranked based on the number of other pages that linked to it. This system is known widely as PageRank. This algorithm, along with other algorithms that analyse the chain of hypertext links leading to sites, enable Google to offer a superior way to search the Web compared to brute-force methods that other search engines primarily use.

Google provides the following explanation regarding PageRank:

*PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at considerably more than the sheer volume of votes, or links a page receives; for example, it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important." Using these and other factors, Google provides its views on pages' relative importance.*

The basic concept of PageRank is that of introducing the notion of page authority, independent of the page content. This type of authority measure is only possible with from the topological structure of the Web (Pujari 2006).

With the continued growth of the web, it can be argued that search functionality based on the topological view may no longer be sustainable. The research problem evaluated as part of this dissertation examines the Google search engine functionality within a limited domain and presents some findings in this field. The next sections of this chapter present some of the issues associated with the Page Authority approach to search rankings.

### ***4.3 Google's PageRank Limitation***

Issues surrounding accessibility of information on the Web has existed in some form since its formation with commentators describing problems with outdated indexing and poor search coverage (Lawrence and Giles, 2000). An extensive year long study was carried out in 2004 with weekly snapshots of 150 Web sites over the course of the year. The sites were measured on the evolution of content and link structure. It was found that the link structure of the Web was significantly more dynamic than the pages and the content. On completion of the study, it was found that only 24% of the initial links were still available and that on average, 25% of links measured each week were new links, as opposed to the 8% for page creation and 5% content creation. This result indicates point to a need for search engines to update their link-based ranking metrics (such as PageRank) frequently. For example, given the 25% new links every week, a week-old ranking may not reflect the current ranking of the pages very well (Ntoulas, Cho and Olsten 2004).

While this indexing latency presents no obvious issues for established sites that themselves are well linked with other well-known sites, it does present problems for new sites. It could take weeks for the new pages to get into the Google index which will depend on the frequency of Google's Deep Web crawl as new sites will not be indexed via Google's regular surface crawls (Madhavan 2008). It will also take time for these new sites to be linked from the established websites. Researchers have found that because search engines repeatedly return the most popular pages at the top of

search results, popular pages tend to get even more popular. The unpopular pages either get ignored by the user or do not make the result set. This “*rich-get-richer*” phenomenon is particularly problematic for new and high-quality pages because they may never get a chance to get users' attention (Cho, Roy and Adams 2005). Research in this area has produced evidence to suggest that this inequality does in fact exist on the web (Cho and Roy 2004). It should be noted that Google use Hypertext-Matching Analysis in conjunction with the Page Rank Algorithm which analyses the full content of a page and factors in fonts, subdivisions and the precise location of each word (Wiguna, Fernandez-Tebar and Garcia). This analysis method was developed to overcome the issues associated with the previous search approach which analysed a Web pages Meta-tag. It was noted that webmasters were intentionally placing misleading content in a Web page's Meta tags in order to boost the page's search engine ranking (Henzinger, Motwani, and Silverstein 2002). Detailed research on the Hypertext-Matching Analysis process does not exist so it is hard to determine how it negates some of the perceived weaknesses of the Page Rank algorithm.

Although seemingly neutral, search engines and directories systematically exclude certain sites in favour of others either by design or accident (Introna and Nissenbaum 2000). Hargittai (2004) argues that commercial interests lie behind the most popular Web sites, and those to which users turn to find their way to online content. Tools such as Search Engine Optimisations are available for these commercial interests to aid them in achieving higher search rankings.

#### ***4.3.1 Search Engine Optimization***

Google has grown from its original function, as an information retrieval tool, to a position where it currently has a large influence on the web scenery. Webmasters are constantly trying to “please” Google, so that their web pages become top ranked results for appropriate search terms (Bar-Ilan 2007). They optimize their web pages in order to improve the rankings of the pages on Google, a practice known as Search Engine Optimization.

Search Engine Optimization (SEO) is concerned with exploiting Google's ranking algorithm in order to achieve the highest placement on the web search results, which

are also known as natural or organic results because they are supposed to reflect relevancy in searchers' standard. IT firms that specialise in the practice of SEO are typically employed by advertisers on behalf of clients in order to ensure their web presence is fully utilised. A second option, known as paid placement, is also available on Google where a pre-specified region of the result page is reserved for sponsored or paid results which match the keywords entered in the search. Google charge placement fees for this service with the cost tied to the price of the relevant keywords, which is primarily determined by auction and measured by cost per click, and the number of click-throughs the advertisement receives.

The concern with SEO is that search engines that are guided by profit may inadvertently direct people away from the most relevant and best quality sites in favour of those that have paid the highest bids for placement on the results page regardless of their quality and relevance to the search query (Hargittai 2004). This same concern applies for pages that have been optimised for placement by specialist IT companies. This concern is compounded by analysis of large-scale search engine usage data, which suggest that users mainly rely on the first page of results to a search query. A study analyzing almost one billion queries on the AltaVista search engine carried out by Silverstein *et al.*, (1999) showed that in 85 percent of the cases users only viewed the first screen of results. Another study carried out in 2006 by Jupiter Research and marketing firm iProspect (source iProspect 2008) found that people will go through three pages of results before giving up and found that a third of users linked companies on the first page of results with top brands. Perhaps most interestingly of all, 41% of the 2,369 people surveyed admitted to changing search engines or their search term if they did not find what they were searching for on the first page.

What these findings highlight is that web users heavily rely on search sites for presenting them with the desired content. They are likely to try alternate search sites rather than using sophisticated search techniques to fine-tune their queries if they cannot find what the relevant content. This implies that content prominently displayed on Google, Yahoo and other popular search sites has a reasonable chance of being the destination of web users.



#### **4.3.2 The Deep Web**

While search engines results may sometimes be skewed by commercial interest, they can also suffer from their inability to retrieve results from the non-indexable web or Deep Web as it is more popularly known. The Deep Web refers to any pages that remain outside of the search index. These pages are inaccessible either as a result of poor linkage with other websites which leave them on the periphery of web indexes or because the web page content is stored in backend databases which cannot be indexed in the standard format. Mike Bergman, credited with coining the phrase, has said that *"searching on the Web today can be compared to dragging a net across the surface of the ocean; a great deal may be caught in the net, but there is a wealth of information that is deep and therefore missed"* (Bergman, 2000). As mentioned in the introduction of this dissertation, one of the more recent studies estimates the size of the web as of 2005 to be 11.5 billion web pages (Gulli and Signorini, 2005). This estimate is already outdated, and in any case it only measures what Lawrence and Giles (2000) called the *"indexable Web,"* which excludes the Deep Web. The Deep Web itself was estimated to be at least 450 times larger than the indexable web (Bergman, 2000). These figures highlight that a staggering amount of content is in essence lost on the Web and show the problems facing search engines in crawling this content.

Google is tackling this issue on a number of fronts, however, and research carried on by Google in 2008 found a number of potential approaches to retrieving content from the deep web. Cafarella *et al .*, (2008) firstly explored the possibility of using smaller search engines based on specific domains using semantic mappings technology but it was found that the overhead on this approach, both in maintaining the domains and semantic mappings, made it an unrealistic option. The second approach involved a surfacing approach whereby partial indexing was run to collect a subset of data on the most popular search submissions. Full indexing on the submissions carried out off-line enabling Google to run deeper crawls on the web thus presenting better result set for future search submissions. This research shows progress has been made in the area of Deep Web exploration.

Sites that use backend database technology also contribute to the Deep Web, as it is difficult for the current generation of search engines to access and index this content.

Certain institutions, however, have given Google the access they require so that the database content can be indexed. An example of this is Google Scholar, where academic institutions have opened up their repositories of educational content to Google so that this content can be retrieved through one of Google's customised search engines (Jacso 2005). Similar initiatives need to occur in other domains so that search results become a truer reflection of the Web content. As Google Scholar illustrates, it is not a necessity to provide the content free of charge, but simply, to make the web user aware of the content.

As with Search Engine Optimisations, the question exists as to whether search engines such as Google can overcome the problem of the Deep Web on the Webs current architecture. Search Engine Optimisation and the Deep Web were not explored as part of the research problem but they are an issue for knowledge retrieval on the current web and thus form part of the rationale for undertaking the research.

#### ***4.4 Google Challenge***

Web users must be educated about the influence commercial interests could have on search result listing. If users do not possess the knowledge on how content is organised and presented to them online, then they are essentially reliant on what search sites decide to feature prominently and present to them. They must be made aware that the most prominent results are not necessarily the best quality results, merely the most popular. Google argues that the most popular websites should equate to the highest quality websites based on the democratic nature of the web where link structure should server as an indicator of a websites quality or value.

Non-profit organisations such as educational groups and charities lack the resources that are essential to obtaining the web exposure necessary for reaching users and resources which are needed to expose search engine optimisations. As a result of this, diversity of web content could suffer as non-profit organisations and individual websites are in essence hidden from the potential audience. Signs are appearing though that Google does see the weakness with an algorithm based on popularity. New Google features, such as Google Grants, which award free Ad Words advertising to charitable organisations based on the organisations meeting certain requirements,

demonstrate that Google is trying to balance the books (Stabile 2008). Davies *et al.*, (2005) suggest that the future role of search engines is in supporting the information management process as opposed to concentrating on incremental improvements in ranking algorithms. The search software should no longer return document links, but instead possess the intelligence to interpret and analyse the relevant documents that are returned by the search. It seems Google has begun to release the potential of this approach and recently announced support for RDF (source Google 2009), a semantic technology that will be discussed in the following chapter.

#### **4.5 Conclusion**

This chapter discusses the role of Google on the contemporary web. The position it holds on the search market is researched and analysed in terms of the company's history. An in-depth examination of the underlying PageRank algorithm, which determines the search results, is undertaken and the weakness of this approach is discussed with relation to the Deep Web. The commercial interest of Google is also discussed in terms of Search Engine Optimisation and placed advertising, which can compete directly with organic results. Finally the future of search on the web is discussed and changes in Google's strategy are examined.

Question marks were raised here with relation to Google search, but Google is simply exploiting the current structure of the web and is likely to continue doing so until the underlying structure of content changes. This dissertation does not seek to find fault with Google, but simply illustrate some of the failings of search engines that rely on rankings algorithms. The search engine needs to move to an information centric approach, away from the document-centric approach of today. As Davies *et al.*, (2005) states "*Corporate knowledge workers need information defined by its meaning, not by text strings ('bags of words'). They also need information relevant to their interests and to their current context. They need to find not just documents, but sections and information entities within documents*". In order to achieve this, the content of the web has to change through semantic technologies.

## 5 WEB 3.0 – THE INTELLIGENT WEB

### 5.1 Introduction

*“People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty - on Web 2.0 and access to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource.”* (Berners-Lee, 2006).

Web 3.0 is the third stage of the web evolution that is beginning now. It is envisaged as a World Wide Web of meaning and knowledge where both machines and web users have the ability to exploit the full potential of web content. It is not a re-invention of the web but a progression of the web from the social web that emerged during web 2.0 to architectures of learning and knowledge that will make the web more connected, open and intelligent. This chapter will evaluate how this can be achieved through new levels of intelligence to the applications and infrastructure of the web for an improved interface for the web user. Finally this chapter will look at the framework and core technologies of the Semantic Web which also form part of the research experiment.

### 5.2 The Semantic Web

The semantic web or Web 3.0 as it is popularly known came about as a result of researchers and web developers proposing the augmentation of the Web with languages that would describe the content of web pages thus making their meaning clear. Tim Berners-Lee, inventor of the Web, coined the term “*Semantic Web*” to describe this approach.

*“The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”.* (Berners-Lee, Hendler & Lassila, 2001)

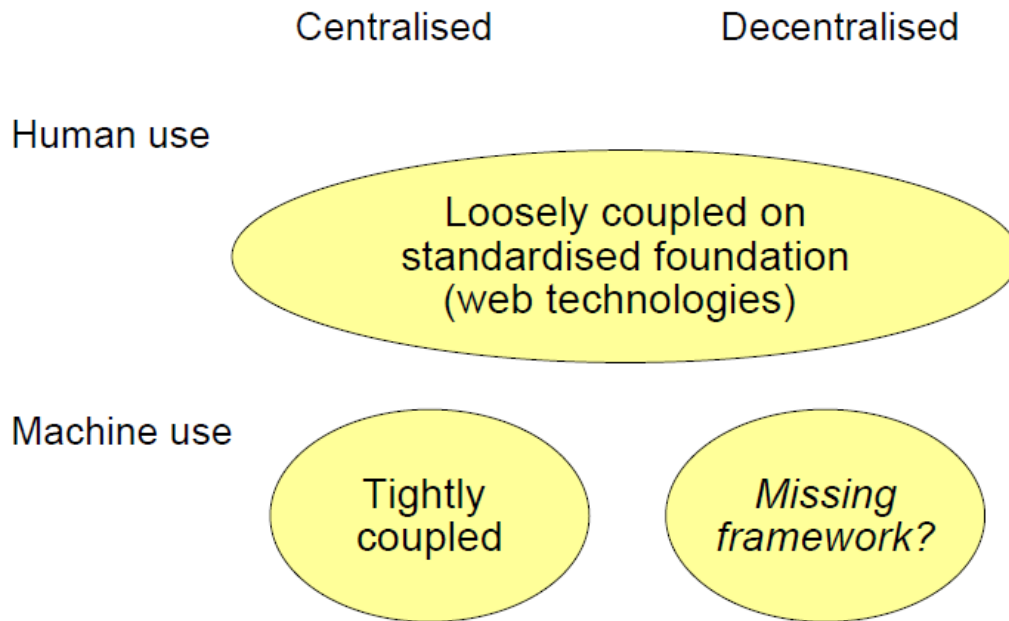
The ultimate goal of the Semantic Web is to transform the existing web into “*. . . a set of connected applications . . . forming a consistent logical web of data . . .*” (Berners-

Lee, Hendler & Lassila, 2001). Put another way, it is the transformation of the Web from a linked document repository into a distributed knowledge base and application platform. This platform will allow for the vast range of content and services to be effectively exploited (Peter, Patel-Schneidel and Horrocks 2006).

W3C provides the following definition of the Semantic Web

*The Semantic Web is a Web of data. There is a lot of data we all use every day, and it is not part of the Web. For example, I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar? Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.* (Source W3C)

When O'Reilly (2005) stated in his Web 2.0 manifesto that “*data is the intel inside*”, he recognised the untapped potential of the web but the web of participation that he championed did not present any concrete solutions to unearthing the data at a fundamental level. Web 2.0 enabled web users to share and create data more easily through tools like Folksonomies and Blogs but the underlying data remained unreadable to the computer outside of the context of presentation. A recent talk by Tim Berners-Lee at TED (TED March 2009) highlighted this gap when he called for the freeing up the raw data held within government and organisation repositories in order for it to be utilised on the World Wide Web for full exploitation.



**Figure 5.1 The Semantic Web (Olsson 2007)**

Figure 5.1 shows this distinction between the human use and the machine use in terms of the web. The web's current content is designed for human consumption, and web tasks typically require web users to combine data from different sites on the web. For example, hotel and travel information may come from different sites. This does not provide any issues for web users as they can combine these sources of information easily even if different terminology is used across the array of web sites.

The automated use of web content by computers to carry out similar tasks to a web user is difficult due to the construction of the current web content. Content is difficult to combine as no distinct relationship between the content is defined i.e. no linkage exists between content that may be related. Coupled with this, content such as images and videos are complicated to represent in machine readable format. This is because web content is primarily intended for presentation to and consumption by human users: HTML mark-up is primarily concerned with style, layout, size, colour and other presentational issues (Peter, Patel-Schneidel and Horrocks 2006).

The Semantic Web aims to overcome the issues the current web presents for machine agents by making web content more accessible to automated processes and by

illuminating the hidden meaning of web content through semantic annotations. This new programmable data web will be based on an interlinked network of data where programs will know how to relate to content and make use of this content. These new complex applications will provide a new class of solutions over the web that will enable the web user to utilise the web to its full capabilities.

For this to become a reality however; a common framework is required that allows data to be shared and reused across application, enterprise, and community boundaries. This paper will now discuss this framework and the underlying architecture (the semantic stack) of the Semantic Web as outlined by the World Wide Web Consortium.

### ***5.3 Semantic Web Technologies***

Semantic Web technologies allow us to build applications and solutions that were previously impossible and unfeasible. The combination of semantic concepts with new technologies makes it possible to model data and capture the relationships between the data for machine learning. Semantic technologies tap new value by modelling knowledge, adding intelligence and enabling knowledge (Mills Davis 2009).

Semantic Web technologies can be utilised across a variety of application areas. Data Integration involving the merging of data held in different formats across various repositories can provide better, domain specific search engine capabilities. The combination of these new semantic technologies, it is hoped will lead to the creation of intelligent software agents that will help facilitate knowledge sharing and exchange across the Web.

The Semantic technologies are based on the semantic stack which illustrates the architecture of the Semantic Web as demonstrated in Figure 5.2. It describes how the technologies that are standardised for Semantic Web are organized to make the Semantic Web possible. The stack is built on the underlying technologies of the Web specifically the Universal Resource Identifier and the Unicode Character Set. XML, an improvement on the original HTML syntax of the web is the mark-up language of the Semantic Web providing the desired structure for the content. These core technologies

reiterate the fact the Semantic Web is an extension and not a replacement of the classic hypertext web.

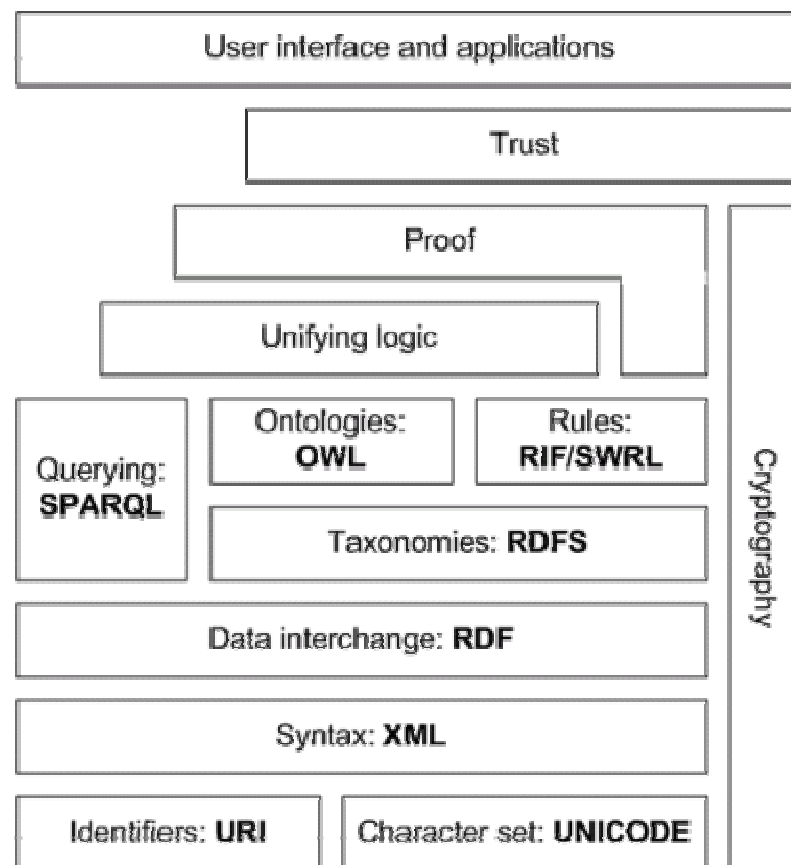


Figure 5.2 - Semantic Stack – World Wide Web Consortium

The middle layers of the Semantic Stack are the technologies that the World Wide Web Consortium is seeking to standardise to enable building semantic web applications. Technologies within these layers consist of RDF, RDFS, OWL and RIF and these will now be discussed in greater detail as together they form the core technologies of the Semantic Web

### ***5.3.1 The Data Model (RDF)***

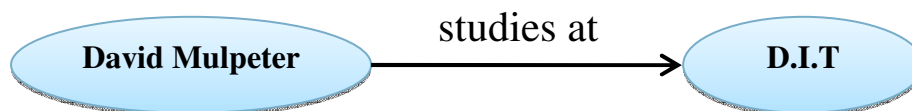
RDF stands for Resource Description Framework, the data model of the Semantic Web. The Worldwide Web Consortium (W3C) created the framework as a recommendation for the formulation of metadata for World Wide Web resources and was designed to provide interoperability between applications that exchange metadata.



