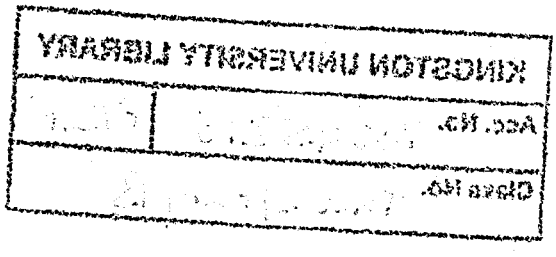
A Framework for Semantic Information Search and Discovery in Enterprise E-Commerce Applications

Biju John

A thesis submitted in partial fulfillment of the requirements of Kingston University for the degree of Doctor of Philosophy

Faculty of Computing, Information Systems, and Mathematics
Kingston University

February 2011



Abstract

Semantic web is an extension of the current web which is envisaged to offer a machine readable information space, where intelligent engines perform sophisticated tasks that can help in managing the massive and complex information on web. The shift towards the use of semantic web and semantic information is important in many areas including E-commerce enterprise applications. However, major issues remain particularly in terms of semantic information grouping, maintenance, storage and automated reasoning and retrieval support that is required for the verification, ranking and validation of E-commerce products and services. Moreover, many E-business and E-commerce approaches are deficient in a number of ways. These include the lack of semantic guidelines for building websites into a business, lack of semantic information search and real-time information leading to a lack of appropriate levels of interaction between the customers and the entrepreneur. Hence, the main aim of this research is to develop a framework for semantic information search and discovery in E-commerce enterprise applications. The proposed framework offers a new approach for semantic grouping, storing, retrieval and ranking of information semantically.

Thus, to achieve its aims the research work started with an investigation into existing E-commerce websites and their evaluation using a software quality model which showed that many sites suffer from a lack of appropriate information linkage and real-time interaction that are required for building semantic E-commerce systems and for information searches. To address these issues a semantic architecture including a semantic layer with an intelligent network based engine is proposed and implemented in this thesis. The architecture supports an automated semantic grouping and retrieval system and produces semantic ranking of information while continuously checking the information semantically and suggesting possible enhancements to the user. The experimental evaluation results, using a case study based on an E-commerce business application, demonstrated the benefits of the approach particularly in terms of semantic search and semantic ranking of the products. Moreover, although this information search approach was applied in the context of E-commerce it is generic which makes it appropriate for many online applications and websites.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to extend my gratitude to my director of studies, Dr. Souheil Khaddaj, for his continuous support and devoted cooperation throughout the course of this work. I also would like to thank Dr Andrea Hoppe for his ideas and objective criticism which have significantly contributed to this work.

Chapter 1 - Introduction

1.1 Historical Background

During the last decade the internet has rapidly evolved into the main medium of communication and information exchange. Its use has spread into almost every aspect of society, ranging from entertainment to commerce. Today the internet and other forms of web technologies provide access to information and are used by millions of people around the world. The web is considered the most popular medium of communication; even for the laymen it is easy to publish or access information. According to a survey made by Pingdom (Pin09), it is estimated that there are around 1.4 billion internet users worldwide, and approximately 210 billion e-mails are sent each day, which this does not include spam e-mails. The number of websites added during 2009 was 31.5 million (Pin09, Won10). The internet growth has been massive and it is continuously expanding.

However, the internet had a humble beginning as part of a research project at the Massachusetts Institute of Technology (MIT) in the early 1960s (Kle61, Kle62, Kle64). It began from the research work on packet switching at MIT and then the Defence Advanced Research Project Agency (DARPA) undertaken by Bolt, Beranek, and Newman (Lic68, Law67). The development of the internet is linked with the cold war. In 1957, when the Soviet Union launched Sputnik 1, the first satellite, into space, the United States of America was alarmed, that this could mean that the Soviet Union could launch bombs into anywhere on earth. These concerns in the US military did lead to the establishment of DARPA in 1958. DARPA's first priority was to protect American research in technology, safeguarding information against a space-based missile attack and thus taking a technological superiority over the Soviet Union. As part of these efforts DARPA appointed Joseph Licklider as the head of the new Information Processing Techniques Office (IPTO) and his job was to create programs and develop technologies to protect the US against space-based nuclear attack which was the first direct contribution to the development of the internet (Lic68, Ste08, Mit97). It was Licklider's vision to have a countrywide communications network, and therefore DARPA asked Lawrence Roberts, who was doing research with networks, to carry out further development in this field. Thus, Roberts founded the ARPANet (Advanced Research Project Agencies

Network) architecture based on the new idea of packet switching (Law67). The ARPANet first went live in October 1969, with communications between the University of California in Los Angeles (UCLA) and the Stanford Research Institute (SRI). The first networking protocol used on the ARPANet was the Network Control Protocol (NCP) (Cro70, Kle69). In the late 1960's it was even named as Resource Sharing Computer Networks. Later in 1970's it was generally considered as the inception of the internet.

Although the ARPANET was established only in 1969, it is generally regarded as the forerunner of the present day internet and its first successful demonstration took place of the International Computer Communication conference (ICCC) in Washington in 1972. The placement of the initial ARAPANET sites at research institutions and universities were regarded as ground-breaking research. The key services the first sites provided were: UCLA - Network Measurement Centre, SRI - Network Information Centre, UCSB - Culler-Fried interactive mathematics. The initial presentation of ARPANET in Washington paved the foundation for the development of present day internet and the web.

In 1974, TCP (Transmission control program/protocol) was formed and later it was developed into TCP/IP protocol, which is still the standard used today (Vin74, Vin78). In 1990, the National Science Foundation took over management of what was then called the NSFNet, and significantly expanded its reach by connecting it to the CSNET in universities throughout North America, and later to the EUnet throughout research facilities in Europe. With NSF's free-thinking management and the internet growing popularity, the nature of the internet changed quickly. In 1992, when the U.S. government began pulling out of network management, commercial entities offered internet access to the general public for the first time.

In 1980, Tim Berners Lee at the European Organisation for Nuclear Research (CERN), Switzerland, built ENQUIRE, a personal database for people and software models, and also to experiment with hypertext (Ber96). New information in ENQUIRE had to link with an existing page. In 1984 CERN had taken seriously the problem of information exchange and presentation, that is physicists from around the world had to share information, with no common machines and no common presentation software. In 1989 Bernard Lee made a proposal based on a large hypertext database with typed links (Ber92, Ber96). This project had generated little interest and later, asked by Mike Sendall, Bernard Lee tried to implement his system on a newly acquired NeXT workstation (Ber94, Ber96). Several names were

proposed such as Information Mesh, The Information Mine or Mine of Information, but the name World Wide Web was accepted (Ber96, Ber99).

According to Berners Lee (Ber 98) the web is an abstract information space. Within this space one can navigate and seemingly move around the world of information. The World Wide Web (WWW) is global information medium which users can navigate, read, write, publish, and retrieve information by the use of internet. The term is very often used together with the internet and therefore the World Wide Web is often mistaken for the internet or the other way. The World Wide Web is only a service over the internet which is a large number of networks grouped together and capable of sharing information (networks of networks). The networks are made up of computers and cables, which can be seen and therefore it is physical. But Vinton Cerf and Bob Kahn (Vin74) tried to make the concept more informative by sending around little "packets" of information around the network. Vinton pointed out that a packet is a bit like a postcard consisting of an address that is capable of identifying the destination and be delivered around the world.

The internet can be used by different types of programs such as e-mail, video conferencing, streamed audio channels, voice over IP etc. These programs are able to encode information and use protocols and languages to communicate with each others. The web, however, is an abstract space of information or imaginary space of information. For example over the web we may find information of various kinds such as sounds, pictures, videos and documents. The connecting links between the websites are hyperlinks. The web is unable to communicate without programs and protocols and it cannot exist without the internet, in fact it made the internet more informative and interesting.

1.2 Challenges

The internet and web massive growth have lead to the expansion of information technology itself. However, the web was originally designed purely as an information space useful for humans and processed by machines. One of the most difficult part of this understanding is that most of the information on the web is designed for human consumption (Ber98, Ber01). This means that the automatic integration of information from across the web and meaningfully linking information between computers remain a challenging task. Moreover, many of current web technologies are not able to efficiently integrate information about the

real world real objects, relationships and abstract concepts e.g. people, publications and the relationship between them (Sha05, Euz02).

The internet and other forms of web technology are continuously evolving and indeed have massively influenced the society, and the instant and easy availability of information is also revolutionising the peoples' imagination and lifestyles. However, the immense growth and availability of information makes it increasingly difficult to find and extract meaningful information. As pointed out by Petra Wilson (Pet02) "trying to get information from the internet is like drinking from a fire hose, you don't even know what the source of the water is." As a result, unpredictable expansion of the internet is taking place, because of the high demand for information by the humans (Oec08). However, the overload of information can create a number of problems including the accuracy of information, trust and authorship, origin of data etc. (Cra09, Pet02).

These problems, including scattered information and the question of trust, have affected the web usage, particularly the behaviour of search engines. For example a search engine crawling over stored information would not know how to effectively interact with sites, indeed search engines cannot penetrate deep into the pages and see what is trustworthy, reliable and meaningful. This is due to the fact that the traditional/current search engines use an Information Retrieval (IR) technology based search which is purely on the occurrence of the word in documents. Google, augment this in the scenario of web with information about the hyperlink structure of the web (Guh05). Google like searches are often based on Crawler-based search engines, that means this search engine can create its lists automatically and it will crawl or spider the web and come up with listings and then the user has to search through what the search engine has found. Another category of search engine is the human powered ones, which rely on humans to submit information that will be indexed and catalogued, and only the input information is indexed. These search engines can only produce the output from its index. The user search result can be narrowed or manipulated, the user is not searching the web, and rather the search engine searches the previously stored information from a giant database. As a result the user sometimes receives links which are not active, working anymore or broken links.

Moreover, the question of trust is leading to the meaning of information and the enormous volumes of data over the net increasingly make it difficult to access, present and maintain the information. In fact, computers have no reliable way to process the information (Pet02, Ber 01).

The scattered information and the question of trust over the internet experienced by the present day internet users were already under consideration when the paper by Bernard Lee, Hendler and Lassila containing a vision of the future of the internet was published in Scientific American in 2001 (Ber01). In this vision, 'The Semantic Web' was described as going to bring structure to the meaningful content of the web pages, and thereby sophisticated tasks for users would be performed by software agents and the scattered information problem would be solved. The semantic web was not considered a separate entity, but as an extension of the current web that would bring structure to the meaningful content of web pages, thus, computers become much better able to process and understand the data.

The World Wide Web Consortium (W3C), since its establishment, is trying to improve, extend and standardize the web, and they have developed a number of languages tools and other applications [W3C04]. However, it is also noted that semantic web technologies are still developing and it is early to predict the web's future (Sea09). Clearly, the future of the internet will be different from the present day internet. The information provided on the internet needs to be contextualised in order to be used more reliably by an automated information processing system.

In this thesis, a new framework and architecture for semantic information provision and management for the web is proposed. The information is associated with contextual semantic information which is provided in a machine readable and understandable format. This will allow the information to be self organising and be able to recognize its context and meaning. This is in analogy to the early network structure of a self organising network. In this approach, the semantic information can be used to find and establish links between information on relevant sites. Thus, improving traditional search by placing structured, machine understandable information in objects (web pages linked to objects). This self organising information linkage allows the automatic gathering of information about entities/products/services within a local environment as well as globally on the web. Furthermore, the semantic information can be used to adapt of individual's needs by reorganising the site and providing a specific semantic display of say products and services.

Ultimately, this thesis makes a significant contribution in development of semantic search and discovery approaches which could be used as the basis for a self organising hyper-information system.

However, in order to develop the new framework and architecture for real-time information searching on the internet and the web, it is necessary to define the architecture and technologies. It is also important to investigate technological functions and its various usages. When making use of the existing technologies and functions, the first priority is to establish the adaptations required to develop an efficient single point framework and architecture for real-time information search web sites. Moreover, to be able to develop the framework a number of architectural principles and design concepts need to be considered and the criteria for evaluating and designing websites should be defined. Furthermore the adaptation of the proposed framework and architecture to attain real-time semantic information search websites should be investigated.

Thus, the thesis aims to contribute to this stream of research regarding the use of semantic information searches on the Internet. Moreover, of particular interest to this research is the identification and description of the current position of the web in the business industry (E-commerce) and its search results. As will be discussed in this thesis, the real-time semantic information search methods described in this research refer to any organisations and concerns both the use of semantic technologies and the semantic information search within an organisation. The research is done by using a qualitative and evaluative study on semantic information search methods for the logical and semantic conclusion of the findings and results. The qualitative data is analysed and discussed in order to reach the research results.