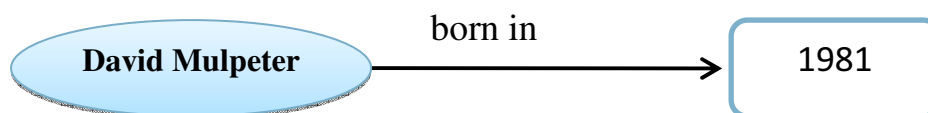
It provides an authoritative, triple-based representation language for Universal Resource Identifiers, the language used to create statements each with a subject, verb and object (Bojars *et al.*, 2007).

The following diagrams illustrate the representations with subject, verb and object

- “David Mulpeter studies at DIT”



- “David Mulpeter was born in 1981”



- “D.I.T is headquartered in Dublin”

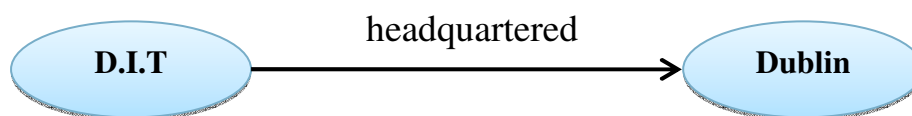


Figure 5.3 RDF Representations (Author)

These representations can then be joined together to form labelled directed graphs as illustrated in the below diagram

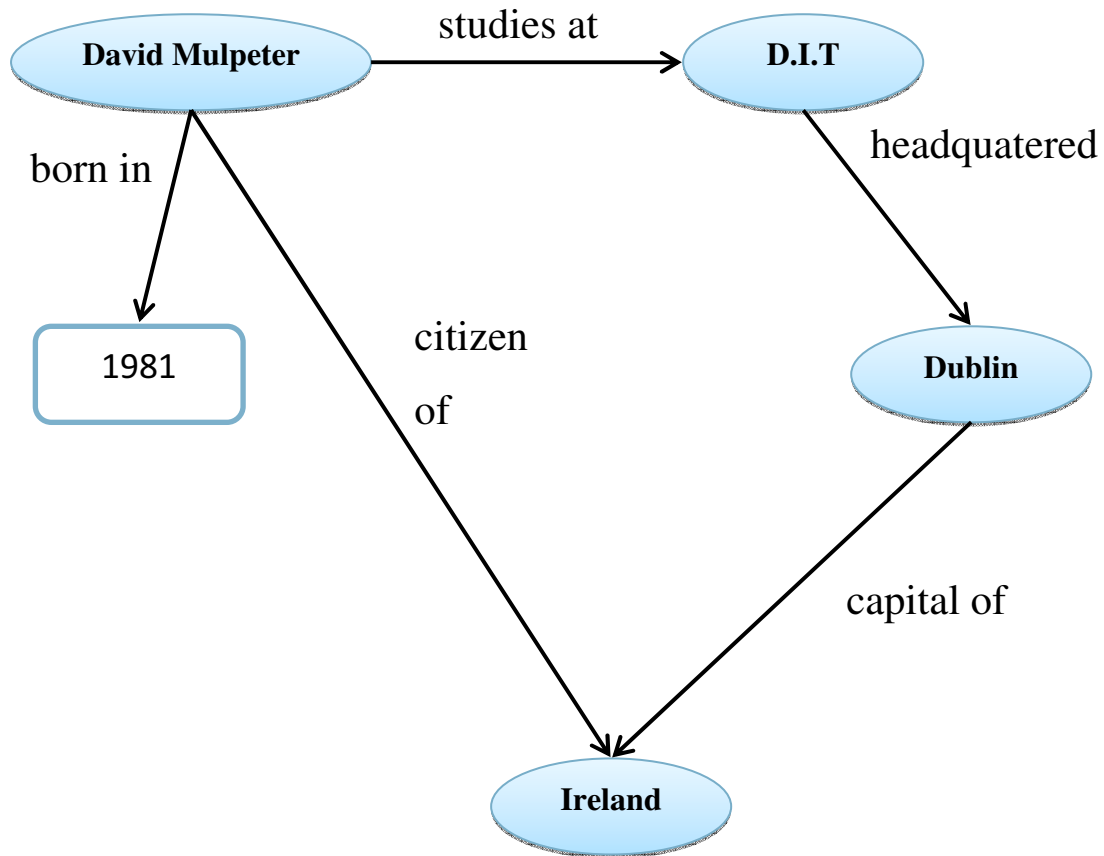


Figure 5.4 RDF Graph (Author)

Where the original Web took hypertext and made it work on a global scale; the vision for RDF was to provide a minimalist knowledge representation for the Web. The framework itself has existed since 1999 but was revisited and revised in 2004. The original specification was extremely formal and was conceived as a solution to a problem people didn't think existed back when it was first created. It was only later when the Semantic Web movement gained prominence that it came back into focus.

The basic model of the RDF allows us to do a lot on the blackboard, but does not give us many tools. It gives us a model of assertions and quotations on which we can map the data in any new format. We next need a schema layer to declare the existence of new property (Tim Berners-Lee 1998). The RDF Schema was created for this purpose and it provides the basic vocabulary for RDF.

### 5.3.2 Structure & Semantics (RDFS, OWL, RIF)

Three of the components of the Semantic Stack are about describing a domain well enough to capture the meaning of resources and relationships in the domain.

- RDFS
- OWL
- RIF

#### **RDFS**

The Resource Description Framework Schema extends the RDF standard with added functionality to specify domain vocabulary and object structures. It provides mechanisms for describing clusters of related resources and the relationships between these resources (Brickle and Guha 2000). It is an extensible knowledge representation language which provides the basic elements for the description of ontologies, otherwise called Resource Description Framework (RDF) vocabularies. RDF Schema provides the framework to describe application-specific classes and properties. This allows resources to be defined as instances of classes, and subclasses of classes as the below example illustrates with the resource class horse a subclass of the resource animal (W3C)

```
<rdfs:Class rdfs:ID="horse">  
  <rdfs:subClassOf rdfs:resource="#animal"/>  
</rdfs:Class>
```

RDF, in combination with RDF Schema, offers modelling primitives that can be extended according to the needs at hand. Basic class hierarchies and relations between classes and objects are expressible in RDF Schema (Stojanovic, Stabb and Studer 2001).

## OWL

In order to integrate content from different web sources, a shared understanding of the web domain needs to exist. Knowledge representation formalisms provide structures for organizing this knowledge (RDFS), but provide no mechanisms for sharing it. Heflin *et al.*, (2001) sees Ontologies as providing this through common vocabulary that supports the sharing and reuse of knowledge. Hepp, Bachlechner and Siorpaes (2006) describe Ontologies as unambiguous representations of concepts, and the relationships between these concepts. In order to represent these Ontologies formally, a language is required and OWL is the language used for authoring these Ontologies on the Semantic Web. It is a language that is used to describe the concepts and the relations that exist between them that are inbuilt in the content of the Web.

McGuinness *et al.*, (2002) in their work for the W3C Web Ontology Working Group, describe the OWL language as a revision of the DAML+OIL web ontology language incorporating learning's from the design and application use of DAML+OIL. Horrocks and Bechhofer (2007) in their research on Ontologies, have seen OWL become a *de facto* standard for ontology development across a number of different fields such as geography, geology, astronomy, agriculture, and life sciences. Sure and Iosif (2002) feel this growth and popularity is largely due to what Ontologies promise: "*a shared and common understanding of a domain that can be communicated between people and application systems*". This continued growth in the field of Ontologies since the creation of the OWL language in 2001 lead to the creation of OWL 2. OWL 2 is an update to OWL with new features such as extended support for databases and extended annotation capabilities. These extra features are valid in both the field of knowledge management and the semantic web. Ding and Fensel (2001) feel that Ontologies provide the opportunity to improve the knowledge capabilities within large organisations and the semantic web is a tool that be used to exploit this capability.

The use of Ontologies on the semantic web differs from their conventional use of where they are used in specialised domains such as specific fields of science. Historically, Ontologies were created and maintained by a group of domain experts in the relevant field and this group defined strict guidelines upon the ontology in order to

maintain quality. The challenge of creating and maintaining Ontologies on the semantic web is sizable given the structure of the web. Auer *et al.*, (2007) recognise this but feel the incremental, collaborative approach of Web 2.0 can help in meeting this challenge. This iterative collaborative approach is needed if the Ontologies are to evolve with the web and is discussed in greater detail in Chapter 6, which explores the meta-knowledge framework.

## RIF

RIF stands for Rules Interchange Format, a proposed component of the semantic web to be used in conjunction with OWL. It is a developing standard for exchanging rules among different systems, especially on the Semantic Web and is represented in the XML language.

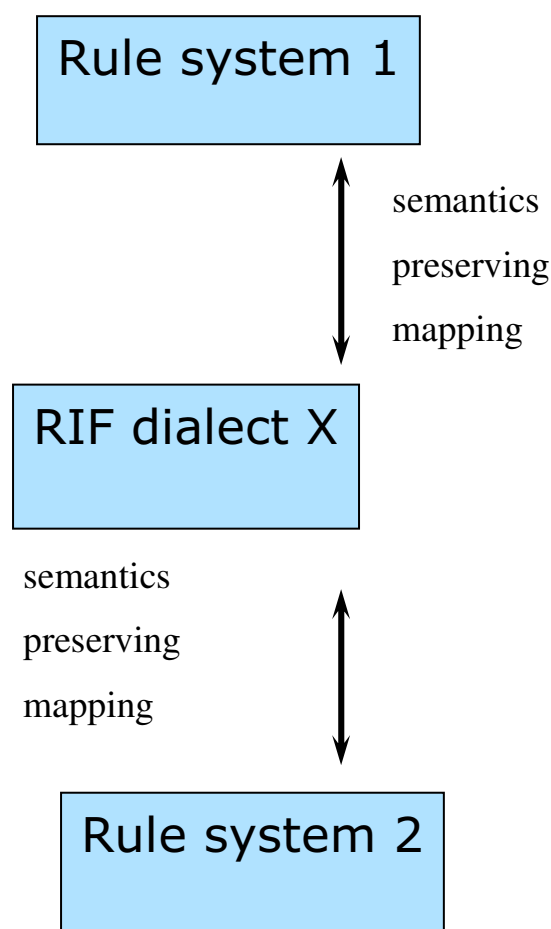


Figure 5.5 RIF Interaction (Kiefer 2007)

Figure 5.5 demonstrates how this rule sharing and exchange is facilitated through RIF dialects. Dialects are needed due to the different semantics and rule languages that exist but a carefully chosen set of interrelated dialects can serve the purpose of sharing and exchanging rules over the Web. It is hoped that the standard will allow systems to inter operate independent of the different rule based languages and technologies they use.

### ***5.3.3 The Query Language (SPARQL)***

*"Trying to use the Semantic Web without SPARQL is like trying to use a relational database without SQL"* – (Tim Berners-Lee 2008 SPARQL Protocol)

In 2004 the RDF Data Access Working Group (part of the Semantic Web Activity) released a first public working draft of a query language for RDF, called SPARQL (Perez, Arenas, and Gutierrez 2006). SPARQL is a SQL-like language for querying sets of RDF graphs natively and the queries themselves look and act like RDF. It can be used to query and search disparate data sources containing both structured and semi-structured data and to explore data by discovering unknown relationships (Prud and Seabourne 2005).

Due to the relative youth of the technology, however, some drawbacks do exist. A large number of data stores exist that cannot be directly queried with SPARQL and the language lack the optimisations that mature technologies such as SQL and XQuery possesses. This is likely to improve as study and implementations of the technology contribute to a body of research surrounding SPARQL (Perez, Arenas, and Gutierrez 2006). Section 5.4 will evaluate the Semantic search in more detail as it is a key factor in the research project.

#### 5.3.4 *Linked Data*

*“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data”* (Berners-Lee 2006).

Linked Data is the latest development on the semantic web and put simply it is the organisation of the resources on the Semantic Web. Linked Data is the method of connecting the resources of the Semantic Web such as RDFs or Ontologies enabling the web user or web agent to navigate between these resources. It accomplished this by following the principles that Tim Berners-Lee outlined at the outset of the Linked Data Project.

These principles are as follows

- *Use URIs as named for things*
- *Use HTTP URIs so that people can look up those names*
- *When someone looks up a URI, provide useful information*
- *Include links to other URIs, so that they can discover more things*

(Tim Berners Lee W3C 2007)

The key to these principles is the utilisation of URIs as resource identifiers which allows the HTTP protocol to retrieve resource descriptions and thus aid in forming the links between the RDF's.

The true power of the linked data becomes apparent through the use of web agents. Auer *et al.*, (2007) describes various, new semantic web agents such as Semantic Web Browsers (Disco) and Semantic Web crawlers (SWSE and Swoogle) which can

interpret the linked data and make semantic connections between the resources to infer potential semantic content.

The Linking Open Data Project, which began as a community project, started within the W3C Semantic Web Education & Outreach group in 2007. Figure 5.6 and 5.7 illustrate the rapid growth in linked data since the project inception. Bizer *et al.*, (2009) argue that the linked data methodology is the foundation for meeting the original goal the Semantic Web; a global Web of machine-readable data and is the stepping stone for the creation of intelligent semantic agents.

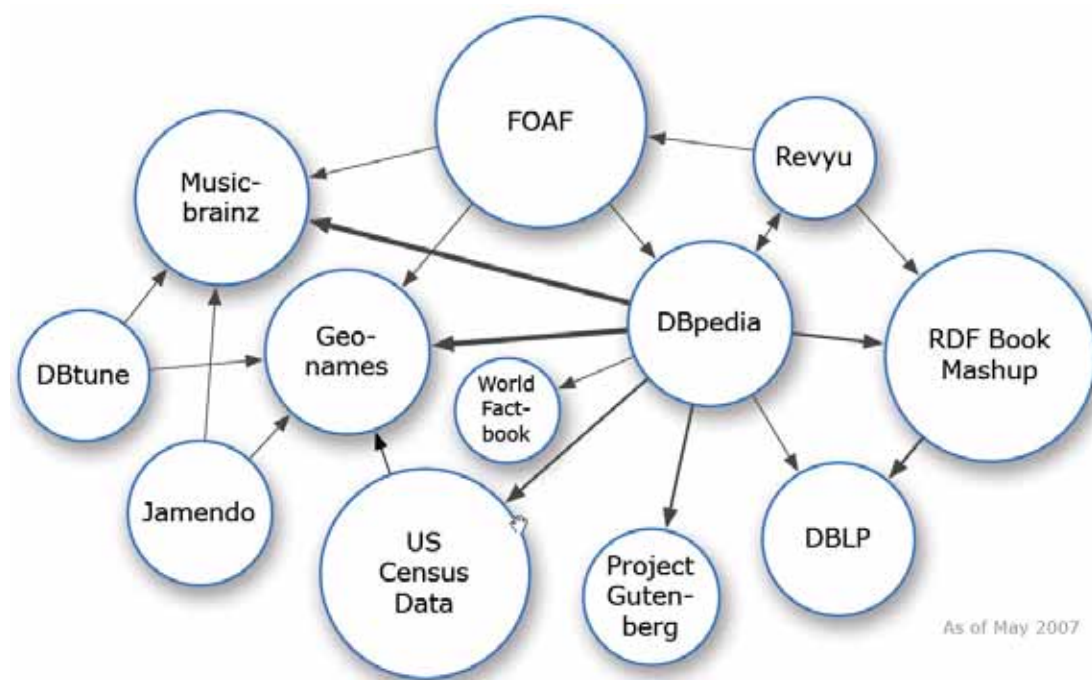


Figure 5.6 – Linked Open Data Cloud Cambridge Semantics 2007



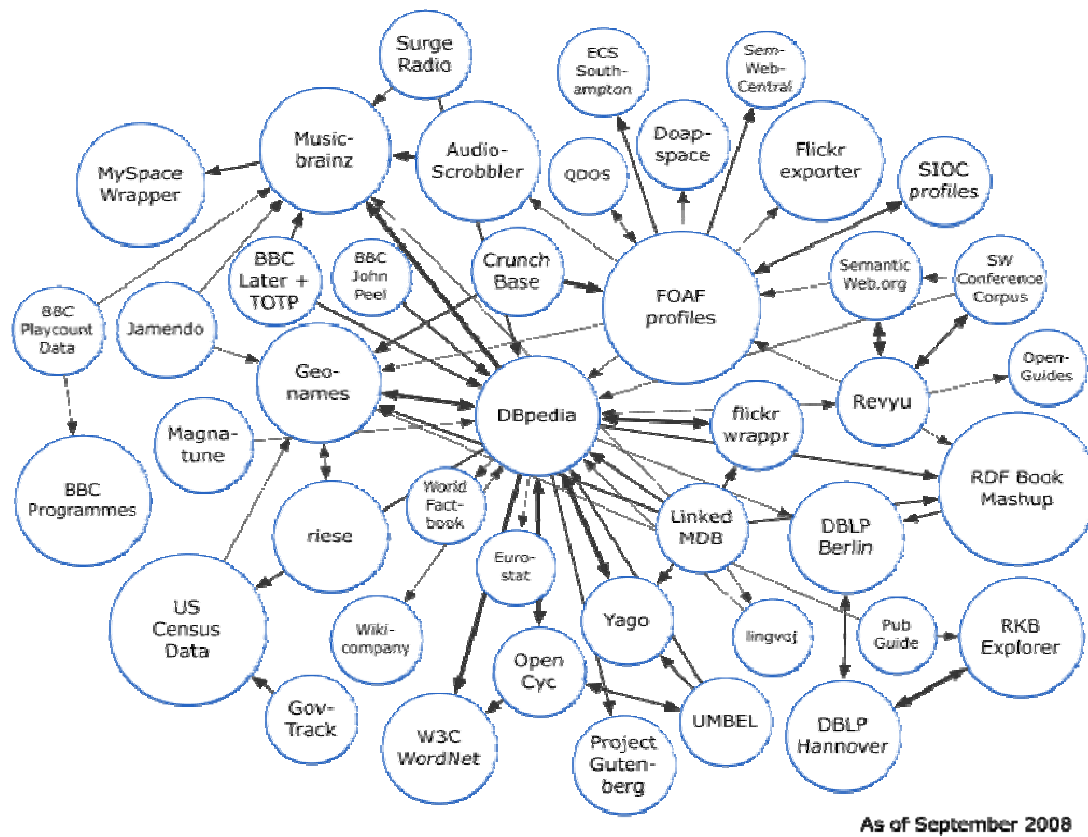


Figure 5.7 – Linked Open Data Cloud Cambridge Semantics 2009

Many within the Linked Data Community and the larger Semantic community believe that Linked Data lies at the heart of bringing the different semantic technologies together and is a crucial breakthrough in the Semantic Jigsaw (Hausenblas 2009), (Bizer *et al* 2009).