1.3 Structure of the Thesis

The thesis is structured in the following manner:

Chapter 2 starts by considering the network infrastructure and communication protocols which are the fundamental blocks on which the internet and the web are built. Then, web languages including the mark-up languages and architectures including the tree tier systems are discussed. The different types of web sites are then considered showing some of the advantages and limitations of some existing sites. This is followed by reviewing E-commerce systems with particular emphasis on their design aspects. Then, a brief discussion on

information search and the issues of information overload is presented. Finally, a summary of the chapter is presented.

Chapter 3 starts with a discussion of the definitions and architectures commonly used to describe semantic web. Then, the knowledge representation languages and their contribution in the designing of a semantic web site and a number of common patterns that these languages follow are also discussed. This is followed by descriptions of semantic web documents and formation. Web services and their potential semantic links are then discussed. Finally, a summary of the chapter is presented.

Chapter 4 starts with a brief discussion of the approaches commonly used for the evaluation of websites which are normally based on one or more aspects of the quality characteristics. Then, the software quality factors which can be used for the evaluation of websites for their semancticity are considered. This is followed by proposing a quality model which is then used for the evaluation of some leading E-commerce websites. Moreover, some specific quality characteristics directly related to the indentified quality factors are considered together with the different types of websites structures and their impacts on quality characteristics. Finally, the chapter's findings are summarised in the conclusion section.

Chapter 5 starts with a discussion of the meaning of the semantic information together with semantic encoding and indexing, which explain and define, using examples, the approaches used in web application techniques particularly in terms of indexing web based objects and information according to the user relationship and the semantic information search. The semantic architecture which can be used for the implementation of the proposed approach is then presented, together with some descriptions of semantic items relationships and processes. This followed by a discussion of a self organising hyper linking system. Finally, the chapter's findings are summarised in the conclusion section.

Chapter 6 starts by presenting a more suitable definition of semanticity, in the context of artificial intelligence on which the new approach has evolved. Then, the intelligent engine in the semantic layer is described together with all the required processes. Neural Networks are then briefly described and the choice of a network based intelligent engine is rationalised. The probabilistic network of the engine together with the ways how the connection weights are calculated are then presented. followed by examples on how the approach may be used in

an E-commerce system for semantic information search and discovery. Finally, the chapter's findings are summarised in the conclusion section.

Chapter 7 presents the integration of the proposed framework and architecture together with the indexing, semantic storage, semantic search and semantic display concepts. It starts with a discussion of E-commerce systems and the requirement for product E-catalogues/lists, then, it carry on with a description of the components of the system, conducts an experimental study with the aim to evaluate the system using a case study based on a retail E-commerce business application, and it presents the results and analysis of the evaluation. The chapter concludes with a summary highlighting the results of the experimental study and showing the significant benefits of the new approach.

Chapter 8 concludes the thesis by specifying the main contributions of the work. It gives a summary of the work done as well as its evaluation results. It also outlines possible research areas that can be carried out in future work and identifies some guidelines and directions of future semantic information research.

Chapter 2 - Internet and Web technologies & Applications

2.1 Introduction

The fast growth of the Internet during the last few years has been one of the dramatic changes in the world. In 1999 Andy Grove (Gro99) was convinced of the idea that in five years time or so, all global companies will be internet companies. In fact, the prophetical statement of Andy Grove (Gro99) was fulfilling in few years time, because the advancement of internet has become a global phenomenon and has paved the way to do business and other transactions regardless of their interests. This advancement has brought new internet technological applications in businesses and communicating knowledge and information.

There have been plenty of attempts to improve the quality of internet and websites and many approaches, architectures, technologies have been developed and deployed in the last few years and still developing and deploying. Some are focused on technologies and infrastructures and others are focused on software tools and architectures. Initially, the new developments were driven by the massive improvement in IT network infrastructures including new communications hardware and protocols. However, this is now combined with the development of new languages such as mark-up languages and new design tools as well as improvement in data storage and retrieval and new software architectures. All these development contributed to the development of the web. Therefore this chapter reviews the recent developments that have led to the popularity of the web.

This chapter starts by considering some of internet communication protocols which are the fundamental blocks on which the internet and the web are built. Then, web languages including the mark-up languages and architectures including the tree tier systems are discussed. The different types of web sites are then considered showing some of the advantages and limitations of some existing sites. This is followed by reviewing E-commerce systems with particular emphasis on their design aspects. Then, a brief discussion on information search and the issues of information overload is presented. Finally, a summary of the chapter is presented.

2.2 Protocols, Web Services and other Techniques

Protocols are necessary methods of communicating within the networks. The computer programs concentrate on communicating with each other in terms of standardised protocols (Ido08). The quality of web networks is stated in terms of protocols, and it is difficult to view the web pages without the assistance of protocols or one would suffer the problem of not viewing the content of the web page properly. It is also difficult for the programmers to communicate with the network and relate with organisations without the proper use of protocols. Therefore, incorrect use of protocols can cause chaos and problems in the web-system. A solution is to use more robotic protocols in terms more understandable to system processing. One such approach is to find ways for describing and implementing new protocols (Ido08). The new protocols have the task of enterprise application integration (EAI) and cross-organisation integration (B2B). Though EAI and B2B are related, they vary in their details. EAI is mainly working on a single administrative domain and it can be removed at any time and replacing it is also possible, whereas B2B is extended to customers and therefore removal can affect the customers. The frequent change of the protocol can create problems within the system. EAI and B2B can be mistakenly understood as similar, however, both of them differ, and therefore they require different technical strategies to deal with them (Ido08).

SOAP (Simple Object Access Protocol) and WSDL (Web Service Description Language) can overcome the problems of B2B and EAI. SOAP exchanges XML based messages over computer networks, and it uses HTTP.SOAP to form the web services stack, and it provides the basic messaging framework for building up more abstract layers. Many types of messaging patterns are in SOAP, with the most popular one being based on the Remote Procedure call pattern (RPC). The function of sending messages is quite dynamic, in which one network node sends a request message to another node and the server quickly responds to the client (Don00, Dan02).

SOAP works on the internet application layer and as a transport protocol; it also tunnels other protocols. SMTP (Simple Mail Transfer protocol) and HTTP (Hyper text transfer protocol) are application layer protocols used as transport for SOAP. HTTP works well on today's internet communication architecture, and SOAP works well on network firewalls. SOAP also is used widely with the HTTPs though it is similar to HTTP; it also uses an encrypted transport protocol underneath (Elf02, Don00).

In order to describe the web services there are description languages such as WSDL (Web Services Description Language), which is an XML based service description that lays down the descriptions on how to communicate with the web services. WSDL describes services as collections of network endpoints. It gives an XML format for documents. Though the collection of ports defines the particular service, the port is only defined by associating with network addresses with a re-usable binding. Each of these messages are abstract descriptions of the data being exchanged and port types are abstract collections of supported operations. Thus WSDL is often used in combination with the SOAP and XML Schema and it can provide web services over the internet. The client programme can connect to the web and see what is in the server by using WSDL and the client can make use of SOAP to call on the function of the service (Edo05).

The critics of SOAP point out that SOAP is, being degraded as a framework. Its goals have changed so drastically that the word SOAP went from being an acronym for Simple Object Access Protocol to being merely a meaningless name. It is becoming a syntax to a framework from an RPC (The Remote Procedure Call) protocol to an abstract, one way messaging environment (Nik04).

2.2.1 HTTP

In order to access and display of information between any two systems over the internet, the hyper text transfer protocol (http) is developed as the standard communication protocol for a distributed collaborative hypermedia information system over the internet (Yve03). The W3 consortium defines that the transaction over the internet consists of four basic phases which are the connection, request, response and close phases (Yve03). In the first phase, the web user establishes a connection with the http server. When the connection is established the communication begins and the user sends his/her request to the web server. The request specifies the protocol, object, and the responding method. The server will respond according to the user's request, otherwise it will send back an error message. As a response to the request, the browser displays the object and the connection is closed.

2.2.2 URL

Uniform Resource Locator (URL) is the addressing system which is used to describe, locate and identify the type of files or servers. In addition to the address of a particular web page, the Uniform Resource Locator consists of the protocol, domain name, path and file name, and

even the information of what protocol to use for handling a file and the location of the resource. The addressing definition is considered as the most important part of the URL, the information in it defines the method for transporting or handling a file over the Internet. After identifying the protocol, the web browser would know which port and server it is talking to for the information, then it directs to the selected hyper link. HTTP, FTP, Telnet, news, etc. can be specified by the URL to access resources from the Internet. The location of the server hosting the file that the hyperlink is directing to (domain name) is identified by the URL and is also included in the function of a URL. For example a meaningful company can be located by looking at its URL, say http://www.shop.co.uk or IP address for this company could be "175.675.75.119" which refers to the same web server address. The URL also specifies the pathname or filename which the hyperlink is pointing to. When specifying the path to a file, it is normally separated by forward slash, which separates each directory.

2.3 Security and Integration Issue

The complexity of E-commerce and E-business compels the protocols to cope with new security issues. IP protocol has many weaknesses, and that enables attackers to break the security of internet communication. IP spoofing is an area where IP fails. The attackers pretend to send data from an IP address other than the proper address of the source server. In return, the destination server will send information back to the addressee. This is due to the lack of encryption and authentication. Packet sniffing (Electronic Eavesdropping) is another issue requiring more vigilance. The packets on the network can be under attack. If the card is put on promiscuous mode, it can collect all the packets that are on its way and this is mainly done where there is an Ethernet-based network. IP is also vulnerable to session-hijacking. In order to overcome these problems IPSec (Internet Protocol Security) was created by Internet Engineering Task Force (IETF) with three IP extensions: authentication header (AH), encapsulated security payload (ESP) internet key exchange (IKE), with two mode operations such as tunnel mode and transport mode (Dim00, Bur02).

It is important to realise, however, that there is no guarantee as to whether other protocols, such as IPsec, provide a better service to the internet, than other IP protocols (Bur02). For this reason, investigations exist to prove, for example, the limitations of IPsec. Many are complex and not completely mature, incompatible with NAT (Network Address Translation), dynamic addressing is not well supported, and there are interoperability issues. There is no

centralised dynamic management system to handle security issues and it lacks functional and recognised PKIs (public key infrastructure) (Chr07).

The traditional method of communicating over the net has gained wide popularity. Despite this, worries and concerns have been raised over the construction of a simplified communication system and the technical and commercial correctness of the approach. The invention of too many protocols in the future might cause confusion in the commercial as well as the telecommunication community, because the communication must be simplified and straightforward so that it can overcome any security, consistency and reliability problems. Usually, in order to tighten the security, new protocols and methods are invented which are inevitably too complicated for the ordinary customer/entrepreneur. For example, involving multiple vendors into a cohesive solution integration application has been a challenging issue, thus integration involves both in-depth application programming, interface knowledge of the systems and peer to peer integration between each system (Jor07).

Point to point communication also does not appear to be satisfying in an expected way. Alternative methods will, for example, use enterprise application integration (EAI) and efforts have been developed to overcome the complexity of merging different systems by creating a common layer to manage these integrations thereby simplifying each component system (Jor07). However, the simplified method of enterprise application integration has also created problems in the industry such as being complex to create, and difficult to maintain and upgrade (Jor07).

Although the intention was for an integration system technology independent of the traditional method, concerns have been raised that such independence would not be possible. For example, if new communication and information technology is introduced, then the communication, data transfer, capturing etc. cannot be possible unless a new set of protocols and layers is invented or created in line with the new technology. Moreover, if new technology permits the client to do new ways of communication and commerce, or to do things differently, then that communication mode is different and presumably it uses new technology. As a result, its protocols must also change. Thus, if an entrepreneur uses a new technology, and wants to continue with the E-business, then it must produce a new set of protocols and layers for communication (Rom08).

Therefore, many organisations and industries are adopting new approaches to integrate business systems such as service oriented architectures (SOA) (Cla07). SOA is not a product or technology; it is a technique for designing large, enterprise applications. Service oriented architectures are dependent on a number of open standards: Web services, SOAP, XML and HTTP (Tho06). These will help the software developers view their applications as services and to call other systems (Cla07). Moreover, Uncle Bob (Unc07) pointed out that SOA is supposed to confiscate the core business functions and its practices into independent services without frequent changes and they are glorified functions. These services are performed through presentation programs. The presentation programs are meant to present and accept data from various users. Furthermore, Rich Seeley (SEE06) noted that the Digital Business Architecture emerges from SOA. There are concerns about the usefulness of SOA, other web services and techniques. In one enquiry, for example, it was noted that SOA has two limitations in Synthetic SOA transactions; complex transactions are hard to imitate and problems tend to get hidden in the background (Ora09).