#### 5.4 Semantic Search

Semantic search engines operate differently to the ranking based search engines such as Google as described in Section 4.3. Whereas traditional search engines use algorithms such as Google's PageRank for predicting relevancy, semantic search examine the RDF representation of the web content to provide meaningful answers to the search query as opposed to a result set containing the relevant web pages. Guha, McCool and Miller (2003) clarify this difference from the user perspective by categorising the two different types of search. Navigational search is used to describe the common search that users carry out where they provide the search engine with a

combination of words in order to retrieve related web content. The combination of words is read as a search string with no interpretation on the actual words. Research searches, on the other hand, follows the semantic model where the web user provides the search engine with a phrase which represents an object. The users aim here is to retrieve web content directly related to this object i.e. web content that contains attribute information on the object.

Semantic search engines achieve the objective of the research search as defined by Guha, McCool and Miller due to the tightly coupled relationship between RDF and SPARQL. While these illustrated that both technologies are still at a relatively immature stage, it has been demonstrated in limited domains containing structured content, that the semantic search retrieves high quality accurate results. Davies and Weeks (2004) illustrated the possibility of a semantic search model with RDF alone using the QuizRDF application, while the more recent DBPedia application has demonstrated the envisaged interoperability between RDF and SPARQL (Auer *et al.*, 2007). Google has begun to recognise this and recently announced support for the resource description framework. The question as to whether Google can search a database in the same way as SPARQL remains open, but Sheth *et al.*, (2007) has argued the demands that the Semantic Web will place on search engine technology will require Google and its competitors to evolve if they are to handle the demanding information requests of this iteration of the Web.

### ***5.5 Semantic Web Outstanding Issues***

The web technologies developed over the last number of years to meet the needs of the Semantic Web are still in a premature phase. This, in turn, has delayed the implementation of the technologies and languages represented in the upper layers of the semantic stack. Standard solutions remain outstanding for these elements which include privacy, trust and proof.

Of the semantic technologies developed to date, it is recognized that the various aspects of Ontologies require further attention from the research community. Ontology management concerned with the storing, aligning and maintenance of ontologies is an area very important in the broader context of the semantic web (Ding and Fensel

2001). With the growth of Ontologies, the methodologies and measurements required to evaluate them is a discipline that has to be managed carefully to ensure standards are maintained across all ontologies (Bozsak *et al.*, 2002). Sure *et al.*, (2002) recognised this discipline as it applied to the field of knowledge management and developed a methodology for maintaining ontology based KM systems but work is still needed for the new ontologies emerging. Chapter 6 will explore the area of ontology management in greater detail in terms of a meta-knowledge framework.

As the semantic web evolves and semantic web agents advance the elements of the upper layer of the semantic stack will gain increased importance. With these agents integrating information from multiple different sources, the area of trust and proof with relation to content sources will be very important (Heflin 2001). The machine agents that traverse the semantic web will rely on the integrity of the web content, as they lack the intelligence to determine the contents validity. Goldbeck and Hendler (2004) advocate using ontologies as a method for describing the entities and importantly the trust relationship between them, again highlighting how ontologies are central to the semantic stack.

The efforts to address the issues discussed, however, could be futile if the fundamental issue on the semantic web is not tackled; the lack of semantic content on the web. Research by Ding and Finin (2006) estimated the number of semantic documents on the web to be 1.7 million comprising of 300 million RDF's. While these figures are large, they are dwarfed in comparison with the estimations of the overall size of the web as described in section 4.3.2. Many potential factors exist for the lack of semantic content, primarily the complex nature of the semantic web. Outside of the research community, it would be hard to find people with appropriate skill sets for designing and building Semantic Web solutions. Research has also shown that not a large variety of semantic authoring tools and browsers exist which compounds the challenge of creating semantic content (Heflin and Hendler 2001).

The recent emergence of semantic technologies within Wiki's is encouraging in terms of annotation and authoring but work is still needed in create new interfaces for the web user and making existing interfaces more user friendly. Issues with authoring tools and browsers will be discussed in chapter 7 in relation to the research problem.

Benjamins *et al.*, (2002) feel that the migration of HTML to XML will help bridge the gap from the current web to the semantic web. They do raise the risk of content loss on the semantic web should any existing HTML content remain in its current format.

## **5.6 Conclusion**

This chapter discusses the new web for machines and software agents that will plug in to the future World Wide Database. This Semantic Web is seen as the next iteration of the Web and is commonly referred to as the intelligent web. The key technologies of this new intelligent web are discussed in this chapter. These technologies unlike the organic technologies of Web 2.0 are defined by standardisation bodies, such as the W3C and OASIS. The chapter describes these technologies in detail with emphasis on the search technologies, as they relate to the research project. Finally the chapter outlines some of the outstanding issues with the semantic web, in particular, the field of ontologies. Some of these issues are tackled in the following chapter on the Meta-Knowledge Framework.

.

## 6 META KNOWLEDGE FRAMEWORK

### 6.1 Introduction

As discussed in Chapter 5, the Semantic Web is based on accessing and using RDF content with Semantic Query languages such as SPARQL. Both these technologies necessitate ontologies which support the sharable and reusable representations of knowledge. This ontology framework provides the backbone to the Semantic Web but issues with its implementation on the World Wide Web have caused concern within the Semantic Community. This chapter discusses these concerns and takes a detailed look at the different taxonomy representations of web content through the different iterations of the Web. It examines the strengths and weaknesses of the different methods employed to represent metadata. Based on this analysis, a taxonomy compromise is proposed for the Web.

### 6.2 The Knowledge Management Viewpoint

Despite there being an emerging consensus that the use of technology can be used to facilitate many aspects of knowledge management such as the knowledge extraction, codification, representation and transfer of knowledge, there is still debate about whether in fact the use computing technology is central to the success of knowledge management initiatives (Walsham 2001). This has given rise to two distinct views of knowledge management: the "*cognitive*" view and the "*community*" view. The community view emphasises knowledge as socially constructed and is managed primarily by encouraging groups and individuals to communicate and share experiences and ideas at a tacit level. The cognitive view considers knowledge in objective terms which can be expressed and codified, and is often expressed by the capture and codification of knowledge in computer systems.

Issues exist with relation to expressing tacit knowledge within computer systems and this presents obstacles for the representation of tacit content in a computer understandable format. Codifying knowledge with the power of the existing IT and without the support from sociocultural inputs, will result in de-contextualization, i.e.

“*knowledge dilution*”. Mohamed, Stankosky and Murray (2006) argue that this knowledge dilution can be avoided through merging IT with social networks such as communities of practice to create a synergy effect.

This debate is also applicable in the web community where argument exists on the optimal means by which to organize the web content. Halpin, Robu and Sherperd (2007) notes that the argument often pits formalized classifications (cognitive view) against distributed collaborative tagging systems (community view).

### **6.3 Community Metadata**

The process of distributed collaborative tagging or social tagging is commonly referred to as folksonomy. Grouping of related tags are determined programmatically by the system and the cluster of these tags together form the folksonomy (Mathes 2004). Merholz (2004) termed it as “*metadata for the masses*”. Tagging has emerged as a popular method for categorising web content during the Web 2.0 era due of its ease of use. Initially, tagging was used as a tool by the individual web user but the benefits of sharing the categorisations with the larger web community became apparent, and the field of folksonomy emerged (Peters and Stock 2007).

While tagging has proven to be a straightforward and intuitive way to organise web content, it does not describe the full semantics of the data in terms of the relationship of the data to other data. W3C argue that a folksonomy tag is essentially two thirds of the RDF triple representation as outlined in section 5.3.1. They contain the subject and the object, but the verb to link both is absent. Peters and Stock (2007) also discuss the lack of authority on folksonomies which controls the terminology the tag creators use. Without this control, inconsistencies emerge among the tagging with each tag is arbitrarily made by different users and this causes taxonomy inconsistency, such as “*weblog*”, “*blog*”, and “*blogging*” which Mathes (2004) feels is a crucial weakness of the folksonomy classification approach. Certain controls have being put in place in an effort to eliminate the inconsistencies in tagging such as tag prompting but inconsistencies still occur due to human error. Ohmukai, Hamasaki and Takeda (2005) discuss the expanding scope of folksonomies into content areas such as photos, movies

and music but agree that the quality of these new tags remain suspect as a result of the bottom-up approach used to generate them. This they feel, in turn, severely affects the quality of a tag-based search.

Despite these weaknesses, important data are tagged through folksonomies and a viable solution must be found that meets the requirements of both the web user and the web agent in terms of metadata representation.

#### **6.4 Cognitive Metadata**

The formalised classification falls into the cognitive field of knowledge management. The availability of standardised metadata is a requirement for the machine based agents of the semantic web, and many argue the only way this can be achieved is through classification in strict cognitive terms. This classification is achieved on the semantic web through ontologies.

Shirky (2005) argues, however, that this formal classification in the structure of ontologies only works in small or limited domains managed by expert users. It has also been argued that ontologies are very complicated to create and use due to the strict formalisation present within them (Echarte *et al.*, 2004). Mathes (2004) re-enforces this argument stating that it would be impossible to get web users to use complex, hierarchical, controlled vocabulary such as ontologies to tag content. This complication was recognised at the Ontology Summit in 2007, and it was suggested that ontological engineering should be thought of as a discipline complementary to software engineering and to virtually any discipline dealing with data and information exchange (Gruninger *et al.*, 2008).

These difficulties have stunted the growth of ontologies outside of the traditional domains. Hepp, Bachlechner and Siorpaes (2006) argue that many of the existing ontologies have a poor community grounding as they were not designed collaboratively with potential users; instead they were designed by single individuals or small groups of individuals. This silo development of ontologies is in contrast to the evolution of natural language vocabulary which develops within communities over time. Hepp *et al.*, (2007) feels that this insufficient involvement of users in the

construction of ontologies is a primary reason for the shortage of ontologies across domains.

### ***6.5 The Taxonomy Compromise***

In this chapter the two distinct views; community and cognitive, have been described as direct opposites. However, as Gruber and Thomas (2007) state, this is a “*false dichotomy*” and the two can co-exist in what they term “*an ontology of folksonomy*” where value can be gained by applying semantic web technologies to the content of Web 2.0. Zhdanova *et al.*, (2007) argues that this can also work in the reverse manner, by adding community-support to semantic applications that will facilitate the creation of large amounts of semantic data and metadata. This would inherently extend the Semantic Web. Involving the web community in the Semantic Web will aid in its growth, in much the same way it did for Web 2.0. Van Damme, Hepp and Siorpaes (2007) take this a step further and advocate putting the community at the centre of the ontology engineering process to drive the ontology maturing process.

Theoretical research promoting this cross-pollination of semantic technology with social web behaviour, however, needs to demonstrate that it is possible. It is clear to see how weaknesses of folkonomies could be overcome by the use of ontologies and how the social structures that help create folksonomies could aid in the creation of ontologies, but new tools and techniques will be needed to achieve this. Figure 6.1 illustrates at a high level how this could work.



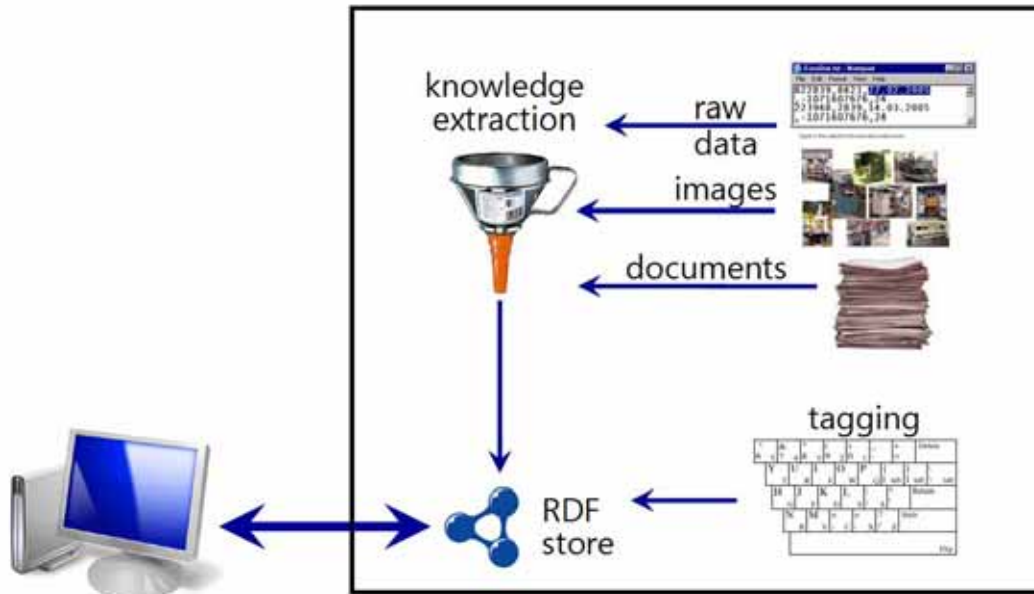


Figure 6.1 – General Application Scenario (Schueler et al., 2008)

Semantic web content consisting of images, documents and data would be represented in ontologies while Non Semantic Content or Traditional Web Content would be tagged and mapped for RDF representation. Knowledge Extraction can then occur for all web content irrespective of its original representation through web agents utilising semantic technologies. Collaborative algorithms for suggesting tags such as the algorithm proposed by Xu *et al.*, (2006) are examples of research concerned with bridging the gap between the social web (Web 2.0) and the semantic web (Web 3.0). Other examples such as the SHOE (Simple HTML Ontology Extensions) and FolksOntology tools demonstrate how the gap is being tackled from a number of different technical approaches (Van Damme, Hepp and Siorpaes 2007, Heflin 2001).

Section 7.4.2 will describe the meta-knowledge framework implemented as part of the research problem of this dissertation. The research problem is based on web 2.0 tools implementing web 3.0 technologies without the use of a formal ontology. The framework implemented was a compromise on taxonomy representation.

## **6.6 Conclusion**

This chapter presented an overview of the two distinct types of knowledge representation – cognitive and community and applied them to the field of taxonomies. Taxonomy representations including folksonomies and ontologies were evaluated and issues relating to both were discussed. Finally, a proposal for a synergy of the two representations was put forward citing recent research in this field. It was argued that a cross-pollination of semantic technology with social web behaviour could alleviate some of the issues present in the respective taxonomy representations.

## **7 SEMANTIC WIKI PILOT**

### ***7.1 Introduction***

As part of the research for this dissertation, a Semantic Wiki called “Compass” was piloted with the IT department of a financial services organisation. The purpose of this pilot was to investigate the possibility of generating relevant knowledge content for the end user through the new Semantic Wiki technology.

### ***7.2 The Organisation***

The research project was undertaken with the Information Technology department of an independent global asset management organisation. The IT department follows the traditional demarcation of IT Infrastructure, IT Development, IT Governance, Project Management Office and IT Application Support. The research project concentrated on the trading system content held within the organisation and the relevant IT teams who manage and maintain this content were invited to partake in the research project.

### ***7.3 Existing Technology***

A following is a summary of the different technologies within the organisation that are used for the purpose of sharing knowledge.

#### ***7.3.1 ITECH (SharePoint)***

The organisation utilises the Microsoft SharePoint Application for all project documentation. Non-technical users are familiar and experienced with MS office applications and SharePoint is an extension of the Microsoft Office tool set in this regard. It provides a rich set of features for work flows and documentation revision control and is the company standard for all business application documentation due to its suitability for the storage and maintenance of controlled documentation.

However, SharePoint does not provide all the functionality required to document a dynamic Production IT environment due to the following reasons.

- Dynamic content creation an issue due to the structure of the portal, which was project centric.
  - Duplication of information within documents.
  - Business/IT knowledge separated.
  - Information becomes stale very quickly.
  - Difficult for the end-user to extend functionality within the existing framework
- Coupled with the above, the SharePoint portal was more business orientated, subject to regulation and audit requirements, and typically had a defined period of relevance i.e. the duration of the project. It was for these reasons that the Production Systems Team proposed a Wiki solution, a solution that is steadily growing out across the entire IT department.

### ***7.3.2 K-Portal (TWiki)***

The Wiki selected for the deployment of an enterprise level structured Wiki was TWiki. There were a number of reasons for this, but the primary reason was due to the quality and quantity of applications offered within Twiki. As Twiki grew out of the open source community, it benefits from a large number of developers who are constantly creating plug-ins and extensions for the platform with more than 400 extensions now available. These plug-ins and extensions make the platform highly customizable for the needs of the Production Systems team. The product itself is already in heavy use by companies such as British Telecom, SAP and Motorola and the fact that it was Perl based meant it merged well with the skill set of the Production Systems team. These features coupled with the standard tools provided such as RSS feeds tipped the balance in favour of Twiki as the home for the new Production Systems Knowledge Portal (K-Portal).

The team structured the layout of the TWiki installation to provide a full Content Management Database (CMDB) which has the following high level design.

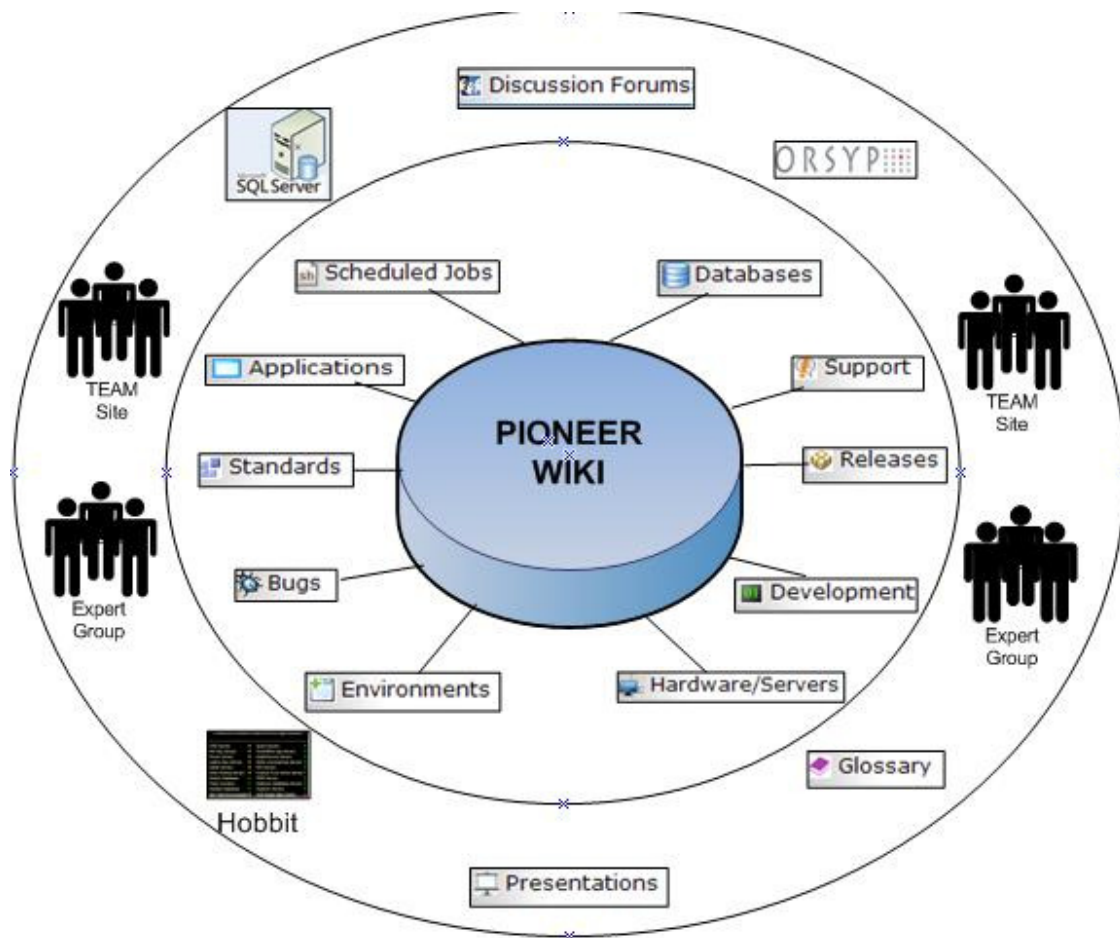


Figure 7.1 K-Portal Wiki Design

There are a number of workspaces (called webs in K-Portal), each providing a function in the CMDB. Each web is self contained, but the ability exists to include information from other webs dynamically within the same level by utilising macro extensions that embed sections of a topic (K-Portal web page) based on the results of regular expressions (or query) searches. This provides a powerful combination of features which provide similar functionality to that of a back end database, while providing all features of a standard Wiki. The inner circles of webs, referred to as “Core Webs” are editable by all users of the Wiki. Editing is locked down to team members on the Team Webs but remain viewable by all registered users. This solution provides a high level of flexibility and ensures that redundant documentation is greatly reduced.

With the structure clearly defined, work is now concentrated on embedding K-Portal into the IT processes, so knowledge content gets created within the day-to-day processes of the department and thus remains constantly updated. Technologies previously discussed such as Tagging, Blogs and RSS feeds are embedded in K-Portal to aid functionality and collaboration on the Portal.

## Welcome to K-Portal - The Knowledge Portal



Welcome to the Knowledge Portal. The aim of this site is to centralize documentation used within the different Pioneer environments and make it available to everyone. The site itself is a Wiki which encourages collaboration between users.

You can browse K-Portal as a guest but if you want to add/modify content please [register](#) if you haven't already done so. This will help identify your contributions which are very welcome.

[aitload](#) [bloomberg](#) [cfa](#) [charlesRiver](#) [Collaboration](#) [DataStage](#) [Derivatives](#) [dollaruniverse](#) [dtsjobs](#) [eipu](#) [ETL](#)  
[EzeCastle](#) [faqs](#) [FIX](#) [FrontOfficeApps](#) [goldensource](#) [hobbitmon](#) [Induction](#) [investment-apps](#) [latentzero](#) [linux](#)  
[madg](#) [Minerva](#) [monitoring](#) [Murex](#) [murex-mo-extracts](#) [Murex\\_BackdatedTrades](#) [Murex\\_PL\\_Filters](#) [MurexCalendar](#) [MurexCFDResets](#)  
[MurexFQHF](#) [MurexOptions](#) [MurexTradeEntry](#) [MurexViewer](#) [nightly\\_batches](#) [oracle](#) [per](#) [PESB](#) [plugins](#) [QAENV.REFRESH](#)  
[replication](#) [request.server](#) [securitymaster](#) [sentinel](#) [smartsockets](#) [solaris](#) [sybase](#) [tagging](#) [Tesseract](#) [tibco-hawk](#) [tradeweb](#)  
[Vernilion](#) [version.control](#) [Wiki](#)



### IT Core Webs

- [Applications](#)
- [Databases](#)
- [Scheduled Jobs](#)
- [Hardware/Servers](#)



### Team Sites

- **Application Support**
- [Corporate & Distribution](#)
- [Investment Systems](#)
- [Security Data](#)

### Links

- [New to K-Portal? Start here](#)
- [Interested? Like an overview?](#)
- [Add suggestions here](#)

Figure 7.2 – Tagging on the Organisations Knowledge Portal

Figure 7.2 shows the intuitive organisation of content on the K-Portal and not surprisingly shows the trading system content tagged most, an accurate reflection for a large investment organisation. Blogs created by pro-active users educate users on the spectrum of Web 2.0 technologies available and encourage the use of these technologies within the department.

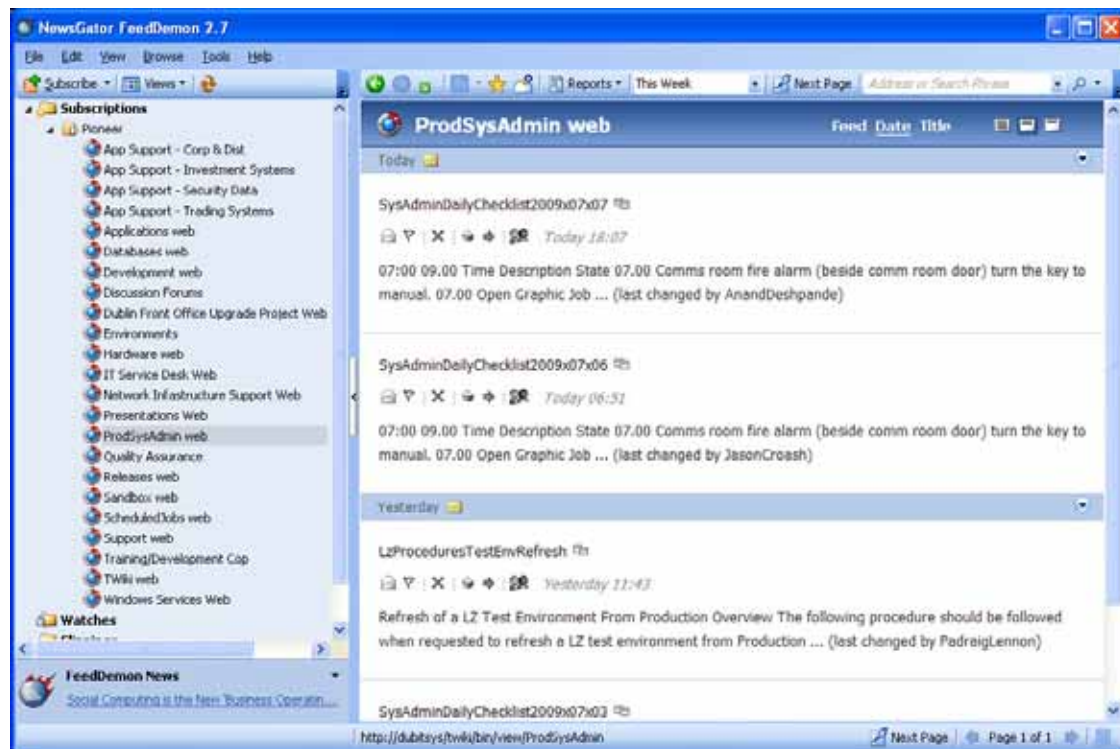


Figure 7.3 – RSS Feeds linked to on the Organisations Knowledge Portal

Figure 7.3 illustrates how RSS feeds provide an overview of content change on the ever-expanding K-Portal. The feeds provide users with a single location from where to view content changes of interest. It has also proven to be beneficial for offshore staff in keeping track of developments and changes on K-Portal as well as giving management an overview of documentation creation and alterations.

However, the built-in search engine within TWiki was not particularly powerful and did not work well with the K-Portal structure. To overcome this application weakness, the Google Search Appliance which existed within the organisation was connected to K-Portal with regular crawls and indexing scheduled against the knowledge base. The roll out was an improvement on the original search engine but too often the number of results returned was excessive, for example, a search phrase containing the name of the primary trading system of the organisation returns 10 pages of results consisting of 408 individual pages. As a result, an exploration into the field of semantics was proposed to evaluate its capabilities in providing an improved content retrieval solution.

## **7.4 Semantic Wiki**

The experiment is concerned with evaluating the benefits and drawbacks of utilising semantic tools in the process of knowledge creation and retrieval within the organisation when compared with the incumbent knowledge repository K-Portal.

### **7.4.1 Product Choice**

In order to carry out a balanced evaluation of the semantic technology against the existing technology, a comparable platform had to be selected for the experiment. A Semantic version of the Wiki system TWiki, which hosts the organisation knowledge repository, did not exist. A number of Semantic Wikis were evaluated at the outset of the experiment and it was found that the majority of these were not in a mature stage and thus proved unsuitable. Among the Wiki's evaluated were the KiWi (Knowledge in a Wiki) Wiki, Ace Wiki and the Semantic MediWiki. The KiWi Wiki was the original choice Wiki for the research experiment and assistance was received from the main developer of the product in installing and configuring the software. However, bug issues with certain aspects of the software during the migration of content to this platform led to a re-evaluation of the software. AceWiki was also evaluated but it was felt the user interface was poor. It was also based on a non-standard language so this ruled it out of the running. Semantic MediaWiki, a Semantic version of the popular MediaWiki, the open-source Wiki system behind Wikipedia, proved to be the most established Semantic Wiki technology to date. The recent commercialisation of the product provided extra confidence in it being the best choice for the experiment.

Markus Krotzsh and Denny Vrandecic developed Semantic Wiki as an extension to MediaWiki in order to make content on the popular website Wikiepdia machine-processable (Krotzsch, Vrandecic and Volkel 2006). Figure 7.4 illustrates the development of the Semantic extension from its origins in MediaWiki to the current commercial enterprise version SMW+ which is distributed by OntoPrise. The version implemented as part of the pilot was the open source version of SMW+.



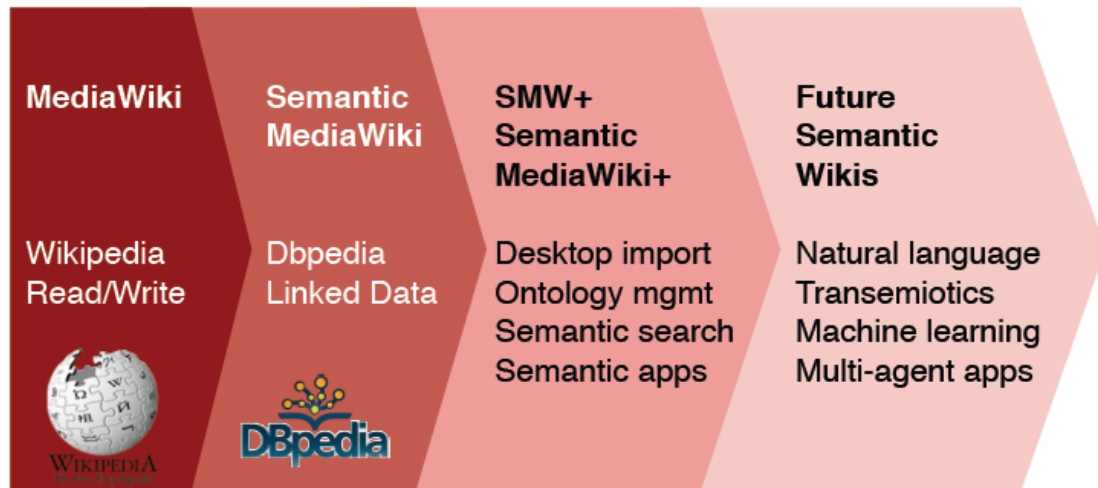


Figure 7.4 – Semantic MediaWiki Development Iterations (Source Project 10X)

#### 7.4.2 Semantic MediaWiki Design

Semantic MediaWiki is an attempt to merge the Wiki philosophy with semantic technology such as RDF(S) and Ontologies in order to capture or identify further information about the content and its relationship to other content. The Mediawiki engine has being enhanced with semantic capabilities that help to search, organise, tag, browse, evaluate, and share the Wiki's content.

The Semantic MediaWiki architecture is based on three core elements – categories, typed links and attributes.

- *categories, which classify articles according to their content,*
- *typed links, which classify links between articles according to their meaning, and*
- *attributes, which specify simple properties related to the content of an article.*

(Krotzsch, Vrandecic and Volkel 2006).

These elements will be discussed in more detail in the preceding section in relation to the web content of the experiment.

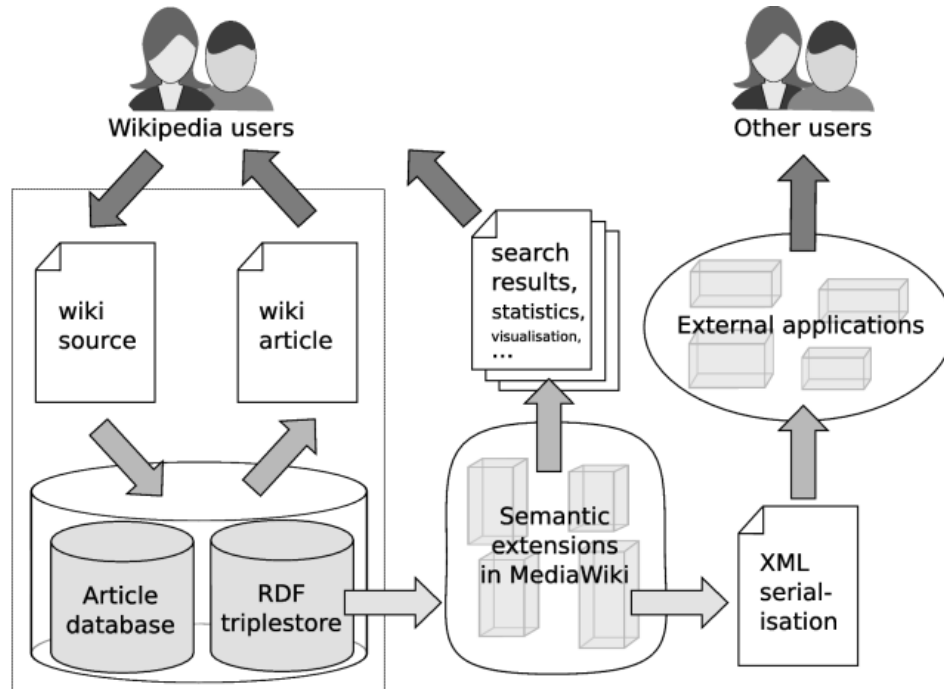


Fig. 7.5: Basic architecture of the semantic extensions to MediaWiki (Volkel *et al.*, 2006)

The architecture of the Semantic MediaWiki is built on the underlying architecture of the popular MediaWiki technology with extra functionality implemented. The backend MySQL database stores the content of the Wiki, be this standard content or semantically annotated content, which is stored in a form of RDF representation. The RDF content within the database is exploited through semantic searching on the database using a subset of SPARQL queries. This subset of functionality is known as the ASK query language. The results of the queries are presented in SQRQL format but it is not possible to use the full SPARQL functionality due to the absence of a fully fledged RDF on the Semantic Wiki (Krotzsch, Vrandecic and Volkel 2006).

### 7.4.3 Semantic MediaWiki MetaKnowledge Design

The technology of the Semantic MediaWiki represents a facade of the underlying technologies of the semantic web. While the Wiki supports both RDFs and Ontologies, it represents them in a different structure within the system.

The Semantic MediaWiki system represents metadata in a similar manner to RDF representation. Table 7.1 illustrates the mapping of RDF representation to the corresponding representation within MediaWiki.

SMW	RDF
Category:x	Rdfs:Class
Property:x	rdf:Property
[[Category:x]] (on category page)	Rdfs:subClassOf
[[Category:x]]	rdf:type

Table 7.1 - SMW: a variation of RDF

The Semantic MediaWiki supports Ontologies but similar to RDF's they must be stored in certain formats in order to work. Table 7.2 illustrates this representation of OWL constructs within the Semantic MediaWiki

OWL Construct	Semantic MediaWiki
Class	Category
Datatype property	Property
Object property	Property also
Class instantiation	Page categorization (e.g. [[Category:X]])
Subclass of	Category subcategorization (e.g. [[Category:X]] on a category page)
Individual	Article (in Main namespace)
Instantiated datatype property	Attribute annotation (e.g. [[X:=Y]])
Instantiated object property	Typed link (e.g. [[X::Y]])

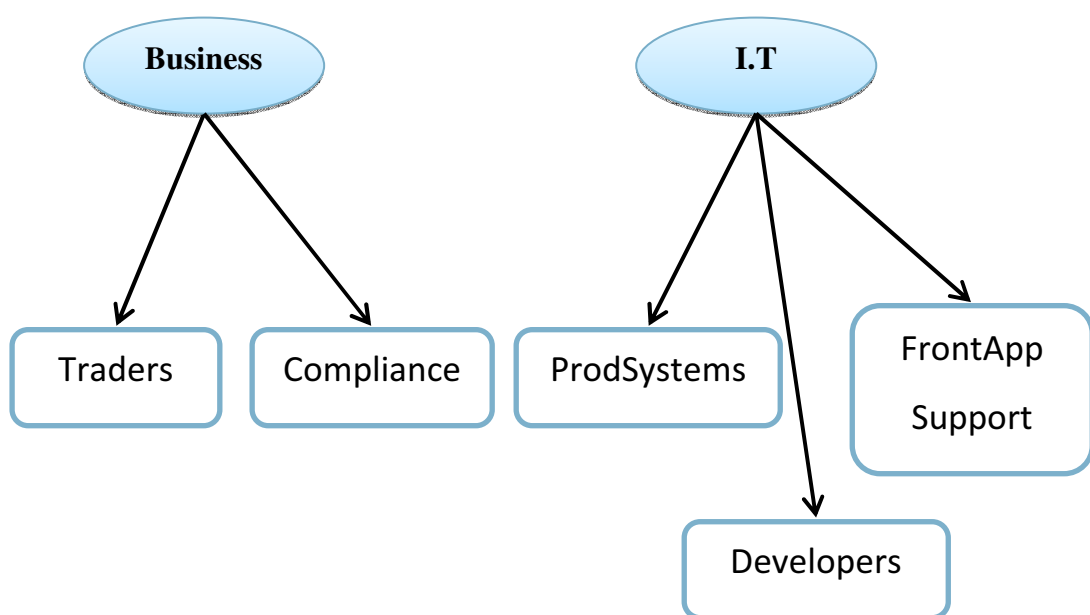
Table 7.2 - Ontology Construct in Semantic MediaWiki ([semantic-mediawiki.org](http://semantic-mediawiki.org))

The common elements in the mapping of RDF and OWL are Categories and Properties and this paper will now discuss both in relation to the research experiment.

## Categories

Categories are a feature used in MediaWiki used to group related pages. Adding a category tag to a page adds it to the relevant category which will list the page in the associated category page which provides a catalogue of the pages within the category (Krotzsch, Vrandecic and Volkel 2006). Semantic MediaWiki extends this functionality by utilising categories as a form of classification. This classification forms a similar function to the Class representation present in RDF.

For the purpose of the research experiment two parent categories or classes were created. These categories were Business and IT and represented a high level super class for all content. The content has then been further categorised into subsections of these areas so those with the knowledge in this area can edit the specific content. For example, content containing details on LZ services can only be edited by the Production Systems team while content relating to specific trade types can only be edited by Traders, Compliance and Front App Support. Figure 7.5 illustrates the categories and subcategories as they exist on the Semantic MediaWiki.



**Figure 7.5 – Semantic MediaWiki Categories (Source Author)**

## Properties

Properties are the standard way of entering semantic data in Semantic MediaWiki. They can be thought of as categories for values in a Semantic Wiki page.

They have the following syntax - `[[property name::value]]`.

As part of the research project a number of predefined properties were created by the author. Below is a list of these properties with brief explanation on each

`[[Pioneer Trading System::value]]` - represents a trading systems  
`[[Pioneer Trading Platform::value]]` – represents external broker platforms  
`[[Pioneer Service Account::value]]`– represents trading system service account  
`[[Pioneer Trade Type::value]]` – represents trade type i.e. bond, OTC  
`[[LatentZeroComponent::value]]` – represents a trading system component  
`[[LatentZeroServer::value]]` – represent a trading system server  
`[[LatentZero Service::value]]` – represents a trading system service  
`[[External Provider::value]]` – represents external vendor

By adding these semantic tags to the content, it enables the Semantic Mediawiki to interpret the data so a search on LatentZero servers will automatically retrieve all data tagged with this property and should return a list of all LatentZero servers if the data has being tagged correctly.