In another investigation, it was cited that one cannot distribute all functions in an application, and therefore there is doubt that SOA is not able to distribute all functions by itself and it can limit the service orientation (Rme07). These findings are confirmed by the use of the SOA architecture in a business environment, in which effort has been taken to improve SOA so that it is close to some of the technologies generated by the existing SOA usage. It is important to realise, however, that there is confusion as to whether other architectures and techniques provide a better business service than the SOA architecture. For this reason, technologies exist, for example, which can enable E-business in a similar way to SOA, and it was concluded that better results are found by moving to other techniques which are close to SOA (And08). Even today, investigations are on the way to find the consistency, reliability and security of SOA, other web services and other techniques (Hel07).

As a result of such issues, new technologies are emerging in E-business environment and new E-business architectures are proposed, for the purposes and improvements of reliability, consistency and security. The new approach is based on the conclusions that a single technology based E-business solution would not overcome all the existing limitations of E-business and E-commerce (Hel07). Different types of web services or collaboration of different technologies are therefore required (Hel07).

2.4 Web Languages and Architectures

2.4.1 Mark-up Languages

A Mark-up language is able to structure text and multimedia documents and includes hypertext links, then connects between documents for use on the World Wide Web. In the late 1960's the idea of developing a structured mark-up language was necessary for various aspects (Gol90, Oxf02) for example the need to describe the contents of a document which include the heading, separating the logical elements of the document, specifying the processing functions which are to be performed on the same elements (Jae03). A Mark-up language is not simply used for word processing but for data structuring both texts and multimedia. As the web and Internet developed, mark-up languages have also developed, because a single mark-up cannot overcome all the needs required for human communication. But, mark-up languages set a common structure for processing, executing, displaying and presenting the content of a document with many benefits, for example it is easier to learn similar languages. The most commonly used mark-up languages are explained in the following sections.

2.4.1.1 HTML

HTML (Hypertext Mark-up Language) was developed by CERN (European Organisation for Nuclear Research) (Yon02). Since SGML (Standard Generalized Markup Language) is a complex language (Oxf02) and it was developed before the widespread use of the Internet, it was deemed too complex to develop documents for the web; therefore HTML was developed mainly to become an application profile for SGML. HTML was developed as a simple language, which defines the content of a page, headings, lines, paragraphs, tables etc. It is able to display information on both hypertext and multimedia, and this information can be exchanged between computers in global interoperability. HTML can present a document structure and it is capable of visual presentation (Rob06). The hypertext technology makes the web more efficient. It is a presentation format that allows clients to view information and it can be linked to other documents as hyperlinks. Hypertext documents are created using HTML which also allows the user to navigate from one section to another. Table 2.1 shows an example of HTML tags used in a web page.

```
<HTML>
<HEAD>
<TITLE></TITLE>
</HEAD>
<BODY>
<H3>homepage</H3>
<P></P>
</BODY>
</HTML>
```

Table 2.1 HTML tags used in a web page

2.4.1.2 XML

The Extensible Mark-up Language (XML) is a meta language as well as a mark-up language, which contains a set of rules for creating document formatting languages for different applications such as application oriented description languages, for example WML (Wireless Mark-up Language) for WAP and GML (Geography Mark-up Language) for dealing with geographical data such as geometry and attributes. Therefore this is a language setting a standard for the exchange of data over networks. As a Meta language it is able to describe another language, and as a Mark-up language it is also capable of describing information for example storing, processing and transmitting information according to the types of applications (Gir00). SGML is the parent language for both HTML and XML, all these languages use tags to explain, present and interpret the contents of a document as shown in Table 2.2. XML is used with HTML, and they complement each other, however XML can deal with some of the features which HTML cannot deal with.

```
<?xml version="1.0"?>
<!-- A simple document -->
< business>
<title>XML, Servlets &amp;amp; JavaServer Pages</title>
<company business="grocery">
<name>shop</name>
```

```
<store>London</store>
<address>14 High Street</address>
</company business>
<day>05</day>
<item>orange</item>
<detail>Orange is imported from South America, and sold in London.
</detail>
</business>
```

Table 2.2 XML Example

2.4.2 Three Tier System

A three tier application separates a networked application into three logical areas such as a client user interface tier, a server-side information process tier often known as the middle tier and the remote tier with a database management system (figure 2.2). The client-user tier manages presentation. In a web application, data is requested by the browser and is sent there from the web server to display. Tier two is responsible for business logic processes. Tier three process the data, from where the data is stored. In a web application tier three is responsible for the back-end database management system (DBMS). Three tier systems are easier to deal with, modify and maintain, because of the separation of presentation, business logic and data processing.

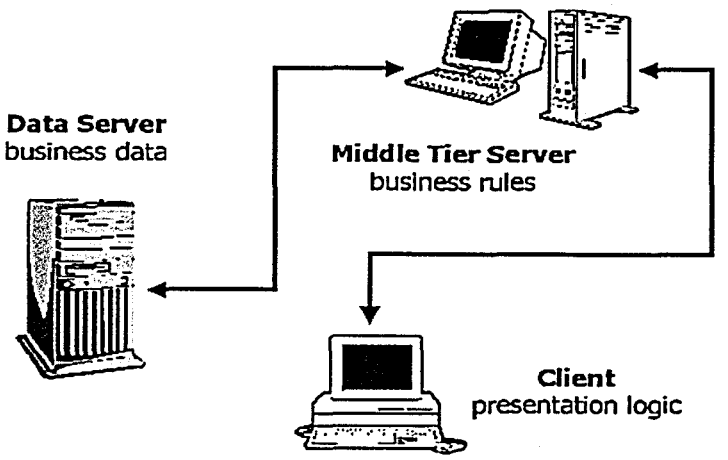


Figure 2.1 A three-tier application (Ram00).

In the three tier model, the presentation layer can be the browser which can interact with the user by HTML. In the web page there will be a provision for the user to enter information and submit the information to the server by making use of web page scripts. The values of the scripts on a web page can be passed to the server for processing. All submissions and requests are passed on to the server through the browser. The web server is the middle tier and the application layer. The web server generates HTML documents and sends them to the user. To access the database, the web server contacts the data layer through a server side script.

2.4.2.1 JAVA Servlets

Java Servlets are JAVA based server-side software components and servlets are protocols and platform independent. JAVA servlets are compiled to platform neutral byte code that would be loaded dynamically into the Java enabled web server. This is independent and therefore not specifically attached to the client server protocol. However, Java Servlets are used with HTTP. Java enabled servers will load the servlets and displays on Java virtual machines. Java servlets do not make use of any graphical user interface, because they run on servers. Servlets responds to HTTP requests from the user and reacts according to the users' need. In most of the graphical user interface programs, Java servlets and applets are used, servlets run within the server and are analogous to applets that run within a browser. The user sends requests from client applets and carries the requests to the appropriate DBMS.

2.4.2.2 PHP

PHP is a scripting language mostly used with server side applications and automated dynamic web pages. It is easy to use with HTML. In comparison with other scripting languages PHP is unique in several ways for example the code is generated from the server which makes data processing fast and the user gets a response efficiently and quickly. The webpage 'us.php.net' explains the historical development of PHP. PHP/FI, which means Personal Home Page or Forms Interpreter, was developed by Rasmus Lerdorf in 1995 and named as personal home page tools (Ben03, sha02) initially it was only a set of Perl scripts for tracking accesses to Rasmus online resume. Later Rasmus wrote a C implementation to communicate with databases, and the users are able to develop simple dynamic web applications. In 1997 PHP/FI was officially released as PHP/FI 2.0 and several thousand users have installed this version of PHP/FI. Though several people contributed to the code, it has largely remained as a one man project. PHP 3.0 was the first version that is very similar to the PHP that we have today. It was developed by AndI Gutmans and Zeev Suraski in 1997 (Ben03). One of the common strengths of PHP 3.0 was its extensibility features; it also provides the end users with infrastructure for lots of different databases and protocols. It was renamed as PHP-Hypertext preprocessor. PHP 4 was released in 1998, and later in 2004 PHP 5 was released with new features. PHP code can be easily parsed at the server into html and then sent back to the browser which will understand the html tags and it will display the webpage. The three main applications of PHP are server-side scripting, command line scripting and writing desktop applications (Meh10).

2.4.2.3 Database

Database can be defined as a collection of data or information organised in such a way for ease of use and speed of management, search and retrieval. This is also known as the data bank. A number of database systems exist for example biographic databases, which provide a record of an item and information about an item, for example author, title, subject, organisation, full-text databases which give the full-text of a publication, sometimes an entire book is stored and its citation is given, image database which collects and stores images, for example CNN collects images, audio information, MP3 files and the combination of the above, meta-databases are there to provide the user with searching facilities for data, the user can search for content that is indexed by other databases (Hun98, Jem10, Joh10).

The Database Management System (DBMS) is one or more software programs which supply functionality to users for developing, using and maintaining a database. DBMS can define, populate, create and manipulate data or databases for different types of functions and applications. DBMS make use of structured query language (SQL) to access or retrieve information from the database. DBMS is well advanced and it has got many advantages such as data abstraction, data independence, transaction model, handling data on several transactions (concurrency control), and robustness for handling crash recovery. There can be many data connectivity applications, for example ODBC (open database connectivity), a database access method developed by Microsoft corporation. ODBC can access any data from any application, regardless of its database management system. ODBC makes use of a middle layer database driver to do this function, between the application and the DBMS. The purpose of this layer is to translate the application's data queries into commands that the DBMS understands. On the other hand Java Database Connectivity (JDBC) works with Java programs and it is a Java API that can execute SQL statements. JDBC makes it possible to write a single database application that can run on different platforms and interact with different DBMSs. The next sections will explain how these internet technologies are used in present day information communication web services, web applications, information search and its presentation over the net.

2.5 Types of Websites

Today's society depends increasingly on the internet to access and retrieve information, millions of people "surf" the Web every day. Therefore the websites have to be created and maintained to meet the expectations of people. This expectation of people can be maintained only by securing a proper communication service.

The main concerns of the business leaders and organisations are, therefore, to build a good information communication service over the Internet. The information communication service can be established by building a good website and thereby improving the efficiency of electronic business. The information communication service ensures the competent nature of an organisation, and the efficient information communication service assures more revenue and thus makes an organisation competitive in today's online market.

A well designed system should have uninterrupted information thus making the service compatible and efficient. Most of the problems connected with web sites are related to

insufficient information or not being able to find the right information. Difficulties in finding the material or service, poor site layout, incorrect content etc. are examples of the problems generated by the inefficient information communication service.

The popular medium of conducting business on the net adapted by industry can be seen as the World Wide Web which is a hypermedia-based-system (explained in the previous section) for browsing Internet sites. The World Wide Web (WWW) is the hypertext-based graphical interface to worldwide resources. It uses the internet as its communication network. The most commonly used service of the internet is the electronic mail service. Behind the rapid growth of internet use, the web plays an important part.

There are a number of ways to describe electronic business. The term "E-business" was coined by Lou Gerstner, CEO of IBM (Rim09). According to Queens-land government (Que05) it can be defined as the conducting of business on the internet. It not only includes buying and selling but also servicing customers and collaborating with business partners. It can also be defined as any business process that is empowered by an information system (Rim09). Today, this is mostly done on the basis of technologies. Therefore, electronic business can be defined as the efficient and dynamic selling and buying of goods online, servicing customers and collaborating with business partners on the Internet. However, there are different types of existing websites with different approaches to E-business which discussed in the following sections.

2.5.1 Information Website

The brochure or the information websites are only accessed by the surfer. As the name signifies it can only be viewed. The information cannot be downloaded, however can be saved on a per page basis. This type of website is the simplest, and is widely used over the net (Tar05, Rob07). The brochure web sites usually contain information about the business, its products or services. There are many features of brochure websites as they are a kind of advertisement and provide an assurance on the net that the company is active and it is doing business over the net. Similarly, brochure websites expose the company to the target audience. Therefore, a well designed brochure website can increase traffic to the business and increase the sales potential. Thus the brochure website can provide a good image of the business.

However, these sites have their limitations in supplying real time information. According to Chris Ramsey (Chr05) information websites severely restrict their functionality, expansion opportunities and the ease with which they could be modified. It is not easy to update the website code.

An information website can sometimes be misleading when it is attached to a search engine and its consequences may be severe especially on health related subjects. Health related information is widely accessed on the internet with millions of people regularly access the internet for health care information and there are massive number of sites that offer health related information. These clearly require methods and tools to improve the quality of information on the websites. Tools include codes of conduct, quality labels, user guides, filters, and third party certification. These tools are necessary for the developers to filter content and help consumers become discerning users of information (Pet02). The technology mainly used by the information website to make the information accurate is a filter. It can be applied manually or automatically to accept or reject the information based on controlled pre-set criteria words. By using these tools online users can organise their access to the internet according to quality and relevance of information related to a particular audience. The data is reviewed, classified and the descriptions are stored in a database (Pet02). Criticisms of relying on the reputation of such an information website will include lack of technology, quality over information (Med01).