The categorisation and semantic tagging of content as described, was carried out solely by the author. An approach needs to be defined for collaboratively categorisation and representation of metadata. Scholz (2008) have enhanced the property function within the Semantic MediaWiki further by introducing the concept of property clustering which allows for the definition of a formal meta model for a knowledge domain. This brings the field of object orientated design into the Semantic Web but the definition of a naming convention for the properties/classes is absent.

A committee approach for defining property/class names and standards within a limited domain is an area for further research. A committee comprising of key stakeholders within the domain could manage and maintain the meta-knowledge representation. This approach presented itself during the research but was not part of the research definition.

## **7.5 Experiment**

The experiment is concerned with evaluating the benefits and drawbacks of utilising a Semantic Wiki titled “Compass” in the process of knowledge creation and retrieval within the organisation when compared with the incumbent knowledge repository hosted on the existing Wiki which is aliased as “K-Portal”. The content base for the experiment is content relating to LatentZero, the trading system of the organisation where the experiment was held. The LatentZero content on K-Portal was migrated to Compass and semantically annotated.

The following tasks were outlined for the Experiment User Base to complete

- Login to Compass (Semantic Wiki) and browse the content using the available menu's, do the same process for K-Portal (TWiki).
- Create test page on Compass (Semantic Wiki) and create similar page on K-Portal (TWiki).
- Edit a LatentZero page on Compass (Semantic Wiki) and edit same page on K-Portal (TWiki). Note - the exact same LZ content exists on both wiki's.
- Search for specific LZ information on the Compass (SemanticWiki) using the normal search option.
- Search for specific LZ information on the K-Portal (TWiki) using the Google search option.
- Search for specific LZ information on the Compass (SemanticWiki) using the semantic search option - help section available here - SearchHelp.

- Browse TeamPage on K-Portal and browse same same on the Compass (Semantic Wiki).
- View source code on page on K-Portal (TWiki) and view source code of page on the Compass (Semantic Wiki).
- Complete the Survey.

### ***7.5.1 Experiment User Base***

The experiment was carried out on a limited domain within the organisation and on a specific subset of content available to this domain. The involvement of knowledge users within this domain was therefore critical in the experiment. Several types of target users were identified and these users were then categorised firstly by role and then by department function.

The five types of users identified were

**FrontApp Support:** The Front Office Application Support team who provide level one Trading System technical support to the business.

**Production Systems:** the Production System Administration Team, who maintain the Trading System Infrastructure and provide support to FrontApp Support in their role.

**Developers:** The IT Development Team who develop new functionality on the Trading Systems, functionality requested from the business through the Front App Support Team.

**Traders:** The primary Business users of the Trading System.

**Compliance:** The secondary Business users of the Trading System who monitor the activity of the Traders on the System

These five types of users were then further categorised into two separate groups as discussed in section 7.4.3; Information Technology consisting FrontApp Support, Production Systems and Developers with Trader and Compliance forming the group Business.

## **7.6 Evaluation**

A number of different techniques were employed in evaluating the Semantic Wiki, which was piloted as part of this research. A survey was developed to collect feedback from those targeted in the pilot. The survey was embedded with the Compass System providing the user base with questions under 4 main areas – Background Information, Tool Evaluation, Search Functionality and Final Thoughts. Following the completion of the pilot a number of informal interviews were also conducted to garner further feedback on the Semantic Tools. The results of both were evaluated in terms of the research definition of the dissertation.

Consideration was also given to the hypotheses for Semantic Tool evaluation as Sure and Iosif (2002) outlined as part of their work on the Onto-Knowledge project. The list of testable hypotheses that was created for this project are outlined below and provide a good reference point for the evaluation of the research of this dissertation.

- 1. Users will be able to complete information-finding tasks in less time using the ontology-based semantic access tools than with the current mainstream keyword based free text search.*
- 2. Users will make fewer mistakes during a search task using the ontology-based semantic access tools than with the current mainstream keyword-based free text search.*
- 3. The reduction in completion and number of mistakes will be more noticeable for less experienced users.*



*4. The reduction in time will also be more noticeable for users lacking a detailed knowledge of the underlying technical system implementation.*

*5. The ontology-based semantic access tools will be perceived as more useful than free text search by different types of persons for a broad range of domains and knowledge-acquisition scenarios.*

*6. The effort for developing and maintaining the ontology and information structure will not significantly exceed the effort to develop and maintaining the free text approach*

(Sure and Iosif 2002)

The initial phase of evaluation concerned the interpretation of the Survey results

## **User Base Survey**

### **Section 1 – Background Information**

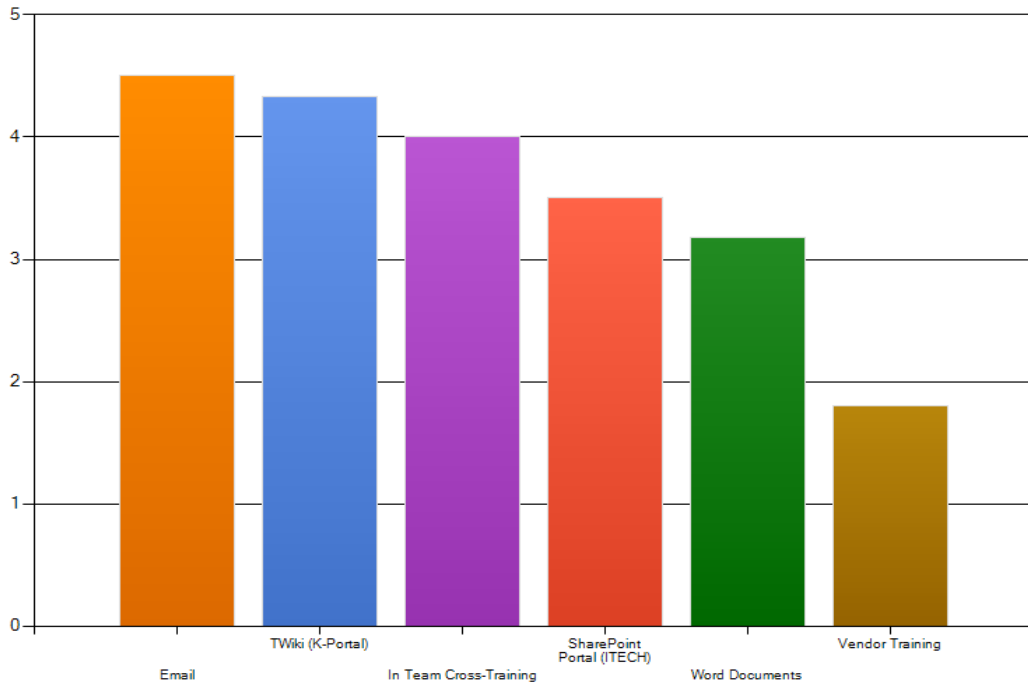
This section was concerned with gathering some initial information of the participant in relation to their role within the organisation and their familiarity around Web 2.0 & Web 3.0 Technologies.

*Question 1 - Which Area of the Organisation (Pioneer) do you work in?*

The Production Systems team made up 66.7% of the respondents to the survey with the two remaining teams within the IT department accounts for 16.7% each. This was expected, as the Production Systems team are the primary users of the existing Wiki system K-Portal and have built up a comfort level with the technology. Due to the relatively low response rate from the other teams within IT, informal interviews were carried out with key individuals from these teams to get feedback on the experiment. It is noted that no users from the Business Teams responded to the survey. These users do not actively use the existing Wiki K-Portal but were included in the survey with the intention of receiving a non-IT view of the respective technologies.

### *Question 2 – Collaboration Methods*

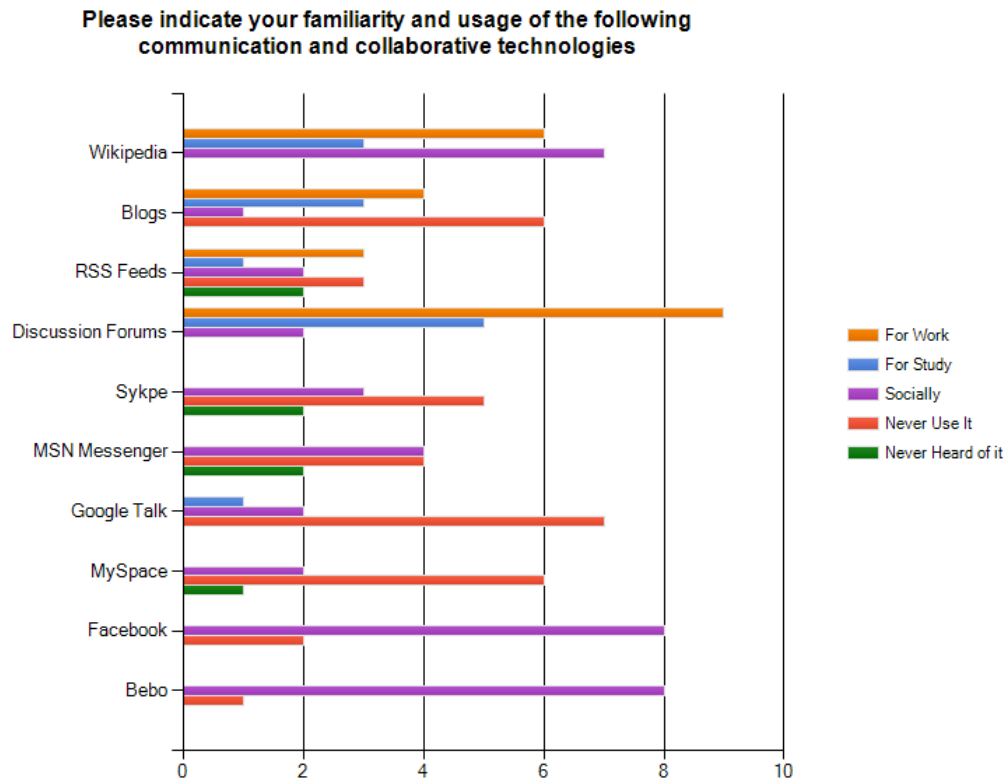
This question asked the respondent to rate tools and techniques used within their team to transfer knowledge.



Prior to the research it was widely known that e-mail was the tool of choice for transferring knowledge within teams but the results did yield positive results for K-Portal and also highlighted the importance of tacit knowledge transfer as part of In-Team Cross Training.

### *Question 3 - Communication and Collaborative technologies*

This question sought to delve deeper into the use of communication and collaborative technologies by the respondent and more specifically in what context they utilised the technology. The most popular brands of the relevant technologies were used in some cases as it was felt users would be more similar with the website\brand name as opposed to the underlying technology.

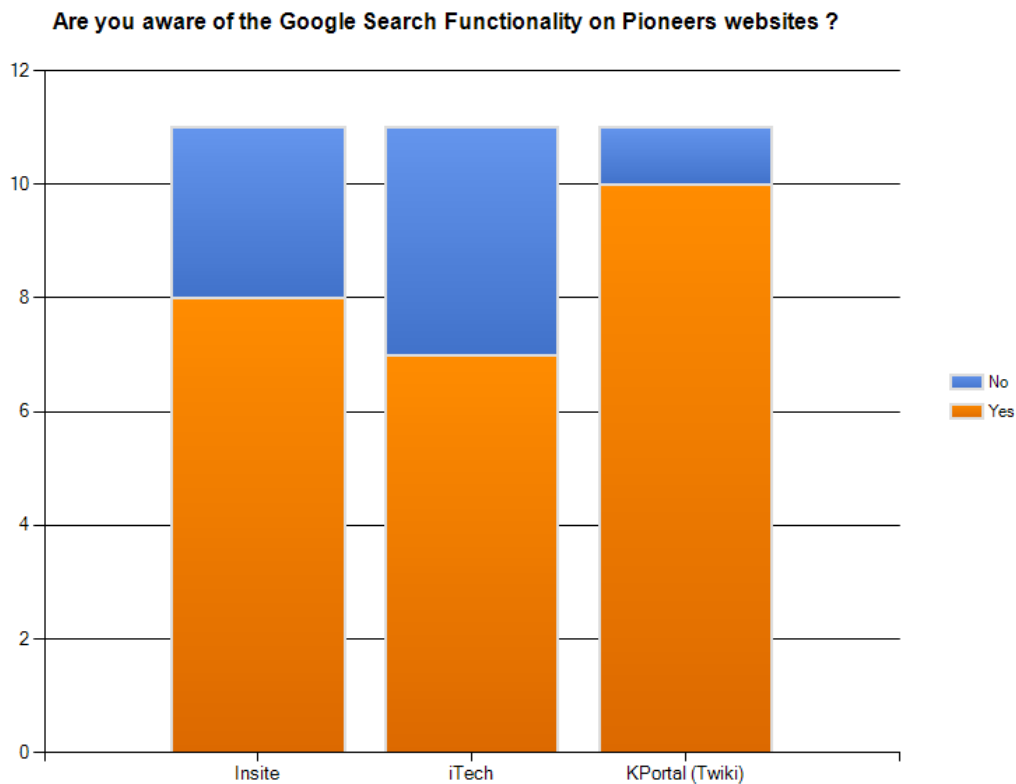


Interestingly it was found that discussion forums were the most heavily used technology for work followed by Wikipedia, Blogs and RSS Feeds respectively. However, in the case of Blogs and RSS Feeds, it is noted that an equal or greater percentage of respondents never use them. Specifically in the case of RSS feeds, a significant number of respondents have never heard of it. This is quite surprising given RSS feeds are used by a number of team members within Production Systems Team to monitor changes on K-Portal. Of the other technologies surveyed, it can be noted that technologies concerned with oral communication such as Skype and Google Talk fare quite poorly in terms of usage. Finally, the strong results seen with regard to Wikipedia infer a high level of familiarity with MediaWiki, the Wiki system on which Wikipedia is built and this needs to be considered when analysing the responses to the Semantic Wiki interface which is an extension of MediaWiki.

#### *Question 4 – Google Functionality*

This question queried the respondent's familiarity with the Google Search plug-in on the organisations different Intranet sites. The organisation has a Google Search

Application server which interfaces to all internal sites to provide full Google search functionality within the organisations web repositories.

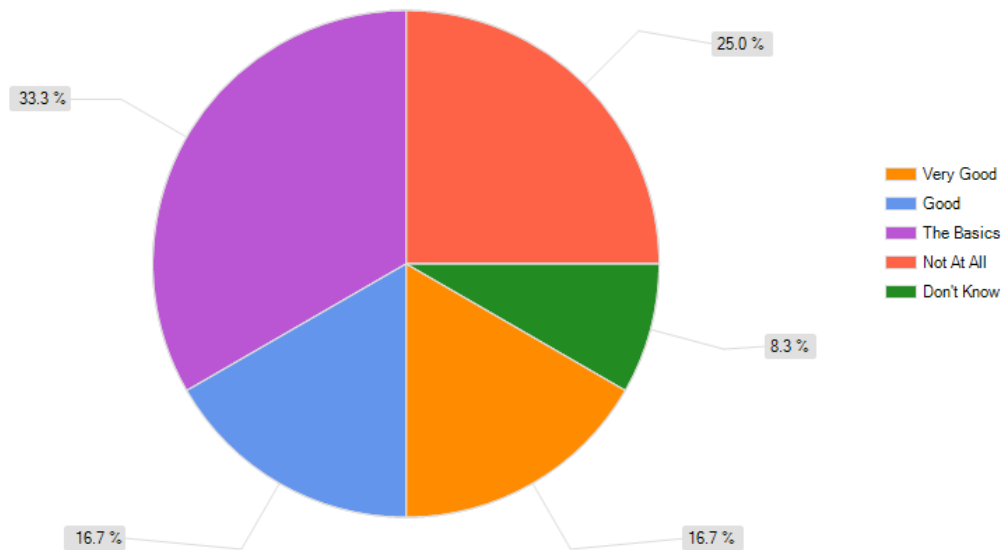


Users were in general familiar with the functionality across the internal sites though the iTech site looks to suffer somewhat from the placement of the Google interface on the site thus resulting in lowest awareness. As part of this question, respondents were also queried as to what they primarily used the Google search engine for with the majority stating they used it primarily on K-Portal and to a lesser extent use it to find official forms on the organisations official intranet portal “insite”.

#### *Question 5 – Semantic Web Familiarity*

Respondents were asked to state their familiarity with the Semantic Web with the majority reporting a basic understanding. However 25% of respondents had no familiarity with the Semantic Web while an equal number responded positively showing a good or very good understanding.

**Please rate your familiarity with the Semantic Web**



Following on from this, those who stated a familiarity were asked to expand on their understanding of the Semantic Web. While responses were not detailed, they did show respondents had a grasp of the technology. Words such as “*intelligent*” and “*understanding*” were common while one respondent termed it “*data mining instead of data search and retrieval where the web can be taught to answer questions instead of search for words*”.

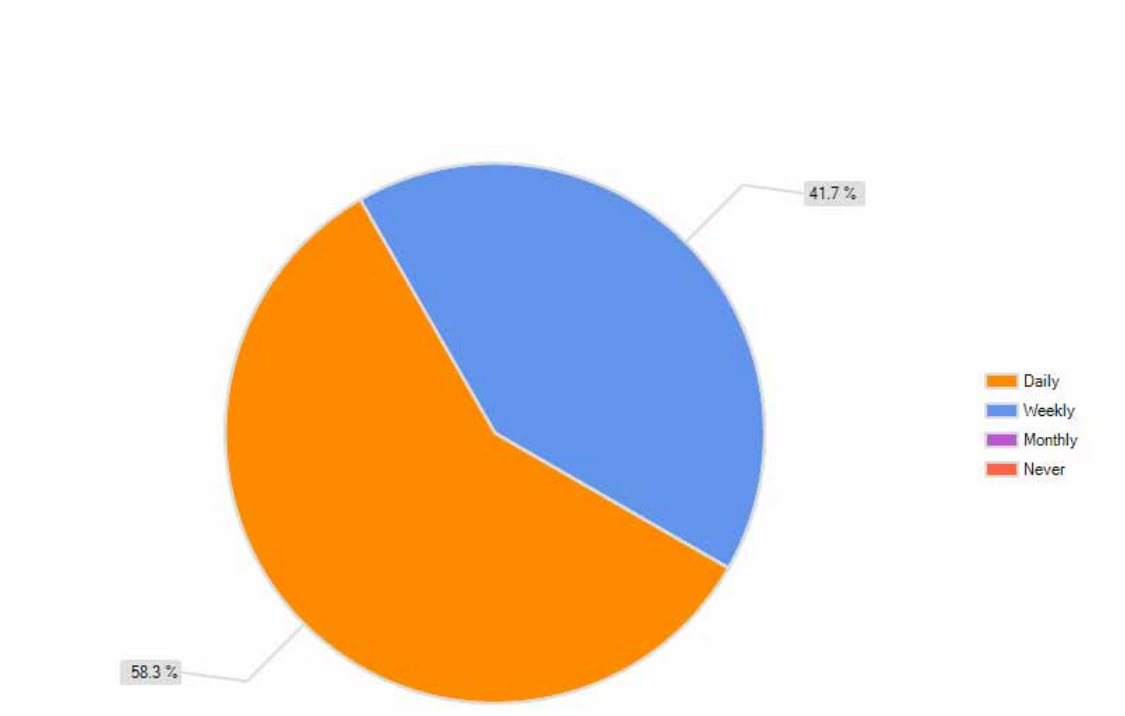
## **Section 2 – Tool Evaluation**

This section involved gathering information on the existing Wiki system (K-Portal) and the piloted Semantic Wiki (Compass).

*Question 1 – How often do you use the existing Pioneer Knowledge Base – K-Portal?*

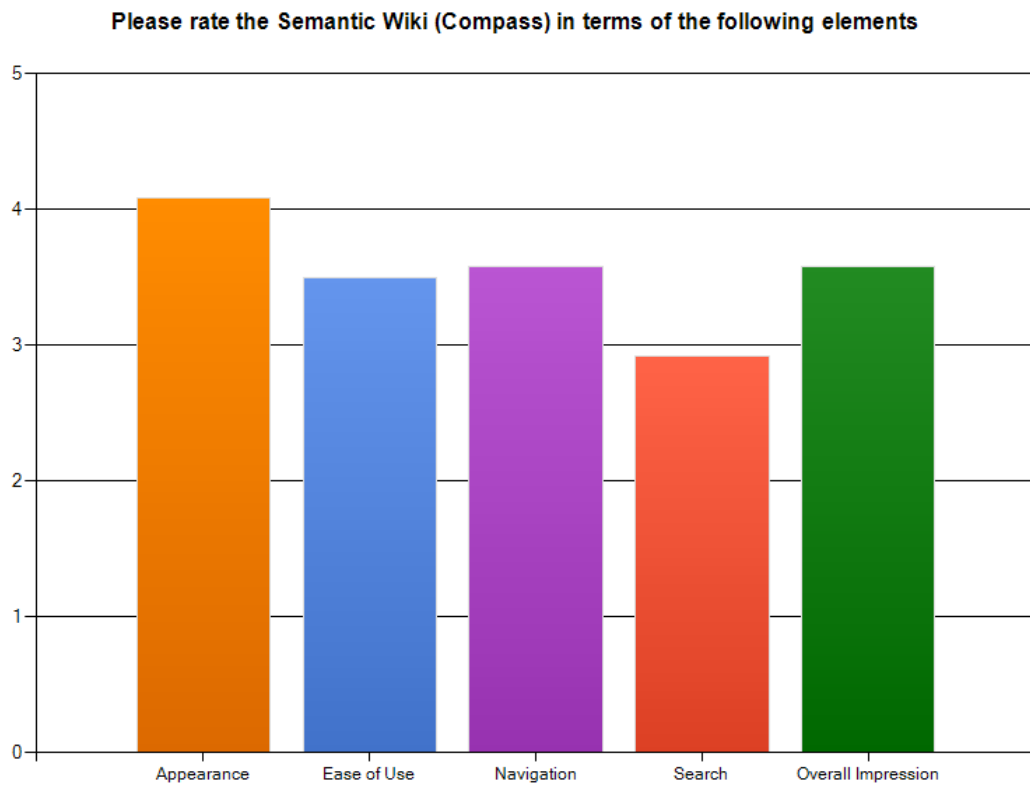
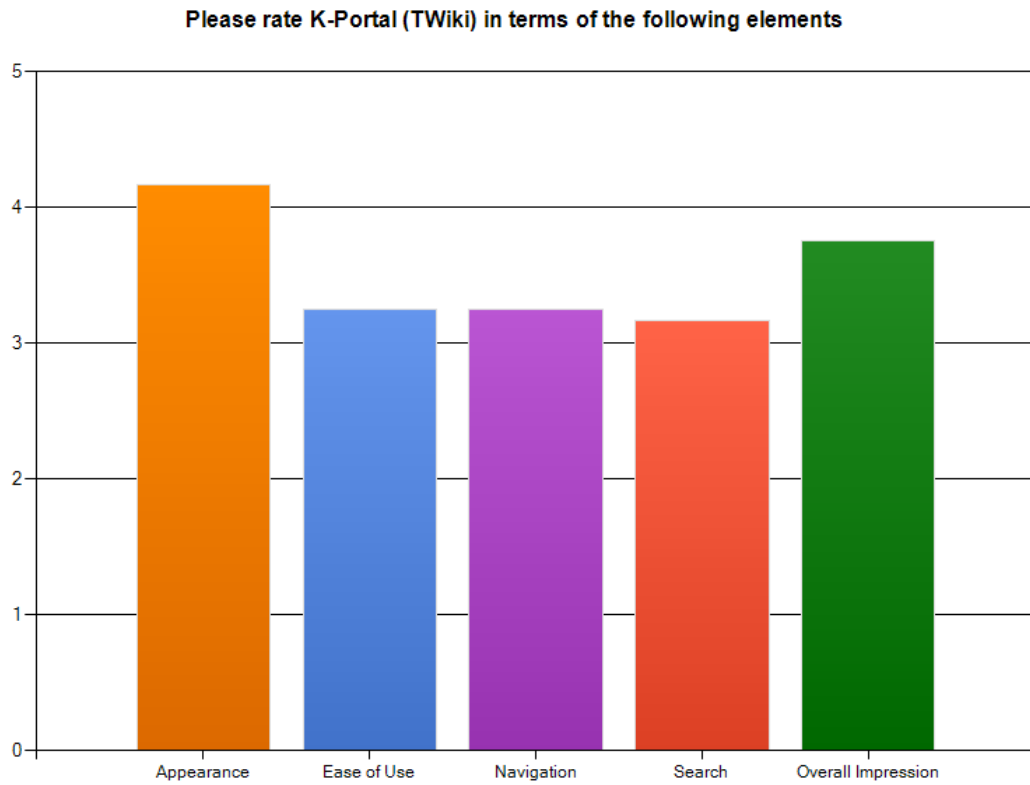
This question queried the respondents’ usage of K-Portal showing the majority use it daily with the remaining respondents using it at least once a week. This highlights the importance of K-Portal within the IT department as a source of information and

knowledge with comment feedback showing it is used by many to aid with their day to day tasks.



### *Question 2 & 3 – Interaction with K-Portal and Compass.*

Question 2 asked the respondents to rate key aspects of their interaction with K-Portal with Question 3 asking the same questions of Compass. The questions related to features including appearance, ease of use, navigation, search and overall impression.



Both systems rated similarly for appearance showing respondents to be quite comfortable with Wiki Interfaces.

Compass did rate higher in terms of Ease of Use and Navigation which could relate to respondents usage of Wikipedia as seen in the results to the Communication and Collaborative technology question in Section 1. As discussed in the analysis of the Communication and Collaborative technology question, the MediaWiki Wiki engine is common to both Compass and MediaWiki so respondents would have had prior exposure to this Wiki technology. Possible reasons for the lower rating received on K-Portal for these aspects will be explored in future questions.

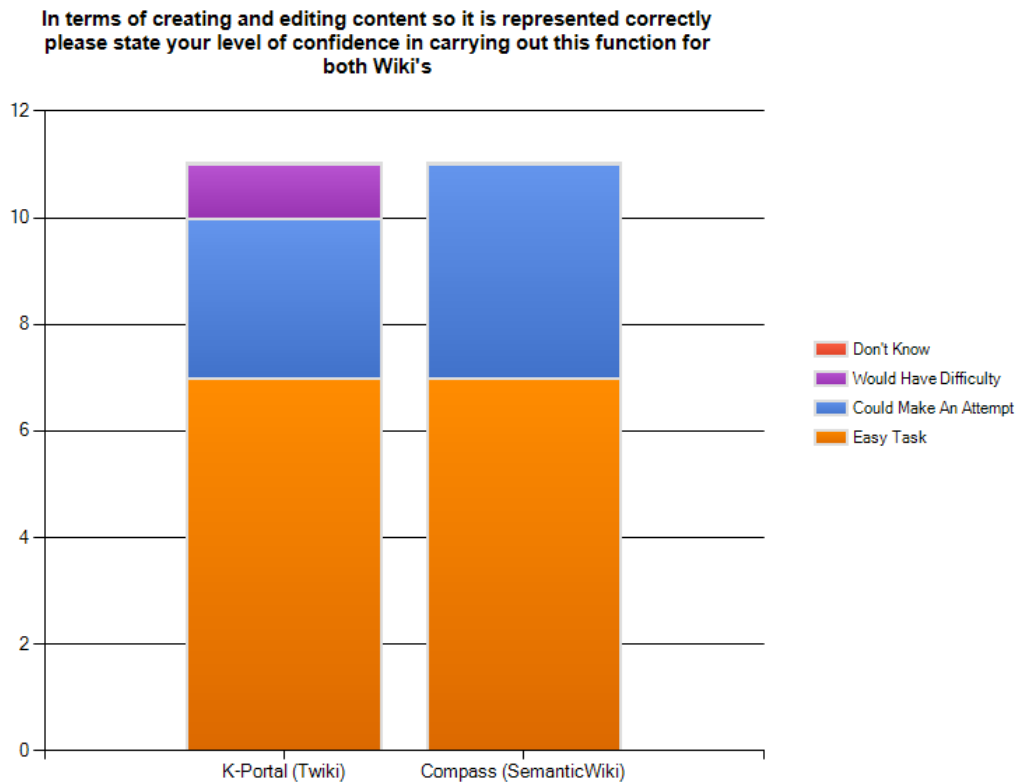
Again for both systems the search functionality rated lowest of all aspects with Compass's search feature receiving the poorest rating. This illustrates an issue around the search functionality of Wiki's and in particular the Semantic search. It shows the benefits of Semantic Search as outlined by Yang *et al.*, (2007) are not yet present in the semantic technology as it stands today.

Both systems received very good results in terms of overall impression which is positive for Wiki technologies role in the corporate workplace.

#### *Question 4 – Content Creation and Editing*

This question asked respondents to measure their level of comfort with content creation and editing on both Wiki Systems. As all respondents are from a technical background, it is not surprising to discover the majority of respondents found the Wiki editors easy to use. Some respondents, however, reported difficulty with “*adding non standard text such as tables and images*” to K-Portal. Hemphill and Yew (2007) reported similar issues with TWiki technology when they rolled the software out to a user community as part of a Wiki evaluation project.



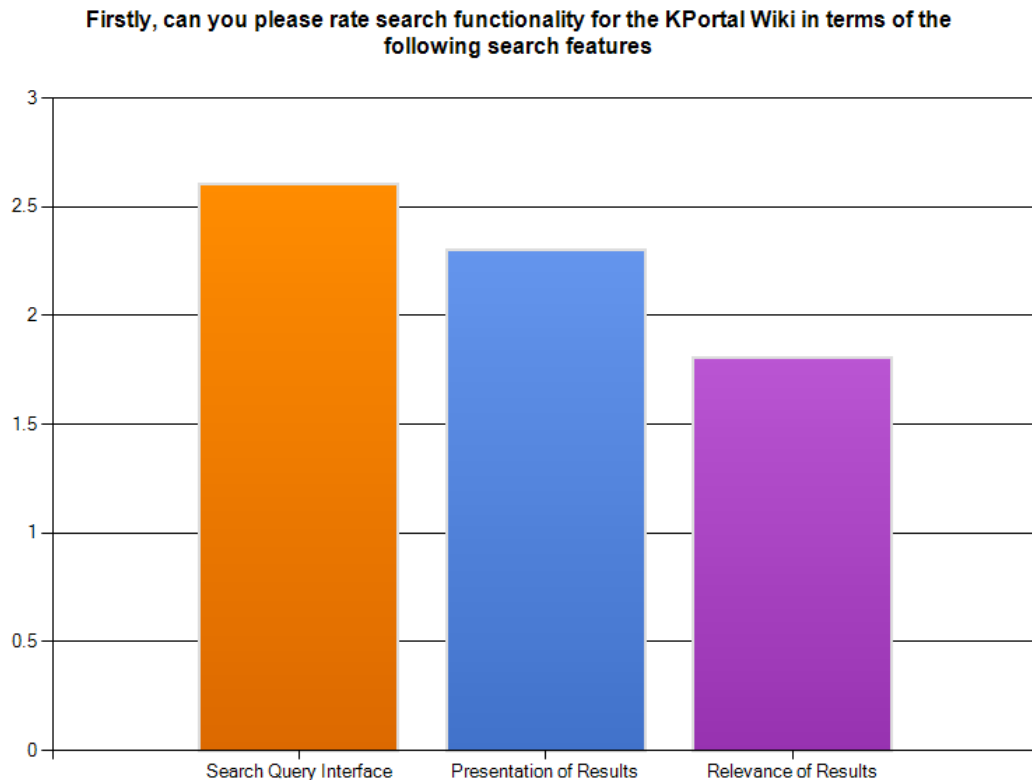


### Section 3 – Search Functionality

This section concentrated on the specific area of knowledge retrieval in order to determine whether the semantic technology provided improved search retrieval capabilities.

#### *Question 1 – K-Portal Search Functionality*

As discussed in the introduction to this dissertation, the original premise for carrying out the evaluation of Semantic Wiki was due to the perceived weakness of the Google search functionality on K-Portal. This was the basis for this question which asked respondents to rate the aspects of the Google Search functionality on K-Portal.



The respondents were asked to rate search functionality in terms of Search Query Interface, Presentation of Results and Relevance of Results. The search query interface rated highly with respondents stating that *“the interface is easy to use, the results are displayed as google results, easy to understand since google is now a standard almost”* which re-enforces the position Google holds on the Web (Cho and Roy 2004). Presentation of results also rated highly with results presented as per the Google standard with which respondents have a *“high level of comfort”*. In terms of relevance of results, however, the ratings were not quite as strong with respondents stating that the search *“returns too much data that may not be relevant to what you are looking for”* while *“case-insensitivity”* and *“wikiwords”* were also highlighted as potential issues for the result set.

#### *Question 2 – Compass Search Functionality*

The same question was presented to the respondents with regard to the Compass Search functionality which is based on a customised query language using inline queries (Ankolekar *et al.*, 2007). As this method of searching would not have being very intuitive to the user base, a special help page on the Semantic MediaWiki was

written outlining out the inline query worked and providing some examples on its usage as illustrated in Figure 7.7.

To list all pages containing LZ services and then list these servers enter the following query on the semantic search page.  
[[Pioneer Trade Type::+]] and ?Pioneer Trade Type will provide same results for Trades

special page

Query

[[LatentZero Service::+]]

Additional printouts (optional)

?LatentZero Service

[Add sorting condition]

Find results

Hide query | Querying help

Previous

Results 1– 5

Next

(20 | 50 | 100 | 250 | 500)

LatentZero Service

Execution Adapter

LatentZeroDerivatives

LZ PROD Minerva 4.0 40 02 Execution Adapter

LZ PROD Derivatives 1.1 01 Bond Calculation Server FINCAD01

LZ PROD Derivatives 1.1 01 Bond Calculation Server FINCAD02

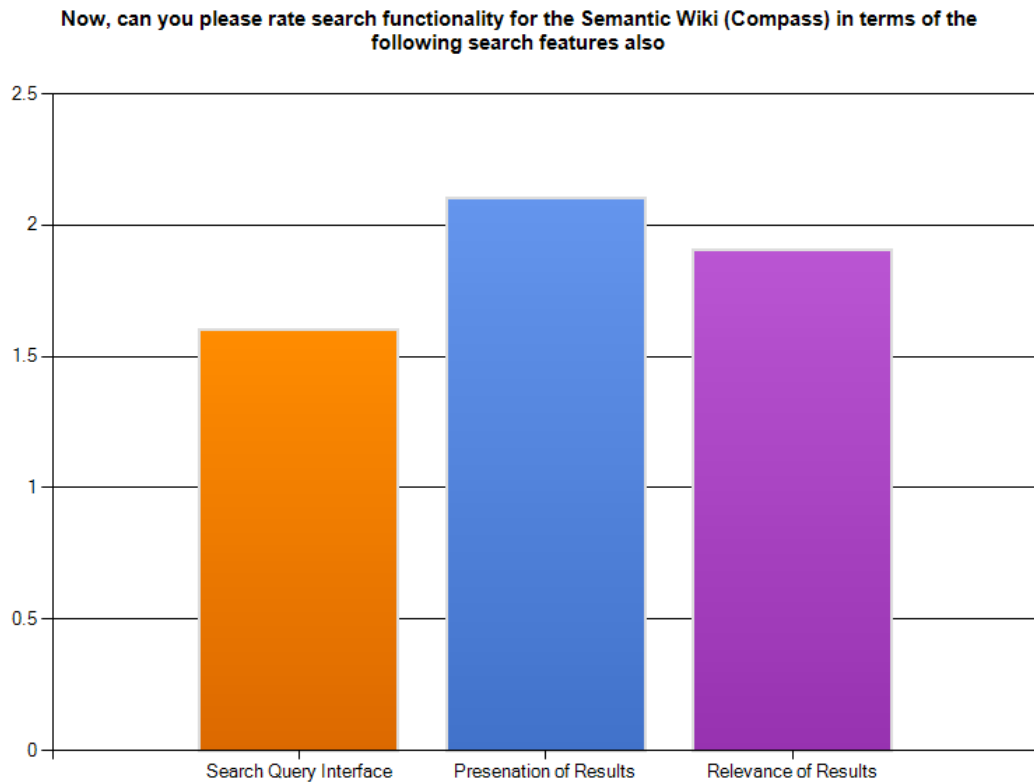
LZ PROD Derivatives 1.1 01 Bond Calculation Server FINCAD03

LZ PROD Minerva 4.0 01 Data Server 01

LZ PROD Minerva 4.0 01 Data Server 02

Figure 7.7 – Semantic MediaWiki Search Help Page

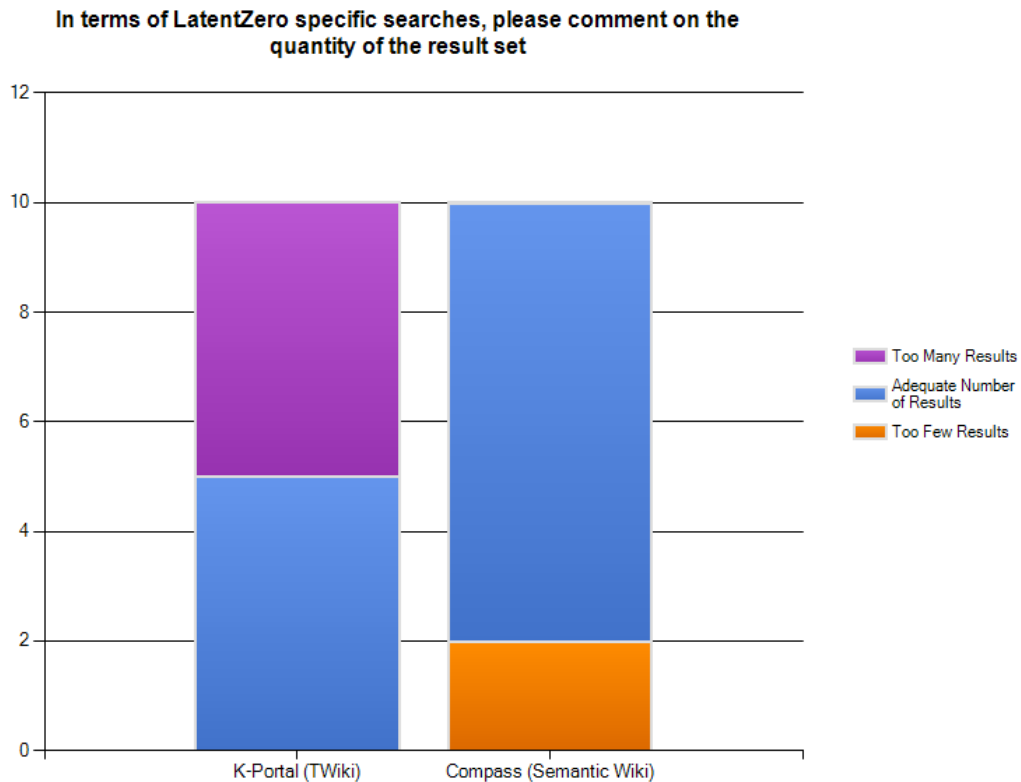
The search query interface, however, still received an unsatisfactory rating notwithstanding this, with respondents reporting that the “*search function is very difficult to use*” and argued that “*having to learn syntax for search is not intuitive at all*”.