2.5.2 Interactive Websites

Interactive websites provide standards-based cross curricular web resources designed to enhance online learning opportunities. These sites interact with the user usually through either a text-based or graphical user interface (Jef08, Moh01). In an interactive website the user can browse and download information. Even software graphics and other multimedia information can be requested and downloaded. Chat rooms and message boards are other forms of interactive sites. Interaction can happen in many ways. Video and audio interactions are mostly used in chat rooms and message boards. Fill-in-blank interactions are often used in education and learning. The customers are often asked to pick up a topic and fill in the blank. The website guides the customers through a good number of websites. The web resources are used for activities and that is in the learner's interest.

Another way of interacting is through visualisation tools, in which people can communicate in a workplace environment (Jos04). A number of organisations that have attempted to use such tools and techniques, however, have found them to be unsatisfactory or individually underperforming, since each technique has different strengths, capabilities and weaknesses (Ing05, Joh06). As a result, some information website', interactive website' and graphical website approaches attempted to incorporate the best features of the available programming languages/tools, such as flash and java applets (Mar04). One example of such a technique is visualisation which incorporates a number of different methods and programmes relating to visualisation in interaction, information and data (Mar04). For example, Newsmap voyage and Diggybigspy are trying to visualise and map the information.

Newsmap allows the website to reflect the continuously changing landscape of the google news-news aggregator. The traditional treemap space-constrained visualisations of information are used by the newsmap tool to divide information into quickly recognisable bands which helps the visualisation algorithm to permit the display of large amount of data gathered by the aggregator. This can provide chances to reveal underlying patterns in news reporting across different environments and cultures. Newsmap doesn't replace the googlenews aggregator, however it demonstrates visually the relationships between data and the unseen patterns in news media and it accentuates the bias. It won't display an unbiased view of the news (Mar04).

Voyage is an RSS (Really Simple Syndication) which is a family of web feed formats used to publish frequently updated work, such as blog entries, news headlines, audio, and video in a standardized format. An RSS document is called a "feed", "web feed" or channel (Dan99). If an RSS feed added to a user's format sites, voyage will lead to all the information, this capability is already a built in feature of flash which is capable of providing a 3D effect for news content. One of the useful functions is that RSS can provide a view option which lists all the items for a feed to the subscriber. This can build up into a library of information, helpful for research or blog post. It can also update the news because of the RSS feed.

Diggbigspy is another technique for arranging popular news stories at the top when people dig them, and the bigger stories have more digs (Vit07). Although such approaches have been influential and encouraging, discontent with visualisation techniques in general is leading to a re-examining of their performances. Programming firms and individuals require certainties, but changing information and therefore data capture techniques provide less accuracy with

real-time semantic information. Well organised programming depends on the availability of real time semantic information, well defined data sources, and consistent information accuracy tools. However, the programming cannot catch up with the real time semantic information because the information is always changing and it is dynamic while programming can be static. As a result, by the time the programme finishes, the information has also changed, and a gap has appeared between the tools and methods used in the programming techniques and the real-time semantic information.

There has also been a tendency to come up with complex programming techniques with little novelty in real time information, with the result that the models and tools are less likely to be applicable to the real time semantic information. Examining these websites shows that without accuracy and consistency to the global as well as the local businesses, tools and methods may fail to come up with real time semantic information (Geo08, Ste06 Mel02).

2.5.3 Virtual Websites

One of the designing techniques on the web is to create a virtual reality. According to (Mar05) virtual webs are a building boom on the web that is still continuing. It is the technique of providing the user with the feeling of real cities, towns or shops on the web. This is constructed with digital bricks and mortar. Many thousands of people are using these virtual cities to carry out everyday activities like shopping, relaxing and meeting friends. Virtual cities are widely used on the web. The virtual designers divide virtual webs into three categories or groups.

a) Web Listing Virtual Cities are web sites which describe themselves as virtual cities, but in reality are merely on-line guides, menus and listings. They are often created solely for advertising purposes, particularly for tourism promotion, and make no attempt to represent the built form of cities. Many examples are available, a typical example being "Virtual Brighton & Hove" (Bri04).

b) "Flat" Virtual Cities use "flat" maps of cities or buildings as an interface to further information. A good example is Virtual Bologna which is making use of stylised town map with familiar landmarks and buildings as a graphical interface making it more realistic to detailed online information (Mar98).

c) 3D Virtual Cities are another concept which uses virtual reality with 3D technologies to model and construct the cities, towns and villages to varying degrees of accuracy and realism (Zsz09, Mar98, Qvo02). The virtual cities provide the user with more real time experience which can enhance the user shopping experience. The digital cities also give the user a genuine sense of living in an urban environment. To achieve this task the designers must make use of all the tools and technologies with realistic user interactive interface (Mar05).

2.5.4 Graphical Real Time Websites

In order to make the web more accessible to all categories of people the designers use more graphics and images. The web can become a better looking place, only when the designers make use of the interaction techniques, tools and other media formats (Pau05). These websites are very common on the net and almost all organisations and business firms make use of these design techniques to make the site realistic. Most businesses today depend on these three categories of website designing models. It is clear therefore, that there is a need for a new, improved approach to E-business and websites that meet all the necessary factors identified.

2.6 E-commerce Definitions and the Designing of a Site

Electronic commerce (E-commerce) simply means doing business over the Internet. It is an action which is performed beyond geographical limitations and is not bound by time; it interchanges data between the customers and owners. E-commerce is used to make business efficient and quick. In most, if not all, E-businesses a special subset of terms can be identified. It can be named basic business models. According to Michael Rappa (Mic07) these business models are mostly discussed but least understood. Theophilus et.al (The03) listed several categories of commercial activity on the internet. These models include B2B which is the commercial activity between two or more business organisations. B2C involves businesses and consumers. C2C simply means Consumer-to-Consumer business activity. C2B can be understood as Consumer to Business e-commerce.

The prime element in E-commerce is the website and an important aspect of delivering goods/ services to customer needs is the ability to interact. Good E-commerce sites can be designed and maintained by understanding customer needs. A good E-commerce site also forms an essential basis for effective E-business, and it can be defined as:

E-commerce = Input through internet/website decides output per unit interaction.

In the case of website designing, input can be readily identified but the output is not quite certain. For a website, suitable input may be decided by its business. Output cannot be decided unless it goes online. Websites differ from one to another on account of their business, however, the functionality and objects of an E-commerce site cannot differ. For example an E-commerce system must have objects such as a shopping cart, login/register, help, account/order, search, link back to home page and links to merchandise. These are the functions of inputs which enable the site to run. Websites' content may differ and each website system is developed according to its business. Content may differ, however, the objects are the same. As a result, it is possible to add the contents of the site later, and therefore contents of the site can be left out until the end of the site designing or at the updating phase. Therefore, it is very difficult to estimate the output of the site. The output of the site can be determined through the interaction by the customer or the users. Website designing, therefore, takes into consideration the content and the objects of the site.

2.6.1 Design Aspects

E-commerce site design may also be defined as the interface for browsing products at different levels, including local and global navigation. Site navigation is given on account of the adopted business processes. The business processes of the site are the decisive factor in selecting the interface for browsing products, for example, the hierarchical navigation, linear navigation, grid navigation, modular navigation, webbed navigation and mixed navigation. Website designing is a decisive element in representing the business for a number of reasons. For example, it can indicate good or poor business and can be used to evaluate the effectiveness of business processes. Since the designing of a site is an important factor in determining the quality of a business, it is therefore essential to choose the interface for browsing products according to the business model. E-commerce site designing also forms an essential and important baseline for best E-commerce. Indeed there is a large variation in E-commerce sites in terms of design, between the best, average and worst.

Moreover, good E-commerce sites should also be concerned with the design of a good shopping cart. For the E-commerce system, there are many traditional business factors that must be incorporated into the shopping cart, and these include easily finding the products/services, placing/adding the products into the basket and checking out. After

identifying the traditional business factors, the shopping cart design itself is often the most difficult part in the E-commerce system. As a result, there are many errors that can occur when designing a shopping cart, including naming the cart, finding the item, adding the item, viewing the shopping cart, visual feed back of the shopping cart's content, the option to replace the items, checking-out desk with registering option, creating an account, shipping address, billing address and reviewing the entire contents of the shopping cart. To avoid errors when designing the shopping cart, one can either develop a site that considers all these traditional business factors, placing and naming them accurately or make use of shopping cart software in estimating the feed-back from the customers. It is important to note that a good shopping cart system should be designed and used to make the E-commerce system function properly and to run the E-business smoothly.

The E-commerce system design must take into consideration different types of E-shopping cart, for example, the intelligence of the shopping cart, its ease to maintain and modify as well as the cost. The shopping cart cost estimation is only a proposal and it can fluctuate at any time once the system is kept running (Bar01). The list of requirements for the E-commerce determines the cost of the shopping cart. For example, features, hardware, software, consultants and licensing fees. Therefore, before designing an E-commerce shopping cart it is important to compare between the best E-shopping cart, such as ASP (active server pages), CGI (cookie shopping cart), java shopping cart and java script shopping cart. The designer has to make comparisons for and between shopping carts, which can be split into best-effective shopping carts (which are fast and easy to set-up and maintain, easy to customise, supported by major browsers, visual interface, supporting large and small E-commerce) and non-effective shopping carts (which are hard to set-up, hard to maintain, slow, waste bandwidth and system resources, browsers that do not support cookies, and are hard to customise (Bar01). The most effective shopping carts are used to design smaller and larger E-commerce systems with a specific scenario, and are also useful for organisations to adopt those shopping carts with easy maintenance and cost effective criteria. The ASP shopping cart can be considered as the most popular and widely used. The ASP shopping cart itself cannot satisfy all the features of most effective shopping carts. The shopping carts are used to design three classes of E-commerce sites: large, medium and small. Since the cost of maintenance and the ability to deal with the customer/s varies depending on the class of the E-commerce, the shopping carts should be chosen accordingly.

Other issues concerned with the designing of E-commerce sites are transactions that simplify the purchase, negotiation, and post-purchase experience (Paw01). These transactions reflect the E-commerce ability of simplifying the selling and buying of items. During the designing process of the site the products/services should be well explained and the links are clearly marked between the parent and child. Designers may be compelled to follow a pattern of designing structure to allow simplicity of selling and buying of products/services. While designing the site the designers make use of a designing model to overcome the transaction problems. One of the most popular approaches is hierarchical functioning which relates the customer to the site content, thus the navigation buttons within the site make the transaction easier. A typical example is shown in table 2.3.

Home-Return-View Cart-Account-Help-Search

Table 2.3 examples of navigation buttons

The designing of an E-commerce site is concerned with the navigation bars to locate the products/services, forward/backward navigation and hierarchical navigation to make the transaction easier.

The design approach can be used to influence the customer which requires certainties in selling and buying the products/services. The best E-commerce sites depend on the ease of locating of products/services, the availability of real time information, and on well defined specifications and consistency of products and services. However, in certain cases inconsistency and non-accuracy appears between what real-time information can provide, and what the E-commerce site requires.

Another issue in the designing of an E-commerce site is the interactivity which allows real-time interaction with sales and service representatives creating online communities of product users (Paw01). However, there has been a tendency to design increasingly complex E-commerce sites with little interactivity for example with little interaction with sales, service representatives and supply channels. Research in this area has shown that without interaction with the sales and service representatives, E-commerce sites may bring less revenue to the organisation (Kay07). There is evidence, too, that smaller businesses with more interactive e-commerce sites may produce revenue that far exceeds their competitors (Dav07). Designing interactive E-commerce sites may also use different business approaches such as B2B, C2C and B2C.

Thus, to address the design issues of E-commerce sites a number of technological improvements have been suggested to enhance the existing websites such as interactive catalogues, content management without IT assistance, 3-D modelling (Ant08,Deb06). Moreover, a number of interactive features can also be used which include advanced compression for fast streaming, live chat, plug-and-play enhancement etc. (Deb06). In facts statistics show that interactive features keep visitors on a site longer and convert more visitors into buyers (Deb06).

The designing of an E-commerce site is mainly dependent on the nature of the business, and therefore the organisation or business should know what sorts of E-commerce site they require, what features they need in their site, what interaction they need, and what information they need to communicate to ensure that the site is used effectively and efficiently.

2.7 Search Engines and Information Search

The web activities groups, services groups, technologies and the semantic web groups are working toward the development of a web (Semantic Web) of distributed machine understandable data and is trying to improve the quality of web searching (Guh05). Indeed using the current common web technologies it is still not easy to define and link information in a way that it could be used for more effective discovery, finding of information, information searching, information cooperation and information exchange. Therefore it is believed that the semantic web will be able to solve the problem by creating a web which will contain not only media objects, (web pages, images, audio clips) but also information on things and objects which are properly linked (Guh05). However due to its complexity, there are various problems for finding information on the current web, such as publishing information on various issues. In a world where anyone can publish anything, a lot of what gets published cannot be trusted (Guh05).

Moreover, due to the present nature of the web there can be large problems with web searching. For example the anxiety of the information searcher could increase as a result of the heterogeneity and the large amount of information available in the web; this can lead to various other problems such as web database selection problems, (which search engine to select and the limitations of the search engines) information overload (how to select the large amount of data retrieved by the search engine) and query coordination (how to search for

information using keywords and other factors for searching) (Ric05). The large amount of information over the net makes difficult to look for information, in fact a great deal of efforts is involved in finding and locating accurate information where selecting the results is complicated and time consuming, and the output can be misleading. Figure 2.3 shows the current nature of search results and terminating page.

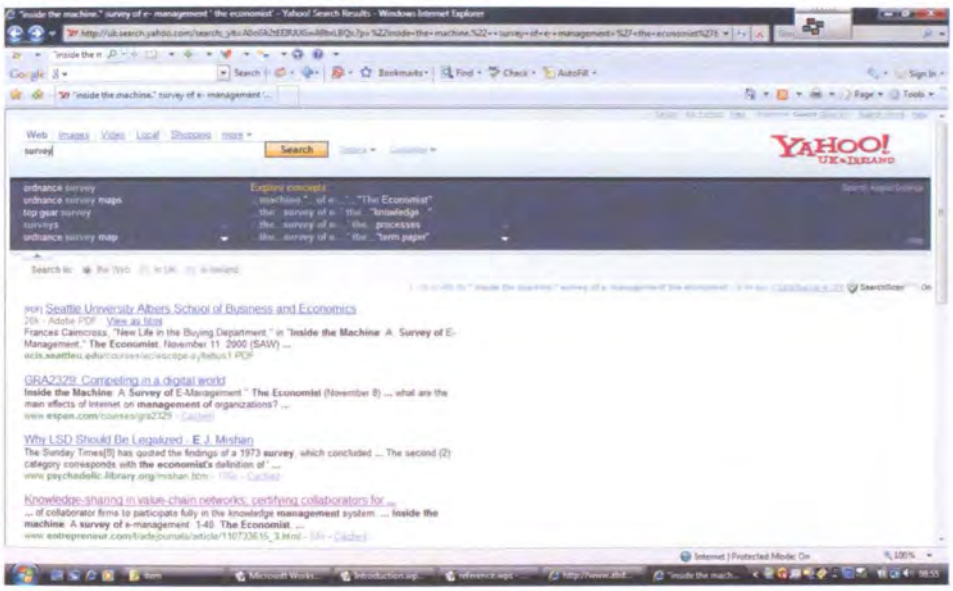


Figure 2.2 Search results

Figure 2.3 shows the search results the search engine produced from the query. When selecting on the fourth link of this page, it leads the seeker to the information page shown in figure 2.4.

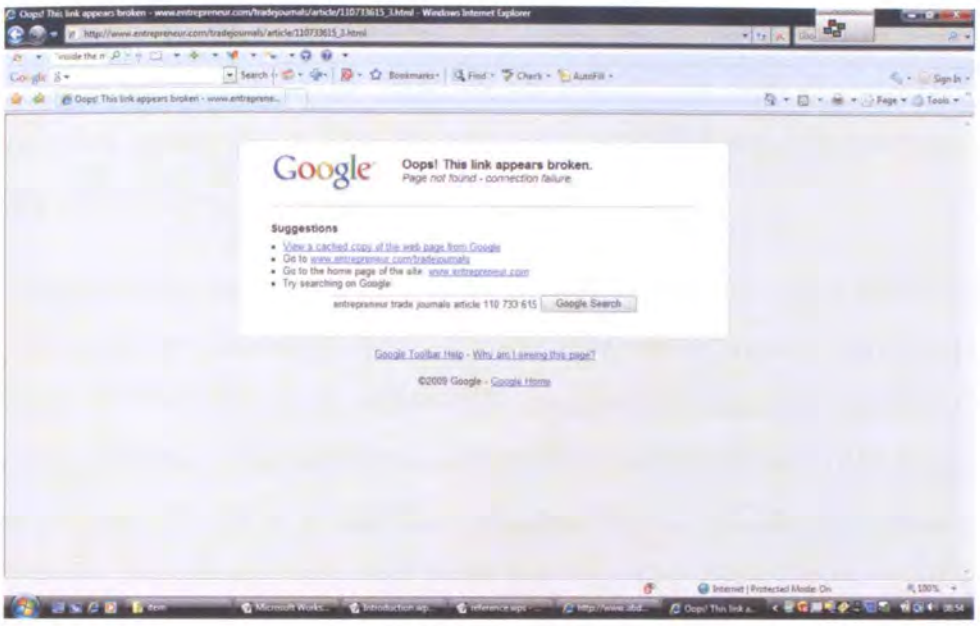


Figure 2.3 Information page

The search engine redirects the information seeker to the google search engine, and the result of the fourth linking page is "Oops! This link appears broken". It is interesting to notice that the search engine results produced many information links, however in many cases, when the seeker followed the link it produced no results and the seeker is disappointed to see that the information does no longer existing. Moreover, many of the existing internet technologies exhibit similar patterns in both types of relationships, furthermore, there might be a strong and evident contrast between the information search and organisation's or individual's real-time information. If real-time information is not found, then such a search finding would suggest that the search engine result could not be real-time and the information is not true to the knowledge of the organisation. If the search engine produces certain results and the information seeker follows the link and finds nothing, then it suggests that the linked pages formed by the search engine are no longer working and therefore there is no real-time information.

2.8 Conclusion

In order to meet the main objectives and the general concepts of this thesis, particularly in term of searching for information and attaining accurate search results in the information world, it is necessary to review some of the technologies the internet and web are based on. Thus, this chapter has provided a background overview of internet and website technologies and their applications with a special focus on E-commerce. This includes reviewing the

internet technologies network structure and protocols, mark-up languages, the three tier systems, databases, the application layer and the information communication services as they are all relevant to this thesis. The chapter also considered the different types of websites identifying their purpose, advantages and limitation.

A particular attention was paid to E-commerce systems in term of their design aspects. In this respect the chapter described how E-commerce sites require careful planning and development if they are to be achievable in terms of semantic, real-time information requirements. The designing and development of the site require good planning in order to be semantic, to give the effect of real-time semantic items, and the modifications to the site should enhance the semantic structure of the E-commerce. There are plenty of approaches to make the E-commerce site more accurate and real-time, including the virtual E-commerce site and the visual shopping cart models. Improving the quality of the E-commerce system includes applying various methods of visual presentation of items and zooming techniques. Altogether, these approaches have never adequately enhanced the E-commerce semantic structures and no one structure can be considered to be better than its counterpart.

Moreover, the way organisations/businesses use webpages to communicate with their customers and other clients and to provide information is very important with information search playing a crucial role in the whole process. Thus, this chapter identified some of the issues in web search, information overload etc., briefly reviewed the existent methodologies that are being used in information searches and has introduced the concept of semantic information search that will be employed in this research. Although, the semantic web and semantic concepts are becoming more popular; they have captured the attention of many industrialists as well as researchers, however many semantic technological issues as well as application methods still have to be resolved.

Chapter 3 - Semantics and Knowledge Representation Languages

3.1 Introduction

According to the Encyclopaedia Britannica (Bri09) the word semantic comes from the Greek 'σημαντικός' significantly originating from sema which means sign. It can be defined as the study of 'meaning.' Semantics is more evident in the science of linguistics, where the study of meaning is based on the historical and psychological significance of words and terms. The word "semantic" is one of those words people sort of understand, but not completely. Like "humility," it is a word that is hard to pin down. Most people do not really know what it means, which is ironic because it means meaning (meaning of, or relating to meaning in language). In the view of Timothy Falconer (Tim03) in computer science, 'mark-up' languages are considered as a type of linguistics. It helps to mark up documents semantically which, in mark-up semantics, means that it is concerned with the meaning of an element, and how that element describes its content (McI01). In contemporary web design, the mark-up languages help to give a meaning of some sorts to an element. However, the internet and the web researchers and technologists might have felt the need for meaningful information over the net, incorporating more semantic information with information searching. Though Bernard Lee (Ber01) called for the re-construction of the web into semantic web (SW) there was no common agreement on the definitions of semantic web, probably because of the ambiguous nature of the word 'semantic' which means 'meaning of meaning'.

Therefore, this chapter will concentrate on explaining the reality of semantic web and all its elements as well at its architectures. It also describes a number of knowledge representation languages and technologies that may be used in the context of semantic websites. However, there are far too many languages, groups and technologies to be discussed in a single chapter, therefore only some of them, which are the most active and commonly used with semantic website development, are considered. Furthermore, semantic web formation, semantic documents as well as web services that are linked semantically are briefly discussed.

This chapter starts with a discussion of the definitions and architectures commonly used to describe semantic web. Then, the knowledge representation languages and their contribution in the designing of a semantic web site and a number of common patterns that these languages follow are also discussed. This is followed by descriptions of semantic web documents and formation. Web services and their potential semantic links are then discussed. Finally, a summary of the chapter is presented.

3.2 The Terms 'Semantic' and Semantic Web

The definitions of semantic web and its developments are undergoing continuous evolution. The current popular definition refers to semantic web as a web of machine understandable information, in some ways like a global database. Bernard Lee describes the semantic web as the extension of the present day web which implies that the future web would be machine understandable information over the net (Ber01). Probably this could be the reason he called it the 'semantic web.' Furthermore, according to Todd et al. (Tod05) commonly used XML tags and schemas have proven to be inadequate solutions, because the meaning is only for humans, and machines would not need any meaning. Solving the alignment problem is the first step to semantic interoperability.

Moreover, in the opinion of Bernardo Cuenca Grau (Cue04) the semantic web is the extension of the World Wide Web and it would be able to use data and semantic definitions which are processed by computer programs. He also argued that the future web would combine existing web technologies with knowledge representation formalisms for infrastructure, so that data could be processed, discovered and filtered effectively (Ber05). On the other hand, Stefano (ste00) defines the semantic web as a network of knowledge other than today's web; the future web is able to define meaning of words and therefore can be defined as a network of information (Ste00).

In 2006, Bernard Lee (Ber06) gave a new definition of semantic web which is based on 1) data interoperability for applications and organizations, 2) a set of interoperable standards for knowledge exchange, 3) an architecture for interconnected communities and vocabularies. Bernard Lee even stated the importance of Artificial intelligence (AI) for the semantic web, in his view "SW is not AI and AI is not SW; AI is a field, SW is a project, SW needs AI and SW owes to AI. SW used much from AI and SW must and should be a great playground for AI and AI projects should use SW to interoperate" (Ber06).

In the above definitions one should get an idea that the semantic web is evolving and developing in various ways, however it is also possible to miss the original vision of the semantic web. Interestingly Bernard Lee himself was not using 'web' in his recent definitions any more (Ber06). Instead he used the phrases, machines, applications on organisations (IT), architecture for communities and vocabularies that means he is concentrating more on concepts, applications and architectures.

However, all the definitions are concentrating on meanings and languages. For the necessary purpose of this research, we consider the following semantic definitions: "Semantic Web is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale" (Sea09). This is one of the most efficient ways of representing data on the World Wide Web, or as a globally linked database. The rationale behind the semantic web is that it needs to find a solution to use the data on a large scale, because at the moment there is no global system for publishing data in such a way that it can be easily processed by anyone. For data processed in HTML, and at times HTML hides the data, and therefore in some contexts, it is difficult to use this data in the ways that one might want to do so, for example, major football statistics, train times, coach arrivals, local community events etc (Sea09).

In the Semantic Web 'data' itself becomes part of the Web and is able to be processed independently of the application, platform, or domain. This is in contrast to the World Wide Web as we know it today, which contains virtually boundless information in the form of documents. We can use computers to search for these documents, but they still have to be read and interpreted by humans before any useful information can be extrapolated. Computers can present us with information but can't understand the information well enough to display the data that is most relevant in a given circumstance. The Semantic Web, on the other hand, is about having data as well as documents on the Web so that machines can process, transform, assemble, and even act on the data in useful ways (Alt08).

The Semantic Web is generally built on syntaxes which use URLs to represent data, usually in triple based structures i.e. many triples of URL data that can be held in databases, or interchanged on the World Wide Web using a set of particular syntaxes developed especially for the task. These syntaxes are called "Resource Description Framework" syntaxes (Sea09). The main components of the semantic web are:

- Unicode and XML
- RDF and other Basic Assertion Languages
- Schema Langauges
- Conversion Language
- The Logical Layer
- A Proof Language
- Evolution Rules Language
- Query Languages for Proof Validation (Aar02).

The semantic web architecture is shown in figure 3.1 which explains semantic web layers.