While the majority of respondents stated that work is needed on the interface, those that did master the search queries stated that *“the results presented are very good”* and *“the manner in which it returns data are better than the Twiki”*. It was further stated that *“when the search criteria are defined correctly it only returns relevant information, which is better than the Twiki which returns too much”*. This positive feedback resulted in the Semantic Wiki receiving a superior rating to K-Portal in terms of result relevance and a comparable rating on the presentation of the results.

### *Question 3 – Trading System Result Set*

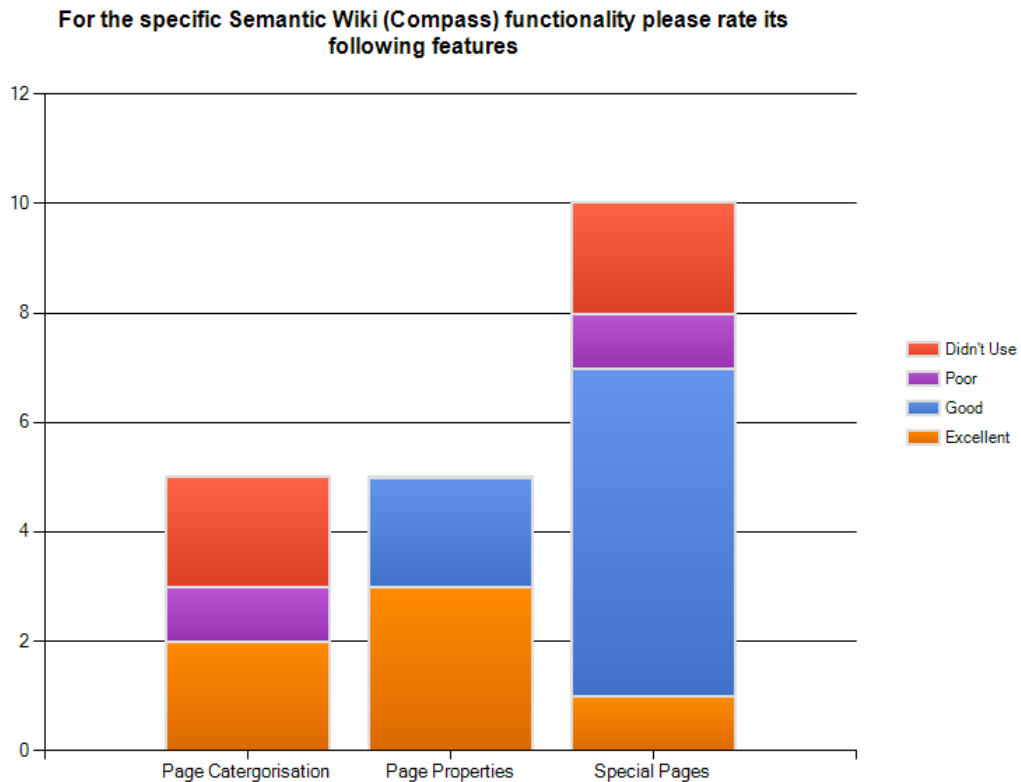
Building on the previous two questions, Question 3 asked respondents to comment on the quantity of the results they received on both systems when querying information on the organisations trading system. The same trading system content exists on both systems, but the content contains semantic annotations on the Semantic Wiki.



50% of the respondents felt too many results were returned on K-Portal while 20% felt that Compass, the Semantic Wiki returned too few results. The qualitative feedback received as part of question stated that “K-Portal provided a lot of duplicate results” but no qualitative feedback was received on the Semantic Wiki returning too few results.

#### *Question 4 – Specific Semantic Wiki (Compass) Functionality*

The final question in the search functionality section queried respondents on the specific semantic technologies that came packaged with the Semantic Wiki. These technologies discussed in Section 7.4.2 provide extra functionality in terms of viewing and searching content.



Firstly it must be noted that all respondents did not answer this question and of those that did, it can be noted that a significant percentage did not use the extra functionality. Of the extra functionality tested, Page Properties rated best with those who used it while Special Pages received more traffic from respondents and received positive feedback to an extent. Page Categorisation, a feature of MediaWiki, which has been reworked for the Semantic Mediawiki did not receive much attention and scored poorly with those who did evaluate it.

#### **Section 4 – Final Thoughts**

The final section of the survey was concerned with gathering qualitative feedback on the Wiki technologies used within the organisations as well as finding people's thoughts on knowledge management and its place in the organisation.

##### *Question 1 – Structure of Content on K-Portal (TWiki)*

Respondents generally felt the structure of K-Portal aided the retrieval of knowledge but some felt its functionality was over extended arguing that in some cases the tool

was used to carry out tasks not associated with a knowledge portal i.e. interfaces exist on K-Portal that enable the user to launch jobs on the enterprise scheduling system. Others felt more training was needed on the technology and argued that not all teams within IT were using it effectively.

#### *Question 2 – Structure of Content on Compass (Semantic Wiki)*

Feedback on the structure of the Semantic Wiki varied greatly with some respondents stating that the “*structure was immature*” while others felt it was “*more organised and easier to find info*” with “*knowledge results more sensible*”. One respondent argued that it was trying to imitate the structure of K-Portal and should have being designed on its own merits and while this is a valid point, it is felt that similarities needed to be present across both systems to undertake a fair evaluation.

#### *Question 3 – Weaknesses of K-Portal*

Respondents were asked what they considered were the main weaknesses of the current Wiki Solution – K-Portal. It was noted that some users struggled with “*with the user friendliness of editing or creating large pages of info*” and that the “*editing of documentation using the WYSIWYG editor was too unwieldy and structured*”. Several respondents stated that the search function is a key weakness and other functionality is overextended. At a high level, it was also stated that K-Portal didn’t have the approval from senior management and “*people simply don't have time to work with it*”.

#### *Question 4 – Weakness of Compass*

Similar to the last question, respondents were asked to consider to weaknesses of the Semantic Wiki Solution – Compass. The overwhelming feedback was the difficulty in using the search function with no mention made to other aspects of the Wiki.

#### *Question 5 – Knowledge Sharing within the Organisation*

This question simply asked whether knowledge sharing should be added to employees yearly goals within the organisation. All respondents agreed they should, which shows

and an appetite for knowledge sharing within the organisation and willingness on the employee's part to share knowledge once it is formally recognised.

#### *Question 6 – Knowledge Management Improvements*

The final question of the survey asked respondents for comments and thought on how knowledge sharing could be improved within the organisation. The majority of respondents cited senior management approval as a mandatory requirement. Training and encouragement were also mentioned as areas where improvements could be made. *“A complete shift in culture and a shared company-wide concept of what Knowledge is to the organisation”* was advocated with *“a single knowledge share promoted across the organisation”*. It was also felt incentives such as *“half day for the employee who enters the most info per month”* should be put in place to promote knowledge sharing. It was also highlighted that knowledge repositories such as Wiki's can't provide all the solutions for knowledge sharing and should be complemented by events such as *“team training meetings”*. In conclusion, it is advocated that knowledge management improvement can come about through the implementation of structures and standards for the KM technologies with time allocated to employees to partake in the knowledge management process. The key support from senior management must also be present for the various knowledge management initiatives.

#### **Informal Interviews**

The informal interviews reiterated a lot of the findings of the survey. Participants were open to the new technology and were positive on certain aspects of the technology. The issues with the search interface were restated but the potential for improved search results was recognised. Again the need for backing from senior management was highlighted as well as recognition for knowledge sharing within the organisation either through formal goals or a change in work practise.



## **7.7 Conclusion**

This chapter describes the phases undertaken as part of the experiment software pilot. It examines the technology already in place within the organisation and provides background on the technology chosen for the pilot. It then describes the design of the Semantic Wiki for the experiment and the framework used in describing the content. The final section of the chapter is a detailed evaluation of the results from the surveys and interviews carried out as part of the experiment. The final chapter of the dissertation will now discuss the conclusions from the research project in more detail.

## **8 CONCLUSIONS AND FUTURE WORK**

### ***8.1 Introduction***

The final chapter of this dissertation presents some conclusions and recommendations from performing this research project. The aim of this research was to evaluate the use of semantic technology in the creation and retrieval of web content in the context of Knowledge Management. This chapter summarises the dissertation by outlining how the research aims and objectives were achieved. The contributions to the wider body of knowledge by the research are presented. Finally the limitations of the research project are discussed and prospective areas for future research are put forward.

### ***8.2 Research Definition & Research Overview***

The primary focus of the research in this dissertation was on the technologies that aid in the field of knowledge management. The research was carried out within an organisation that utilises some of these technologies within certain departments for the purpose of sharing and creating knowledge. The technologies currently in place work well in achieving these aims in the finite domain, but limitations of the software may lead to potential issues as the knowledge base grows. The concern is that the representation of the knowledge content in its current format does not expose itself accurately for full exploitation by search and retrieval technologies such as search engines. The research sought to evaluate new technologies that could pre-empt these issues by providing a more formal structure for the creation of knowledge content; a structure that would merge with search technologies for intelligent content retrieval.

The research project began by performing a literature review on the history of the Web from its inception, through the different iterations of the Web. In particular, it focused on the content and structure of the Web through the evolution. A literature review was then carried out on search technologies with an emphasis on ranking algorithms which are used by the popular sites. The weaknesses with the ranking algorithm approach

were identified as part of this research. The research project finally performed an in-depth literature review of the Semantic Web paying particular attention to the semantic technologies.

The project then explored several semantic tools to prepare for the pilot of the Semantic Wiki. Numerous Semantic Wikis were evaluated with the assistance of the open source community and a choice was made in favour of Semantic MediaWiki. The content from the company's existing Wiki was then migrated to the Semantic Wiki and tagged using an annotation framework defined by the author. Further Semantic Extensions were then installed on the software and a survey tool was embedded in the application. The user base for the pilot was chosen and pages were written on the Semantic Wiki outlining the tasks of the experiment with examples. The pilot ran for a period of two months after which time survey results were analysed followed by informal interview with key users.

As a result, the following objectives have been achieved in this dissertation:

- Performed a literature review to examine the evolution of the Web from its beginnings to the present day. Reviewed topology and structure of the web in the context of web content
- Performed a literature review of current web search technologies and evaluate there strength and weaknesses. Explored Google's page rank algorithm from a number of different aspects.
- Considered and evaluated existing Web 3.0 tools in the context of Knowledge Management. Numerous tools were evaluated and trailed across a broad range of semantic technologies. Some semantic extensions were used with the Semantic Wiki chosen to provide better functionality.
- A framework was developed to represent Knowledge about the Knowledge – Meta Knowledge. A meta-knowledge framework was created for the content

within the domain of the organisation and a taxonomy compromise was proposed for the wider domain.

- The feasibility of generating a Semantic Wiki with AI tools was investigated to generate the relevant knowledge content for the end user. Semantic MediaWiki tool was piloted with limited success. Some AI functionality did exist within the application but it was not utilised.

### ***8.3 Contributions to the Body of Knowledge***

As part of this body of work an in-depth examination of the Web from its inception was performed analysing how its structure developed and how this structure has determined the way we traverse the web. As part of this, the search tools and recent web technologies such as Wiki's and Blogs that have altered our behaviour on the web were examined in a limited domain. The core search functionality of the larger World Wide Web was researched in depth highlighting the strengths and weaknesses of its underlying ranking algorithm while also looking at how the functionality is evolving.

Extensive research on the representation of web content was then undertaken in the context of the Semantic Web evaluating the categorisation and classification of web content. The new breed of technologies in this area were discussed and evaluated in a limited domain presenting interesting findings. Based on this research, a proposed framework for the representation of metadata on the Semantic Mediawiki was put forward.

Together with the research of the web technologies, the body of work also examined the people aspect of KM initiatives such as the Semantic Wiki pilot done as part of this research. A unique aspect to this work was the comparison of two Wiki technologies and the feedback provided some interesting insights on the role of technology with knowledge management.

#### ***8.4 Experimentation, Evaluation and Limitation***

The experiment central to the research project was the pilot of a Semantic Wiki – Semantic Media Wiki (SMW). As outlined in Section 7.5, the experiment involved the user base experimenting with the creation and retrieval of semantic tagged content on a Wiki platform in a limited domain. The Semantic Wiki was provided to the experiment user group with 20 pages of complete semantic tagged content that the author had migrated from existing content on the current Wiki used within the department. The user base were provided with a list of predefined tasks to complete on the Semantic Wiki and were encouraged to partake in a survey on completion of these tasks.

The results of the survey and feedback from informal interviews show that the semantic technology is not yet at a mature stage for rollout with an organisation. The experiment highlighted issues with the existing Wiki which needs to be addressed and lessons learnt on its implementation would be a required starting point for any future potential rollout of a Semantic Wiki. While the semantic technology is not yet ready for the end user, the experiment suggests that semantically tagged data does yield better results for the user and is an improvement on the current search technology. Questions still remain on how to best to semantically annotate the data and the ownership of this task but this is an area well suited for further exploration within a limited domain.

An analysis of the Semantic Wiki pages, post pilot phase, indicated that a number of users who participated did not create semantic tagged content and thus did not fully interact with the new technology. This was a limitation on any evaluation concerning semantic tag annotations. Content that was created by the user base was limited to test pages and very often the content was not annotated with semantic tags. Instead, it was found that existing content was edited retaining the predefined semantic tags that existed on the page at the outset of the experiment. No new semantic tags were created on the trading content was created by the experiment user group.

The user base that completed the experiment consisted of colleagues in the IT department with the business users who were asked to partake in the experiment but

declining to do so. It was established through informal discussions with the business users that the primary reason for not completing the experiment was due to time constraints and a lack of relevance of some of the content to this business. This presents a bias on the evaluation of the technology as the user base would be considered early adopters to any new technology based on their job function.

The maturity of the semantic technology also presented limitations for the project. A number of Semantic Wiki's were tested with support offered from the predominantly research based creators of the various Wiki's. However, it was found that the vast majority of the Semantic Wiki's were still very much in the Beta stage of software. The limitation of a number of viable alternatives narrowed the choice of technology considerably. The technology finally chosen did not employ fully formed ontologies or fully developed SPARQL functionality. Section 7.4.2 discusses the limitations of the chosen technology in greater detail but this must be considered in the evaluation.

An underlying limitation applies to this research as it was focused on one single organisation, with the Wiki technologies being used by internal employees only.

## ***8.5 Future Work***

During completion of the research project, a number of areas for potential future work and research presented themselves. First among these is the requirement for an in-depth evaluation of the incumbent Wiki solution within the organisation. It was discovered through informal interviews and survey responses that the users have strong thoughts and opinion on its structure, layout and design. An anonymous evaluation of all aspects of this technology would be very beneficial for the KM practitioners driving the technology and would help in improving it for its ever expanding audience.

The results from an appraisal of the existing Wiki could provide the foundations for a rollout of an improved Semantic Wiki over the next 18 months provided the semantic technology continues the rate of improvement that has been witnessed through the research project timelines. It would also be interesting to determine if business buy-in

could be achieved for any future rollouts as this could lead to a culture change within the organisation in its attitude to knowledge management. The survey results stress the appetite for KM within the IT department and it would be interesting to discover if this exists across the whole organisation. With this buy-in, the potential exists to use the KM tools in a collaborative environment with partners and other entities within the larger organisation group.

Finally the subject of ontologies is an area that requires further research, in particular, the area of ontological engineering. The creation and maintenance of ontologies is an area of difficulty for the Semantic Web and new methods need to be discovered in order to manage ontologies in the generic domains. The research group DERI have recently concentrated on this area and have created the SIOC ontology which is a major breakthrough for the social web. This ontology enables social data on the web to be semantically represented thus allowing for semantic applications to be created on the social web unleashing new potential for the underlying data. An extension of this research work, for example, would be the creation of trading system ontology for use in a limited domain.

## **8.6 Conclusion**

*“You can’t manage knowledge – nobody can. What you can do is to manage the environment in which knowledge can be created, discovered, captured, shared, distilled, validated, transferred, adopted, adapted and applied.”*

(Collison and Parcell 2005)

This research project sought to discover whether Semantic Web technologies aid in the creation and retrieval of web content in the context of Knowledge Management. It was an attempt to bring further structure to the content of the organisations knowledge portal. It achieved this to an extent and demonstrated improved content retrieval on the new system. The structure imposed on content created was not enhanced by the pilot user group and the experiment may have benefited from a blank canvas in relation to this aspect of the experiment. The improved content retrieval was achieved on content created by the author and it would have been beneficial to have seen these results for

new content. While the research did demonstrate improvements in content retrieval, it has been found that the Semantic Technology is not at a mature phase for rollout within an organisation. Some argue that Semantic Technology is still a lot of theoretical hype but this research project has demonstrated that certain aspects of it do work as designed. Further work is needed, however, to bring it into the early majority phase of the technology adaptation cycle.



## BIBLIOGRAPHY

- Ankolekar, A., Krotzsch, M., Tran, T., Vrandecic, D., 2007 The two cultures:mashing up Web 2.0 and the Semantic Web, in: *Proceedings of the 16<sup>th</sup> International Conference on World Wide Web, Banff*.
- Arasu, Novak, Tomkins and Tomlin, 2002. PageRank Computation and the Structure of the Web: Experiments and Algorithms. In *Proceedings of the Eleventh International World Wide Web Conference*
- Auer *et al.*, 2007. Dbpedia: A nucleus for a web of open data.  
*Proceedings of International Semantic Web Conference ISWC07*
- Bar-Ilan, J. 2007. Manipulating search engine algorithms: the case of Google. *Journal of Information, Communication & Ethics in Society*, 5(2).
- Benjamins,VR.,*et al*, 2002. Six Challenges for the Semantic Web. In *KR2000 Semantic Web Workshop*.
- Bergman, M. 2000. *The deep Web: Surfacing the hidden value*. BrightPlanet,  
Available at: [www.completeplanet.com/Tutorials/DeepWeb/index.asp](http://www.completeplanet.com/Tutorials/DeepWeb/index.asp).
- Berners-Lee, 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*, HarperOne.
- Berners-Lee *et al.*, 2006. A framework for web science. *Foundations and Trends in Web Science*, 1(1).
- Berners-Lee. T & Cailliau R, 1990. *WorldWideWeb: Proposal for a HyperText Project*. Available at: <http://www.w3.org/Proposal.html>
- Berners-Lee, Hendler & Lassila, 2001. The Semantic Web: Scientific American.  
*Scientific American Magazine*.

- Bianchini, M., Gori, M., and Scarselli, F. 2005. Inside PageRank. *ACM Transactions on Internet Technology*, 5(1).
- Bizer *et al.*, 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*.
- Bojars *et al.*, 2007. Using the Semantic Web for Linking and Reusing Data Across Web 2.0 Communities. *The Journal of Web Semantics, Special Issue on the Semantic Web and Web 2.0*
- Bozsak *et.al*, 2002. KAON - Towards a Large Scale Semantic Web, In *Proceedings of EC-Web 2002, volume 2455 of Lecture Notes in Computer Science*, pages 304–313. Springer,
- Brickle, D and Guha, R.V. 2000. Resource Description Framework (RDF) Schema Specification 1.0.  
Available at: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>
- Brin, S & Page, L, 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1).
- Broder, A., 2002, A taxonomy of web search. *ACM Special Interest Group on Information Retrieval Forum*, 36(2)
- Business Wire , 2008. Forrester: Global Enterprise Web 2.0 Market To Reach \$4.6 Billion By 2013. *Business Wire*. Available at: <http://www.forrester.com/Research/Document/Excerpt/0,7211,43850,00.html>
- Cafarella, *et al.*, 2008. Web-scale extraction of structured data. *ACM SIGMOD Recor*, 37(4).
- Casarez *et al.*, 2008. Reshaping Your Business with Web 2.0. In McGraw-Hill Osborne Media.