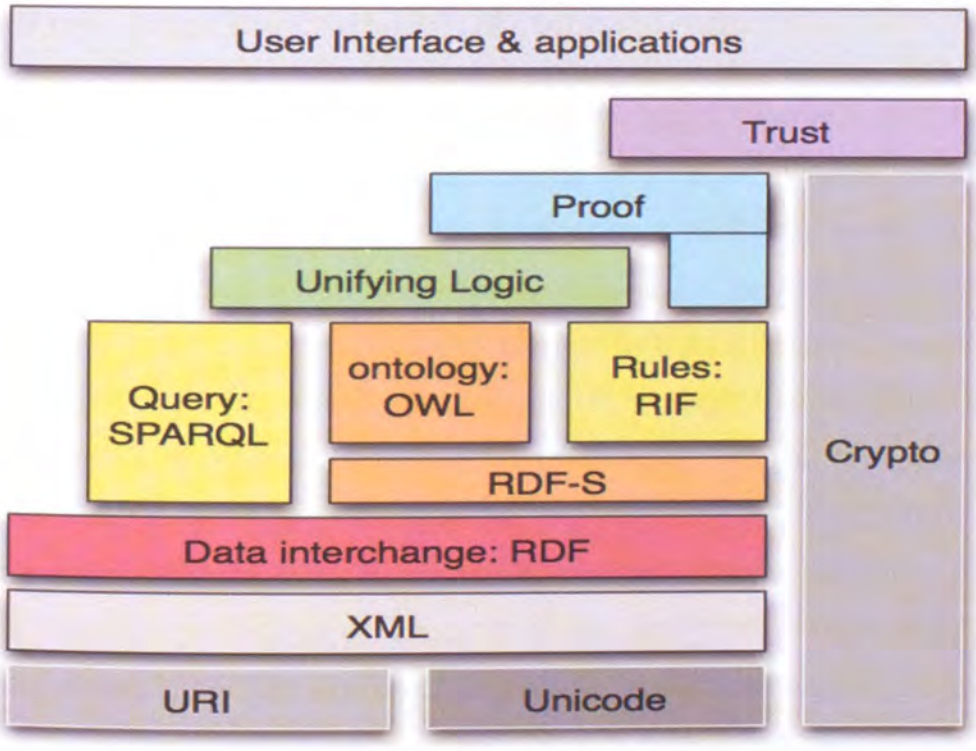


Figure 3.1 Semantic Web Architecture (V4) (Ber06).

Berners-Lee proposed new versions of the semantic web layered architecture in 2005 and 2006, and he referred to as V3 and V4 respectively (Ber05, 06). The V4 of the Semantic Web layered architecture (Figure 3.1) which was presented in a AAAI 2006 keynote presentation

(Ber06) brings some important adaptations. In V4 we can still see the URI and Unicode which are similar to the previous versions. The aim of Unicode is to uniquely identify the characters in all the written languages by assigning a unique number to each character. By assigning a unique numeric value and name for each character it extends ASCII (American Standard Code for Information Interchange). A URI (Uniform Resource Identifier) is defined as an extendable, compact string of characters which identifies a resource. The resource could be physical or an abstract. For example the standard URI specification of the IETF (Internet Engineering Task Force) is RFC3986 (Ber05). The resource is something with an identity and it could be even at a conceptual level. Therefore the resource always remains the same (constant) even if the content changes over time, thus it becomes the semantic web identification problem (SW03). The identifier is an object and it refers to something with identity. For URI, the object is a sequence of characters, ideally Unicode, with a restricted syntax. Whereas the URLs (Uniform Resource Locators) are a subset of URI that particularly identify resources by making use of their network location rather than specifying the resource by name or by other attributes. URL is a compact string representation of the location for a resource that is available through the internet (Ber94).

In layer 2 of the V4 of the architecture the Namespaces is omitted and only XML is specified on this layer. In layer 3 an additional description of RDF is included, which is Data interchange where as in V3 it was named RDF core. RDF-S still resides above RDF. However SPARQL is included which provides a mechanism to query RDF data and resides above the RDF. In order to efficiently extract information from RDF store a query language is necessary (Fra04, Dan04, W3C03, W3C05, Ste10) therefore the W3C has developed SPARQL which can be used in conjunction with a common protocol, so applications can access and combine information from across the Web (W3C05, W3C06, Qui08) In Layer 4, the Ontology Layer, OWL replaces the caption Ontology and Rules, with "Rules: RIF" (Rule Interchange Format) being introduced. Rules remain the same level as OWL in V4 (OWL is the W3C recommendation that are used for the creation of ontologies). In layer 5 the Unifying Logic replaces Logic Framework and Proof is lifted to be above Rules: RIF and Unifying logic, thus residing above both the Rules: RIF and Unifying Logic. The purpose of this layer is to assist with the integration of different formalisms. At present Trust resides above the Proof and it is in Layer 7 of the architecture. In addition a new layer is added above the Trust layer by the name 'User Interface & applications' that seems to represent the notion that all applications and user interfaces of the Semantic Web will reside above all other layers

and it is just above the Trust Layer. The vertical layers in V3 the Digital Signatures and Encryption have been replaced with a single vertical layer called Crypto. Unlike previous versions Crypto no longer starts on top of XML layer but resides alongside of all layers except Trust. Crypto is not yet clearly defined and perhaps it is a combined layer reflecting the security needs of the Semantic Web architecture.

However, it is the knowledge representation languages that give meaning to the web. Multiple languages play a role in the semantic web architecture. As per the semantic web definition, the languages play a part to exchange meaning and knowledge to the web content. The next section explains knowledge representation languages.

3.3 The Knowledge Representation Languages

A knowledge representation language can be defined as any language that is used to represent knowledge. It can be understood in five distinctive ways. According to Randall Davis et al (Ran93) a knowledge representation (KR) is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e. by reasoning about the world rather than taking action about it. It is a set of ontological commitments, i.e. an answer to the question: "In what terms should I think about the world?"

It is also a fragmentary theory of intelligent reasoning, expressed in terms of three components: (i) the representation's fundamental conception of intelligent reasoning; (ii) the set of inferences the representation sanctions; and (iii) the set of inferences it recommends. Moreover, it is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences. Finally, it is a medium of human expression, i.e. a language in which we say things about the world.

The process of tracking information requires a knowledge representation language. It is sometimes the case that knowledge representation languages are not used properly and therefore the knowledge cannot be transferred accurately. In semantic websites there are two primary languages that are used to represent the knowledge. They are RDF and OWL.

3.3.1 RDF (Resource Description Framework)

The Resource Description Framework (RDF) is a language for representing meta-data or information about resources on the Web (W3C04). It also standardises the definition and usage of meta-data, and in the context of Semantic Web it helps to capture the data from web resources. Therefore RDF supposed to exchange meta-data about resources between applications without loosing the meaning. RDF language is used in a sentence that is really in URIs. Each RDF sentence has three parts: a subject, a predicate and an object.

As shown in figure 3.2 <http://www.CISMpapers.org/kingston.html> has an author (creator) and its value is Teaching Staff. This RDF statement describes the first URL as a subject. The subject is the 'http://www.CISMpapers.org/kingston.html' and the object is the phrase 'Teaching staff'. The predicate is the word creator, which relates the subject to the object. Therefore, in this example the RDF statement states that "CISMpapers are authored/created by Teaching Staff." RDF statement can point out to anything. In fact, RDF adheres to the principle 'anything can speak anything about anything.'

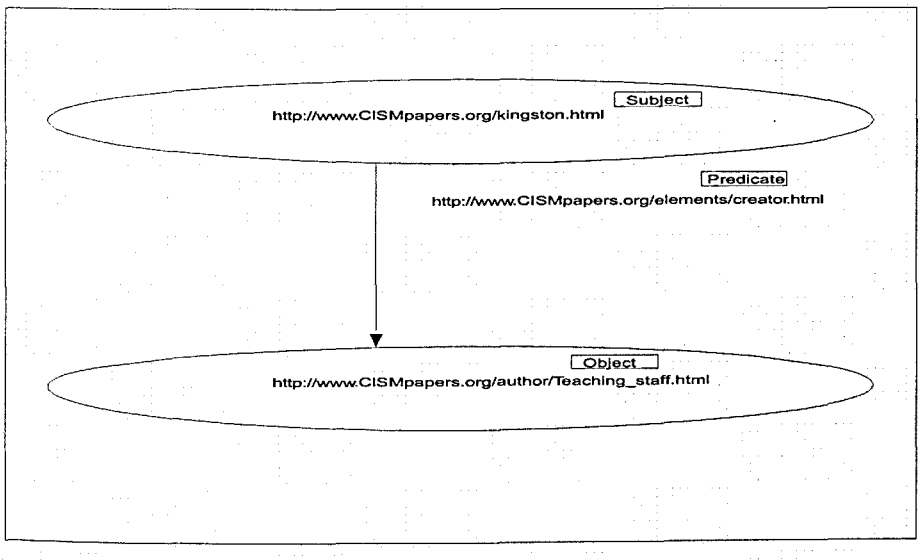


Figure 3.2 RDF Statement Graph

3.3.2 Ontology

The terms Ontology derived from the two Greek words οντος which means *of being* or *to be*, existence and λογία which means to reason, therefore Ontology can be defined as the study of the nature or being, reality or existence, or reasoning about existing things (Ken03). The word Ontology originally used in philosophy where it means the study of the nature of

existence (Gru93) and it is traditionally remained as part of metaphysics. Ontology is a recent introduction into the computer science and a numerous number of definitions have been proposed (Las01), (Gua97). The most common definitions of Ontology is "An ontology is an explicit and formal specification of a conceptualization" (Ant04), "content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge" (Cha99). In this sense Ontology "include a vocabulary of terms and some specification of their meaning" (Usc98). Ontology therefore attached to a domain where it describes the relationships, existing entities groupings such as relationship within a hierarchy etc. and their meaning in terms of concepts and relationships. For the purpose of this study we are dealing with Ontology and its usage in semantic web. In semantic terms, it is the existence of classes, properties attributes and their relationships in Semantic Web.

3.3.2.1 Web Ontology Language (OWL)

OWL stands for Web Ontology Language, and it is built on top of RDF. Web Ontology Language (OWL) was designed to be compatible with the architecture of the Web particularly Semantic Web (McG04). OWL provides constraints applicable to classes and properties (McG04, Hef04). OWL is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans (W3C04). OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. (McG04, Hef04). This representation of terms and their interrelationships is called ontology (W3C06). OWL is therefore used to process information on the web. OWL has three sub languages:

OWL Lite supports those users particularly considering and essentially in need of a classification hierarchy and simple constraints (W3C04). For example, while it supports cardinality constraints, it only allows cardinality values of 0 or 1 (W3C04). It is also noted that it should be simpler to supply tool support for OWL Lite than for its more expressive relatives, and OWL Lite provides a quick migration path for thesauri and other taxonomies and also provides a minimal useful subset of language features (McG04, Hef04). Properties can be made optional or required in OWL Lite. OWL Lite also has a lower formal complexity than OWL DL.

OWL DL (Description Logics) supports those users who want to express themselves maximum while retaining computational completeness and decidability (W3C04). In this case all conclusions are well guaranteed to be computed and all computations will finish in finite time. OWL DL includes all OWL language constructs, but they can be used only under certain restrictions (W3C04). For example, while a class may be a subclass of many classes, a class cannot be an instance of another class. OWL DL is so named due to its correspondence with description logic, a field of research that has studied the logics that form the formal foundation of OWL (Sea06).

OWL Full The complete OWL language is named OWL Full to differentiate it from the subsets and is intended for users who want to maximise expression and the syntactic freedom of RDF with no computational guarantees (W3C04). For example, in OWL Full a class can be considered simultaneously as a collection of individuals and as an individual in its own right. OWL Full also permits an ontology to augment the meaning of the pre-defined (RDF or OWL) vocabulary. It is noted that any reasoning software will not be able to support complete reasoning for every feature of OWL Full (Sea04).

3.4 The Semantic Web formation

The real semantic web formation consists of semantic aspects as well as web aspects (McG04). Semantic aspects may be limited in a normal website, and must be identified and defined, so that appropriate web solutions within these limits can be formed. The process of describing a semantic web requires the collection of semantic data. It is important that semantic aspects of the site are not forgotten during the planning stage.

3.4.1 Methods and Languages

A number of methods, semantic languages and tools are used in designing a typical semantic website. Semantic languages, as explained in previous sessions, provide the primary elements, since they are always required to build a semantic site. The efficient languages used to represent knowledge on the web are, therefore, essential to a semantic website formation. One way of distinguishing the efficient web language is by its compatibility on the construction of a web page and its representation; even a basic web language can be distinguished according to this criteria.

Another important element in designing semantic websites is the 'knowledge based system' (KBS) (Dav08), which has a major impact on semantic websites. KBS records all the sensitive information which could be critical to the success of the semantic website. This includes storing, retrieving and modifying information on the web. KBS also includes all basic information such as code-oriented programs, debugging codes and all information regarding the running of a business (Dav08). One example of a knowledge based system database for Semantic Web is DLDB-OWL (Yua04) (Description Logic, Relational Database- Ontology Web Language). Since there is a need to ensure consistency and scalability between the web and the reality of the information, the effectiveness of the database should be tested against a set of criteria, which may include attributes such as reliability, scalability, security, information retrieval, efficiency, maintaining cost etc (Yua04). The data may be quite sensitive or invaluable, however information is valuable and therefore it is necessary to safeguard the information on the semantic web.

Another important factor which helps in the understanding of the semantic web is the design itself, which contributes to the semantic website formation. By measuring the design method one could identify the real semantic web over the net. The design method which is used in each semantic website will equally be a decisive factor in finding the site on the web, for example, through the usage of the syntax. One important characteristic to consider in the design is the basis of decision making processes, for example, what language will be used? What database? What Semantic documents? What archiving? etc.

Moreover, the number of semantic documents reflects the semantic nature of a web system. For example the query "rdf filetype:html" identifies more than 38 million HTML documents (LiD06) . According to the queries run on 12 May 2006 by Li Dingl and Tim Finin, it was estimated that there are between 107nd 109 Semantic Web documents online," (LiD06).

3.5 Semantic Documents

Documents are used to analyse and verify the quality of a semantic web. Documents express the meaning of a website, and the semantic collection of information. The semantic documents which are hosted on the website of an organisation can be seen as measuring the semantic nature of the site (LiD06). The prime success of a semantic web design model lies in its ability to project the organisation meaningfully to the public. Also in line for consideration are the semantic documents used to design the site. Semantic documents are an

important factor allowing both measurement and expressing the semantic nature of the site. The semantic documents are mainly measured according to the performance of the site, target audience for example consumers. In order to classify the semantic documents over the net, each website needs to be analysed. Various methods are used to analyse the semantic web documents. For example, bootstrapping, google-based meta-crawling, bounded HTML crawling, RDF crawling, Inductive learner and Swoogle Sample dataset can be used for this purpose (LiD06).

Quality assurance of the semantic web is made through the harvesting techniques (Jan03). However, the harvesting processes are not authenticating many websites as semantic, although they claim to be semantic in nature. "The dataset SW06MAY resulted from harvesting data between January 2005 and May 2006 has 3,675,153 URLs, including 1,448,504 (40%) confirmed as SWDs (Semantic Web Documents), 13% confirmed as non-SWDs, 9% unreachable URLs, and 38% unpinged (not yet visited) URLs. The confirmed SWDs are from 162,245 websites are 279,461,895. Although SW06MAY (Semantic Web Ontology) is much smaller than the Web with its 11.5 billion documents, it is much larger than any existing datasets, Eberhard reported (Ebe02) 1,479 valid SWDs out of nearly 3,000,000 URLs (2003) OntoKhoj reported 418 Ontologies out of 2,018,412 URLs after 48-hour crawling (2004) DAML Crawler reported 21,021 DAML files out of 743,017 URLs (LiD06).

The semantic web ontology usage is also needed to be measured. In an evaluation is carried out in (LiD06) to prove the percentage of semantic ontology which is currently used over the net. Dataset SW06MAY resulted from harvesting data between January 2005 and May 2006, which contributes 83,007 SWOs including many unintended ones, such as (i) instance data with unnecessary class or property definitions or references, e.g. 55,565 (66.9%) PML documents (Proof Mark-up Language) from onto.stanford.edu, and 882 (1.1%) semantic blog documents from lojjic.net, and (ii) instance data that has unnecessary instances of owl:Ontology, e.g. 4,437 (5.3%) publication metadata pages from www.aifb.uni-karlsruhe.de and more web portal metadata pages from ontoware. org. Therefore, the "true" number of SWOs in dataset SW06MAY is just 22,123 (26.7%) SWOs after removing the "unintended" ones. Moreover, this number can further reduce to 13,012 (15.7%) since there is much duplication (LiD06).

It is important to note that there is often a difference between the website and the semantic web and the actual designing of the site. The semantic website can be seen as a website to achieve control over business, it can be meaningfully placed over the net, and measured and analysed in terms where the website is not meaningfully expressed. It all depends on the document usage. Research (LiD06) has proven that only 13% of semantic web documents are found over the net which may be due to many factors, including ignorance of the semantic web, delays in the designing processes, poorly achieved designing, etc.

It has also been proven that it is difficult to estimate accurately the semantic nature of the site, and to measure the semantic documents over the net. Research proves that many sites which are claimed to be semantic, in fact, are not semantic (Mar06). The documents used to control and design a semantic web can also be measured and classified. The next section describes the groups and technologies involved with the semantic web.

3.6 Groups and Technologies

A number of semantic web service groups and approaches are available for a business project. Semantic web services groups, the primary web service provider, are essential to a business information service, and in finding new solutions to business. One way of identifying a group is by their technology, and even the most advanced group can be distinguished according to this. One grouping, for example, includes the semantic web interest group, which is a forum to support developers and users of Semantic web technologies such as RDF and (W3C01). This group helps developers to create vocabularies and applications to support web data and it also combines harvesting, syndication, metadata and web service techniques (W3C01).

Web modeling-based approaches are used to automate web services mediation, choreography and discovery. The WebML group was created by researchers in the data base group of the Dipartimento di Elettronica e Informazione at Politecnico di Milano (Web05). This group is based on WebML language. The ESSI WSMO working group, part of the ESSI Cluster safeguard the research and its development initiatives in the field of Semantic Web Services and semantic web which lies in between SEKT, DIP Knowledge Web, and ASG research projects (Mac07).

The Digital Enterprise Research Institute (DERI) Group's aim is to mechanise the interoperability of a system (Mac07). They argue for and try to achieve the mechanisation through the self description of data and its processes incorporating the semantics (Mac07).

Within a semantic web business, different groups have different approaches. For example, the DERI group based its solution on the WSMO conceptual framework for semantic web services. However, DEI (Department of Electronics and Information) is based on well established software Engineering methods (Dei09).

The achievements of each group depends on their technology, therefore technology has a major impact on the semantic business. Technology covers all the semantic web services, from the designing phase to the running of the system. The efficiency of the technology is based on its reliability, integrity, security, consistency etc. Moreover, the architecture of implementing a technology is very much important, also it varies according to different languages and ontology, for example, special artefacts have to be created during the design time phase to apply WSMX (Mic05).

3.6.1 Web Service Execution Environment (WSMX)

The Web Service Execution Environment (WSMX) is a reference implementation of WSMO (Web Service Modelling Ontology) using the WSML (Web Service Modelling Language) family of languages (Mic05). DRI (Digital Repositories Interoperability) rooted its solution on the specifications of WSMO (Rom05) and is based on the web service modelling framework (WSMF). Fensel & Bussler (Fen02) states that WSMO solves the integration problem of web services by describing the services semantically and thus removes ambiguity with regard to the capabilities of a web service and the problem that is trying to solve.

Since WSMX is an execution environment, it provides dynamic service discovery, conceptual framework, mediation, ontology language, architecture and invocation of semantic web services (Mic05). Therefore, WSMX uses WSMO as its conceptual model and it defines its own execution semantics, architecture, and implementation (Arm05, Mic05, Lau04).

3.6.2 WEBML

WebML (Web Modeling Language) language is a high level visual notation for complex data and process centric web applications (Ste07). WebML provides a notation for specifying the complex web sites at the conceptual level. It provides the conceptual modeling of web applications under distinct orthogonal dimensions, it is built above a data schema and it describes the data content such as the structural model, composition model, navigation model, presentation model and personalisation model (Mac06). The WebML data model is the standard Entity-Relationship (E-R) model extended with Object Query Language (OQL) constraints (Ste2004)). In specific technological settings WebML is able to bring forth a site implementation, however, it is independent of both the client side language (application to users) and of the server side platform (bind data to pages). A model driven approach is also associated with WebML, thus it can be made available to build complex sites by way of defining a novel generation of CASE tools. It also supports the advanced features, for example, multi device access, personalisation, and evolution management (implemented in ToriiSoft, a pre-competitive web design tool suite) (Ste04).

WebML uses its well defined semantic for a project, and therefore it is capable of mapping between the model of persistent business objects and a relational data source. WebML language is extendable and permits customised operations and units. For example, it is implemented in the CASE tool WebRatio (Mac05). It generates the complete code of the application in SQL queries, Java and JSP code and Struts files for J2EE platform. It is possible to change the styles of an application. The design scenario is equipped with the facility of automatically producing complete and high quality documentation for the project. The wizard also allows the user to choose from predefined patterns and thus it provides the facility for generating a set of pages, units and links and inserting them into the model. These units and sub-pages inserted into the pages, always refer to the master pages.

3.6.3 The Interoperability Problems

The challenge of modeling the web service mediator with WebML is explained by webml.org/sws (Web05, Tiz07) as follow: The mediator bridging the gap between the process and the data required and the process and the data offered create a data model. In order to overcome the problem of mediation problem is the designing of the RosettaNet

message with the extended E-R model. This created a purchase order entity to store basic details of each order.

A partner entity to model business partners that are involved in the order is also explained. A product line item entity to model the product lines composing each order is an entity to keep trace of the current order.

For each activity required by the WS mediator a proper chain of WebML operations are designed. Once the message is received its order ID is checked against the existing ones. If another order is found with the same ID an error is returned, if no order with the same ID is found, the order is stored by the XML in unit by means of a proper XML adapter. Finally an acknowledgement (ACK) message is sent back to the client. The buyer is selected and a SOAP message is sent to the CRM (Customer Relation Management System) legacy system. The obtained customer ID is stored. Buyer and receiver are selected and a SOAP (Simple Object Access Protocol exchanges structured information) message is sent to the CreateNewOrder operation. The obtained Order ID is stored. For each line in the order a SOAP message for the ADDLine operation is composed and sent to the Moon legacy system. A CloseOrder message is also sent to the moon legacy system. For each LineConfirmation received, the original line is selected in the data and its state is updated according to the confirmation message. If confirmation has been received, the RosettaNet purchase order confirmation message is created and sent to the blue Rosetta net client.

There are differences and similarities between WSMO and WebML. The DERI follows the WSMO and expresses mainly four elements such as explicitly modelling goals, Web services, and ontology. While WebML is based on the modelling of the web mediators, the internal logics of the services and the visual diagrams which represent the execution chains are used to define them. It also gives little emphasis to the design of the goals and of the other semantic aspects. However WebML supplies semi-automatic extraction of WSMO goals and web service behaviour from the designed models and the designer will refine them later (Mac05).

WSMO and WebML are concerned with data models, for example domain specific knowledge and data schema exchanged messages in instance level. WebML specifies the data model by way of describing domain data structure as on Entity-Relationship (E-R) or UML class diagram. WebML-OQL is used to model the domain. Though the logic rules are not

explicitly supported, the expressive power of the model is close to WSML-flight which is an F-Logic (frame logic) programming variant branch of WSML for modeling process ontologies. WSML (Web Service Modeling Language) uses a fully fledged ontology language with rule support. In modelling the ontology SWS-challenge does not make use of complex rules. However in terms of expressiveness both DERI (Digital Enterprise Research Institute) and DEI (Department of Electronics and Information) data models are quite similar. For example WSMO makes use of WSMT and WSMO studio and WebML uses the WebRatio CASE tool (Mac05).

Though the two groups namely, DEI and DERI based their concepts on webservices, the process modelling techniques differentiate between DEI and DERI; DEI-Cefriel models Moon orchestration with software engineering, and DERI makes use of ontology based abstract state machines for specified orchestration (Dum06, Tiz07).

In DERI the orchestrated service is not aware of any incoming or outgoing RosettaNet messages. It specifies messages in its native ontology and it is up to the Data Mediator to resolve and mediate data heterogeneities between the service requester and service provider (Mac05). Both differ in data mediation and tool support. WSMO Choreography modelling still need modifications and it has to be matured just like the other business process modelling offered by other platforms (e.g. WebML graphic process modeling) (Tom07). These two approaches are not capable of modelling Web systems properly at higher levels of abstraction and to make relationship efficiently with models of the business model and processes (Rac04).

3.6.4 WebML+

WebML+ is an extension of WebML developed by David Lowe & Rachatrin (Rac07) in which the developers can express the core features of a system at a higher level. It will not really commit developers to detailed architectural designs. The core functionality of this system is to define both the internal and external information flows within a web system (Rac04).

This is an architectural level specification language and defines web systems based on an extension of WebML. It claims that it can decompose the information into WebML models and thereby affords the opportunity to link the abstract system modelling to the detailed design model in WebML, which can provide a better view of architectural information than

the typical site maps which are in use (Rac07). It also claims that it can connect between the information and functional perspectives of a system.