- Cerf, V & Kahn, R., 1974. A Protocol for Packet Network Intercommunication. *Communications, IEEE Transactions on*, 22(5), 637- 648.
- Chapman, M. 2007. Google's dominance 'will not last forever' - Information World Review. *Information World Review*. Available at: <http://www.iwr.co.uk/information-world-review/news/2191158/google-dominance-forever>
- Cho, J., Roy, S., and Adams, R.E., 2005. Page quality: in search of an unbiased web ranking. In *Proceedings of SIGMOD Conference*
- Cho, J and Roy, S., 2004. Impact of search engines on page popularity. In *Proceedings of the international conference on World Wide Web '04*
- Chopin, 2008. Finding communities: alternative viewpoints through weblogs and tagging. *Journal of Documentation*, 64(4), 552 - 575.
- Chu, H. and Rosenthal, M. 1996 Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology. *Proceedings of ASIS 1996 Annual Conference*.
- Collison, C. and Parcell, G. 2005. *Learning to Fly: Practical Knowledge Management from Leading and Learning Organizations*, Capstone 2<sup>nd</sup> Ed
- Cormode, G. and Krishnamurthy, B. 2008. Key Differences between Web1.0 and Web2.0. *First Monday*, 13(6).
- Davies *et al.*, 2005. Next generation knowledge access. *Journal of Knowledge Management*, 9(50).
- Davies, J. and Weeks, R. 2004 QuizRDF: Search Technology for the Semantic Web. In *Proceedings of the 37th Annual Hawaii International Conference on System*

*Sciences.*

Ding, L. and Finin, T, 2006. Characterizing the SemanticWeb on the web. In *Proceedings of the International Semantic Web Conference (ISWC)*.

Ding, L and Fensel, 2001. Ontology Library Systems: The key for successful Ontology Reuse. In *Proceedings of the the 1st Semantic web working symposium*, Stanford, CA, USA

Duffy, J. 2003. Google calls in the 'language police'. *BBC*. Available at: [http://news.bbc.co.uk/2/hi/uk\\_news/3006486.stm](http://news.bbc.co.uk/2/hi/uk_news/3006486.stm)

Echarte, F. *et al.*, 2004. Ontology of Folksonomy: A New Modeling Method. *Proceedings of Semantic Authoring, Annotation and Knowledge Markup (SAAKM)*.

Etzioni, O, 1996. The World-Wide Web. *Communications of the ACM*, 39(11), 65-68.

Fielding R., *et al.*, 1999. Hypertext Transfer Protocol -- HTTP/1.1. W3c. Available at: <http://www.w3.org/Protocols/rfc2616/rfc2616.html>.

Goldbeck, J., and Hendler, J. 2004. Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-based Social Networks. *Lecture Notes in Computer Science*.

Google 2009, Marking up structured data - RDFa. *Google*. Available at: <http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=99170>.

Gruninger, M., Bodenreider, O., Olken, F., Obrst, L., and Yim, P. 2008. Ontology Summit 2007 - Ontology, taxonomy, folksonomy: Understanding the distinctions. *Applied Ontology*. 3(3)

Gruber and Thomas, 2007. Ontology of Folksonomy: A Mash-Up of Apples and Oranges: *International Journal on Semantic Web & Information Systems*,

3(1).

Guha, R., McCool, R., and Miller, E. 2003. Semantic Search. In *Proceedings of the 12th international conference on World Wide Web*.

Gulli, A. and Signorini, A., 2005. The indexable web is more than 11.5 billion pages. In *Proceedings of the 14th international conference on World Wide Web (WWW)*, ACM.

Halpin, H., Robu, V., and Shepherd, H. 2007. The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*.

Hargittai, E., 2004. Classifying and coding online actions. *Social Science Computer Review*, 22(2).

Hausenblas, M. 2009. Exploiting Linked Data to Build Web Applications. , *IEEE Internet Computing* 13(4).

Hepp, M., Bachlechner, D. and Siorpaes, K., 2006. OntoWiki: community-driven ontology engineering and ontology usage based on Wikis. *International Symposium on Wikis*.

Heflin, J, 2001. *Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment*. University of Maryland, College Park.

Heflin, J & Hendler, J., 2001. A portrait of the Semantic Web in action. *Intelligent Systems, IEEE*.

Heflin, J. *et al.*, 2001. SHOE: A Prototype Language for the Semantic Web. *Computer and Information Science*.

Henzinger, M., Motwani, R., and Silverstein, C., 2002. Challenges in Web Search

Engines. *SIGIR Forum* 36(2)

Hepp, M., Siorpaes, K. and Bachlechner, D, 2007. Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management. *Internet Computing, IEEE*, 11(5).

Hirsch, C. *et al.*, 2009. The Visual Wiki: A New Metaphor for Knowledge Access and Management.

In *Proceedings of the 42nd Hawaii International Conference on System Sciences*

HitWise 2007, Google Share of US Searches Up to 65% in April - MarketingVOX. *Marketing Vox*. Available at: <http://www.marketingvox.com/google-share-of-us-searches-up-to-65-in-april-029570/>.

Hoffman *et al.*, 1995. Commercial Scenarios for the Web: Opportunities and Challenges. *Journal of Computer-Mediated Communication*.

Horrocks, I. and Bechhofer, S., 2007. Semantic Web. *University of Manchester*.

Introna, L.D., and Nissenbaum, H., 2000. Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society: An International Journal*, 16(3).

iProspect, 2008. Search Engine Marketing Research: iProspect Blended Search Results Study. Available  
[http://www.iprospect.com/about/researchstudy\\_2008\\_blendedsearchresults.htm](http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm)

Jacso, P. 2005. Google Scholar: the pros and the cons. *Online Information Review*, 29(2), 208-214.

Kiefer, C. 2007. Imprecise SPARQL: Towards a Unified Framework for Similarity Based Semantic Web Tasks.

In *2nd Knowledge Web PhD Symposium (KWEPSY) co-located with the 4th*

- Kinsella, S. et al., 2008 From web 1.0 to web 2.0 and back -: how did your grandma use to tag? In *WIDM*, pages 79–86
- Klein *et al.*, 2002. *Ontology Versioning and Change Detection on the Web, Proc. 13<sup>th</sup> European Conf. Knowledge Eng. and Management (EKAW 2002)*, Springer-Verlag
- Krotzsch, M., Vrandečić, D. and Volkel, M. 2006. Semantic MediaWiki. *Proceedings of the Fifth International Semantic Web Conference*, pp 935-942, Springer
- Krotzsch, M., Vrandečić, D., Volkel, M., 2006,: Wikipedia and the Semantic Web, Part II. In: *Proceedings of the 2nd International Wikimedia Conference*
- Lawrence, S. and Giles, C.L., 2000. Accessibility of information on the Web. *Intelligence*, 11(1).
- Levy, M., 2009. WEB 2.0 implications on knowledge management. *Journal of Knowledge Management*, 13(1).
- Macaskill and Owen, 2006. Web 2.0 to Go. In *LIANZA Conference, Wellington*.
- Mace, S. et al., 1998. Weaving a better Web. *Byte*, 23(3), 58-66.
- Madhavan, J., Ko, D., Kot, L., Ganapathy, V., Rasmussen, A., and Halevy, A.Y., 2008 Google's deep web crawl. *In the Proceedings of the VLDB Endowment*, 1(2)
- Malhotra, Y., 2005. Integrating knowledge management technologies in organizational business processes: getting real time enterprises to deliver real business performance. *Journal of Knowledge Management*, 9(1), 7-28.
- Maness, J.M. 2006. Library 2.0 Theory: Web 2.0 and Its Implications for Libraries.

*Webology*, 3(2).

Mathes, A., 2004. Folksonomies – Cooperative Classification and Communication Through Shared Metadata. *Computer Mediated Communication*.

McAfee, A.P. 2006. Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review*.

McGuinness, D.L. *et al.*, 2002. DAML+OIL: An Ontology Language for the Semantic Web. *IEEE Intelligent Systems*, 17(5).

McLean D, 2007. The new web: rewards and risks for business. *PC World*.

Available at:  
<http://www.pcworld.ca/news/column/45dcb02d0a01040800cd2bfc8c6a6d79/pg0.htm>

McKinsey, 2007. How companies are marketing online: A McKinsey Global Survey - McKinsey Quarterly - Marketing - Digital Marketing. Available at:  
<http://www.mediamath.com/docs/MQ%20online%20marketing.pdf>

Merholz, P 2004. Metadata for the Masses. *Adaptive Path*.

Available:  
<http://www.adaptivepath.com/publications/essays/archives/000361.php>

Mills Davis 2009, *Web3 And The Next Internet - New Directions And Opportunities For STM Publishing*. Project 10X. Available at: <http://project10x.com/>

Mohamed, M., Stankosky, M., and Murray, A., 2006. Knowledge management and information technology: can they work in perfect harmony? *Journal of Knowledge Management*, 20(3).

Musser, J and O'Reilly, T. 2006 *Web 2.0 – Principles and Best Practices*, O'Reilly Radar, Tech. Rep.



- Nardi, B.A *et al.*, 2004. Why we blog. *Communications of the ACM*, 47(12).
- Ntoulas, A., Cho, J., and Olston, C. 2004. What's New on the Web? The Evolution of the Web from a Search Engine Perspective. In *13th WWW ACM*.
- O'Reilly 2005, What is Web2.0. Design patterns and business models for the next generation of software. O'Reilly Media, 2005.  
Available at: [www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html](http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html).
- Ohmukai, I., Hamasaki, M., and Takeda, H. 2005. A Proposal of Community-based Folksonomy with RDF Metadata. *Workshop on End User Semantic Web Interaction co-located with the Fourth International Semantic Web Conference (ISWC2005)*
- Olsson, O. 2007. The Semantic Web – An overview. *Swedish Institute of Computer Science(SICS)*. Available at:  
[http://www.w3c.se/resources/office/talks/20071113/071113-oredev\\_0\\_3a.pdf](http://www.w3c.se/resources/office/talks/20071113/071113-oredev_0_3a.pdf)
- Perez, J., Arenas, M., and Gutierrez, C. 2006. Semantics and Complexity of SPARQL . *The Semantic Web - ISWC 2006*.
- Peter F. Patel-Schneider & Ian Horrocks, 2006. A Comparison of Two Modelling Paradigms in the Semantic Web. In *Proceedings of the Fifteenth International World Wide Web Conference (WWW 2006)*, ACM, 2006, 3-12.
- Peters, I. and Stock, WG. 2007. Folksonomy and Information Retrieval. In *Proceedings of the American Society for Information technology*.
- Prud, E. and Seaborne, A. 2005. SPARQL Query Language for RDF. Available at <http://www.w3.org/TR/2005/WD-rdf-sparql-query-20050217/>.

- Pujari, P 2006. Independent Study Report on Reputation Systems.  
*Johns Hopkins University., USA*
- Quint, B., 2002. Searcher's Voice- Librarians and Google.  
Available at: <http://www.infoday.com/searcher/sep09/voice.shtml>
- Raggett, D. and Jacobs, I. 1999. HTML 4.0 Specification. W3C. Available at:  
<http://www.w3.org/TR/1998/REC-html40-19980424/>
- Relihan, L., Cahill A., and Hinchey, M.G. 1994. Untangling the World-Wide Web.  
*ACM Special Interest Group for Design of Communication.*
- Ruthfield, S., 1995 The Internet's history and development: from wartime tool to fish-cam. *Crossroads*, 2(1).
- Schueler, B. *et al.*, 2008. Querying for meta knowledge. In *Proceeding of the 17th international conference on World Wide Web.*
- Sheth *et al.*, 2007. SPARQ2L: towards support for subgraph extraction queries in rdf databases. In *Proceedings of the 16th international conference on World Wide Web.*
- Shirky, C., 2006: Ontology is Overrated -- Categories, Links, and Tags. *Shirky*.  
Available at: [http://www.shirky.com/writings/ontology\\_overnated.html](http://www.shirky.com/writings/ontology_overnated.html)
- Silverstein, C. *et al.*, 1999. Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33(1).
- Singel, R. 2005. Are You Ready for Web 2.0? *Wired*. Available at:  
<http://www.wired.com/science/discoveries/news/2005/10/69114>
- Stabile, SJ, 2008. Google Benefits or Google's Benefit  
*Journal of Business and Technology Law, University of Maryland*

- Stojanovic, L., Stabb, S., and Studer, R. 2001. eLearning based on the SemanticWeb. *World Conference on the WWW and Internet (WebNet)*, Orlando, Florida
- Sure *et al.*, 2002. OntoEdit: Collaborative ontology development for the Semantic Web. In *First International Semantic Web Conference (ISWC 2002)*, volume 2342 of *LNCS*, pages 221–235. Springer
- Sure and Iosif, 2002. On-To-Knowledge: Ener Search Virtual Organization Case Study: Evaluation Document. On-To-Knowledge Deliverable
- Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. 1st ed. New York: Doubleday
- Tapscott and Williams, 2006. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio, Penguin, 2006.
- Tredinnick, L, 2006. Web 2.0 and business: a pointer to the intranets of the future? *Business Information Review*, 23(4), 228-34.
- Van Damme, C., Hepp, M., and Siorpaes, K., 2007. FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. *Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)*
- Vetter R. J., Spell C., and Ward C., 1994. Mosaic and the World- Wide Web, *IEEE Computer Magazine*, Vol. 25, pp45-60.
- Volkel, M., Krotzsch, M., Vrandečić, D., Haller, H., and Studer, R. 2006. Semantic Wikipedia. In *Proceedings of the 15th international Conference on World Wide Web*
- Walsham, G. 2001. Knowledge management: the benefits and limitations of computer

systems. *European Management Journal*, 19(6)

Wiguna, W.S., Fernandez-Tebar, J.J., Garcia, A., 2006, Using a Fuzzy Model for Combining Search Results from Different Information Sources to Build a Metasearch Engine. In *Proceedings of the Computational Intelligence, Theory and Applications International Conference*.

Wikipedia Wiki , *Wikipedia*. Available at <http://en.wikipedia.org/wiki/Wiki>

White, D. 2005. Technical Approaches for the SPIRE project. *JISC Repositories Programme, Oxford University*.

Wilson, 2002. The nonsense of 'knowledge management'. *Information Research*, 8(1).

WordPress Blog Stats. Available at: <http://en.wordpress.com/stats/>

Wusteman, J.2004. RSS: the latest feed. *Library Hi Tech*, 22(4), 404-413.

Xu, Z. *et al.*, 2006. Towards the Semantic Web: Collaborative Tag Suggestions. *WWW2006*, May 22–26

Yang, C., *et al.*, 2007. Improving the search process through ontology-based adaptive semantic search. *The Electronic Library*, 25(2).

Yu, J. *et al.*, 2008. Understanding Mashup Development. *Internet Computing, IEEE*, 12(5), 44-52.

Zhdanova *et al.*, 2007. Combining RDF Vocabularies for Expert Finding . *The Semantic Web: Research and Applications*.

# APPENDIX A

This was the survey which users who took part in the evaluation responded to.

SemanticWikiEvaluation

[Exit this survey](#)

1. Background Information

This section gathers some background information. Note all responses are anonymous

1. Which area of Pioneer do you work in ?

☐ Production Systems

☐ Compliance

☐ Front App Support

☐ Trading Desk

☐ IT Development

Other (please specify)

List any other tools/techniques here

**3. Please indicate your familiarity and usage of the following communication and collaborative technologies**

	For Work	For Study	Socially	Never Use It	Never Heard of it
Wikipedia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Blogs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RSS Feeds	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Discussion Forums	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sykpe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
MSN Messenger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Google Talk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
MySpace	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Facebook	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bebo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**4. Are you aware of the Google Search Functionality on Pioneers websites ?**

	Yes	No
Insite	<input type="radio"/>	<input type="radio"/>
iTech	<input type="radio"/>	<input type="radio"/>
KPortal (Twiki)	<input type="radio"/>	<input type="radio"/>

If yes, please describe primarly what you use the tool for

**The Semantic Web**

The Semantic Web is an idea of World Wide Web inventor Tim Berners-Lee that the Web as a whole can be made more intelligent and perhaps even intuitive about how to serve a user's needs. To achieve this the web has to be able to describe things in a way that computers can understand. The Semantic Web does this by describing the relationships between things (like A is a part of B and Y is a member of Z) and the properties of things (like size, weight, age, and price).

By doing so it describes the data of the web in a way that computers can understand thus opening up endless possibilities for interpretation and use of the data.

**5. Please rate your familiarity with the Semantic Web**

☐ Very Good

☐ Good

☐ The Basics

☐ Not At All

☐ Don't Know

**6. If familiar with the Semantic Web, please give a brief description of your understanding of it.**

## 2. Semantic Web Evaluation

Evaluation of the existing Web 2.0 technology (KPortal) with the Web 3.0 Semantic Technology (Compass)

### 1. How often do you use the existing Pioneer Knowledge Base - <http://k-portal>

- ☐ Daily  
☐ Weekly  
☐ Monthly  
☐ Never

For given answer, please state how frequent e.g. 1-2 times daily

### 2. Please rate K-Portal (TWiki) in terms of the following elements

	Excellent	Very Good	Good	Average	Poor
Appearance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ease of Use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Navigation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Search	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall Impression	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 3. Please rate the Semantic Wiki (Compass) in terms of the following elements

	Excellent	Very Good	Good	Average	Poor
Appearance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ease of Use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Navigation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Search	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall Impression	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 4. In terms of creating and editing content so it is represented correctly please state your level of confidence in carrying out this function for both Wiki's










	Easy Task	Could Make An Attempt	Would Have Difficulty	Don't Know
K-Portal (Twiki)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compass (SemanticWiki)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please add any additional comments here

### 3. Search Functionalty










Comparison of the search functionality of the two Wikis - Kportal and Compass

**1. Firstly, can you please rate search functionality for the KPortal Wiki in terms of the following search features**

	Excellent	Good	Poor
Search Query Interface			
Presentation of Results			
Relevance of Results			

Where possible, please explain reason for rating

**2. Now, can you please rate search functionality for the Semantic Wiki (Compass) in terms of the following search features also**

	Excellent	Good	Poor
Search Query Interface			
Presentation of Results			
Relevance of Results			

Again where possible, please provide reason for the rating



### 3. In terms of LatentZero specific searches, please comment on the quantity of the result set

	Too Few Results	Adequate Number of Results	Too Many Results
K-Portal (TWiki)			
Compass (Semantic Wiki)			

Please add any additional comments here

The below question only relates to the Semantic Wiki as these features only exist on the semantic platform.

### 4. For the specific Semantic Wiki (Compass) functionality please rate its following features

	Excellent	Good	Poor	Didn't Use
Page Categorisation				
Page Properties				
Special Pages				

Additional Comments

## 4. Final Thoughts

1. Please comment on the structure of the data on TWiki (K-Portal). Do you feel the structure aides or hinders the retrieval of knowledge ?

2. Please comment on the structure of the data on the Semantic Wiki (Compass). Do you feel the structure aides or hinders the retrieval of knowledge ?

3. What do you consider the main weaknesses, if any, are with the current Wiki Solution - KPortal ?

4. What do you consider the main weaknesses, if any, are with the Semantic Wiki Solution - Compass ?

**5. Do you feel Knowledge Sharing should be made part of an Employees goals within Pioneer**

☐ Yes

☐ No

**6. Any final comments/thoughts on how the knowledge sharing could be improved within Pioneer**