It is certainly clear that the use of web services and modelling have to be improved in various levels and stages than those derived using the early semantic systems such as WebML or WSMO (wsmo.org). In WebML, for example, a model of the website is constructed, and monitoring techniques are implemented to check the performance. As long as the website is in use continuously it should demonstrate dynamic and consistent results. Therefore, if the website performance differs substantially from the goal (capability) level, then the website has probably changed in some way. The expected semantic results are determined against the semantic items on the web. The current website is being checked to see the semantic performance. Thus, any website that is below the semantic capability must be examined. By this technique, the efficiency of each semantic website can be determined. For example the semantic items of a website may be dynamic, but may require enhancement to meet the semantic standard.

3.7 Summary

This chapter started by with a discussion of the meaning of the term "semantic" in an attempt to present some definitions and characterisations of semantic web. Therefore, it has presented the historical evolution of the word "semantic" and semantic web. After considering the architectural issues in semantic web the chapter has reviewed several knowledge representation languages and examined their use in semantic web and their associated processes. Moreover, the review continues with architectural issues particularly in term of web services, their interoperability problems and the potential semantic links of web services. Overall, this chapter provides some background, in terms of basic concepts, definitions and technologies to the work presented in this thesis, which has influenced the proposed architecture in Chapter 5, as well the choice of the used technologies particularly in chapter 7.

Chapter 4 - Web Sites Semantic Nature Evaluation

4.1 Introduction

Despite the massive increase in the number of websites representing public as well as private sector enterprises, many still suffer from a number of limitations and problems such as poor navigation and design, inaccuracy of available information, poor meaningful linkage of information etc. which semantic web is trying to address. The purpose of this chapter therefore is to propose a quality model with specific quality factors/metrics, and to conduct an evaluation of existing websites, mainly E-commerce sites, using the identified metrics in order to assess the many aspects that effect their quality and semanticity, with the aim of designing a new approach that addresses those limitations.

Although several attempts have been made to improve the quality of websites, over the years many concerns have been raised with the construction and technical correctness of the used approaches (Bri10, Son08, Hea02, Sch01, Nie00), particularly in term of functionality, for example, the location and choice of the function buttons and the information accuracy behind those buttons or navigation tools. Moreover, the distinction between global and local navigation in the functional buttons is not always very clear. Functional points sometimes do not appear to be leading to the real-time information in an expected way.

There are also limitations in the usability of websites; a poorly designed site with poor navigation can make the use of the site difficult. In fact, many well established businesses may find it difficult to find a place in E-marketing due to the fact that their sites are poorly arranged or designed (Nan05). These include interaction with the customer and providing sufficient information regarding the products and services. Moreover, the reality (real time data) and relationships of presented information are also major issues. The relationships are made meaningful by the accurate location and presentation of websites functions.

Apart from the individual quality factors, functionality, usability etc., mentioned above other important factors are included in the proposed quality model. It is clearly that in order to

produce a broad quality evaluation of a website many different quality factors and characteristics need to be considered. At the highest level software quality factors such as usability, functionality, correctness etc. are considered, however, the relative importance of these factors varies depending on the sector in which a site is operating and the intended users/customers. In this work the evaluation is done in the context of E-commerce therefore additional factors reflecting this sector, needs to be considered. Moreover, in the context of this work it is important to get an idea of the potential semanticity of a website, which means some factors to reflect semantic issues need to be included.

This chapter starts with a brief discussion of the approaches commonly used for the evaluation of websites which are normally based on one or more aspects of the quality characteristics. Then, the software quality factors which can be used for the evaluation of websites for their semancticity are considered. This is followed by proposing a quality model which is then used for the evaluation of some leading E-commerce websites. Moreover, some specific quality characteristics directly related to the indentified quality factors are considered together with the different types of websites structures and their impacts on quality characteristics. Finally, the chapter's findings are summarised in the conclusion section.

4.2 Quality Factors for Website Evaluation

The evaluation of the quality and the type of services offered by exiting or newly constructed websites requires the identification and the definition of the corresponding quality factors/characteristics such as usability, functionality, correctness, integrity etc. In order to analyse the nature of existing websites, this section starts by reviewing some of the traditional and current trends in websites evaluation.

The evaluation of websites for their quality of service has been considered in a number of papers (Spre08, Kim05, Goo87, Dra99, Wol03, Zha01) Goodwin proposed an approach to evaluate a website purely based on its functionality in order to meet the users' requirements i.e. giving them the ability to perform the necessary tasks (Goo87). However, Daren et.al suggested the use of Kano's Model of Quality as a theoretical framework to evaluate the websites (Shi10, Eme04, Dra99, Kan84) Kano's model divides products, services and quality into three levels based on customer expectations such as expected, normal, and exciting (Kan84, Lil05). According to this approach it is argued that websites evaluation should be carried out according to the quality of service received by the customer. Ivory and Hearst

proposed that website evaluation falls into three areas such as testing, inspection and inquiry (Ivo00). This implies that qualitative and subjective information can only be unveiled through user testing, heuristic evaluation and other methods. However, a large number of researchers seem to agree that website evaluation is mainly based on usability and performance (Mat10, Dia10, Wil08, Bra00 Sch99). Moreover, many of the existing approaches do not take into consideration real time information, linkage of information, social-cultural aspects etc. Therefore in the following sections an extended number of quality parameters will be considered with special emphasis on semantic web evaluation.

4.2.1 Quality Model

A quality model links together and defines the various software metrics and measurement techniques that can be used for software projects. In this work the adaptable quality model, or ADEQUATE for short (Hor09, Kha05) which provides a set of standard quality factors, is adopted. The model is adaptable to different projects and includes some global key quality factors (KQF) as well as locally defined factors (LDF) which can be adopted for specific business needs (figure 4.1). In this way, the KQFs represent a common set of criteria that can be used for cross comparisons, whilst the LDFs retain the ability to allow local tailoring. The LDFs are not a replacement for the KQFs. Instead, they define additional quality criteria.

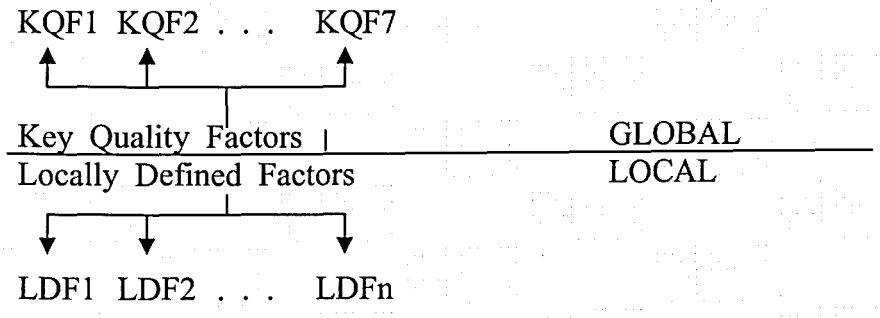


Figure 4.1 Global and Local Quality Factors

There can be many factors, in total eight factors have been identified, which are considered important to evaluate websites. These factors are interactivity/functionality, usability, correctness, integrity, real-time information, information linkage, customer care and socio-cultural aspects. No other factors are included in the model. The first four are typical quality factors used in many software projects including web development. The real-time information and information linkage are important from semantic web review and finally

customer care and socio cultural aspects are very important from user/customer point of view particularly for commercial websites.

In the context of the semantic information search, interactivity/functionality is seen as the ability of a semantic web to function and to interact with the customer and the entrepreneur. This is included as an important factor in semantic information search systems. Almost all of the identified approaches include a reference to interactivity and functionality. In order to offer goods and services to the customer, the site has to be accurate, reliable and the goods and services must be clearly presented in real time. A site with poor interactivity/functionality will not target the desired customers. The interactive nature of any E-commerce site must have a functional stability in real time. As Paul Rand (pau05) states functional stability has two components: getting things right at the design time and then keeping things functioning smoothly over time.

Another factor is usability, which is defined as a website's ease of use. If a website is not easy to use, then it is less likely to be semantic. From this fact alone, usability is a very important for semantic information search. Navigation and ease of use is an issue which needs to be addressed by all E-commerce sites. The customers want to know what they are doing, for example, providing sensitive data such as credit card information and other personal information. The real time information regarding the navigation is an essential factor for a good E-commerce site (Fot04).

Another area of consideration is correctness which is defined as the capability of a website to meet and support its functional objectives. Correctness is described in the semantic information search as the ability of a website to meet and assist its objectives and targets without terminating. Moreover, the accuracy of information has to be checked and the information updated regularly to make sure the business is running in real time. The customers cannot tolerate slow downloads and terminating pages (Pau05).

Integrity in a website is defined as the ability of a website to perform honestly and truthfully, therefore the website has to defend its customers at any cost. Customers are really important assets to a business and if the information is not secure in some form or another, then there is the risk to integrity. Therefore, security is considered as part of a website integrity factor which should be given a very high importance.

The real-time information is also an important factor and it is involved in a website's running time, and therefore this should be considered with higher priority, as should the reality of information required to run a semantic information search system. Its value is constituted by providing the functionality of information, and is determined by evaluating the functionality of a site. For example, each button representing information is real or non-real i.e. representing the accuracy of information at various viewing points.

Information linking is one of the major issues of semantic web (Jaf07). If the information is not linked properly it would be difficult for the semantic web to generate meaningful results to the information seekers. This is very important considering how hard it is to find relevant (meaningful) information over the web.

Customer care is another factor that needs to be considered. The website has to be appealing to its customers; it also has to create a visual opportunity to interact with the customers. According to Bob Baxley (Bob03) web design has to offer constant visual and functional confirmation of the user's whereabouts and options, via graphic design, navigation buttons, or uniformly placed hypertext links, as well as to be able and prepared to respond to the customers' inquiries and/or comments.

Socio-cultural aspect is also important in delivering products and services (ste09). The interface designers of the web sometimes fail to understand the socio-cultural aspects of different nationalities because cultures within some countries are very different. Sacred colours in the Judeo-Christian West (e.g., red, blue, white, gold) are different from Buddhist saffron yellow or Islamic green. Subdued Finnish designs for background screen patterns might not be suitable in Mediterranean climates, in Hollywood, USA, or Bollywood-India. These differences go deeper than mere appearance; they reflect strong cultural values and need to be understood without falling into the trap of stereotyping other cultures (Aar00).

Many other quality factors can be included in the model; however, based on current experience and previous research work, the factors discussed above are the important ones and require further evaluation. Moreover, in line with many software quality factors/parameters commonly used in software projects, people tend to resist plans which evaluate many parameters, due to limited resources or tight schedules. Based on previous research (Hor09, Kha05), the number of key factors should be kept between three and eight. Clearly, the quality factors set were chosen for their obvious importance.

However, it is accepted that only empirical validation across a large number of projects can determine the completeness of this set. The parameters used to evaluate the websites are shown in figure 4.2.

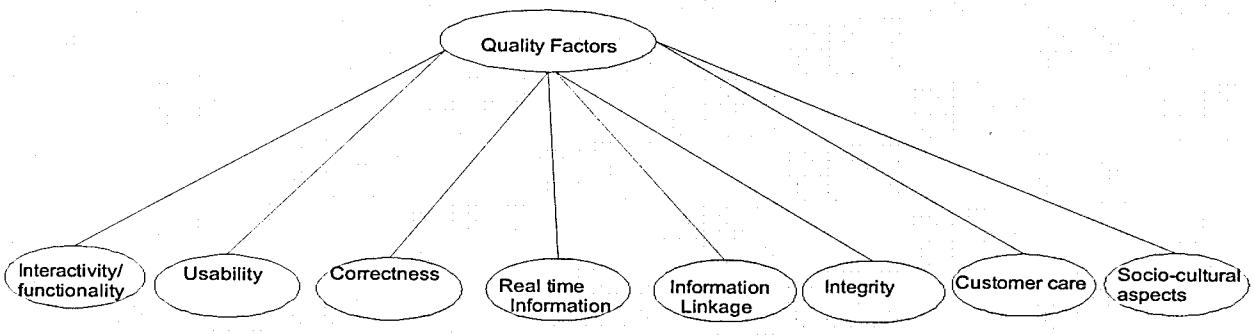


Figure 4.2 Quality factors

In summary the quality factors were chosen to help in determining the reliability, in the broader sense, of websites and in identifying the weaknesses and strengths of E-commerce sites and thereby substantiating their validity.

4.3 Website Evaluation

A number of websites are considered for the evaluation using the quality factors discussed in the previous section. Five company websites are chosen, mainly because they are in active United Kingdom E-business market (Figures 4.3-4.7). They do business in different sectors such as retail, motoring and insurance, financial services, and electronic equipment suppliers.

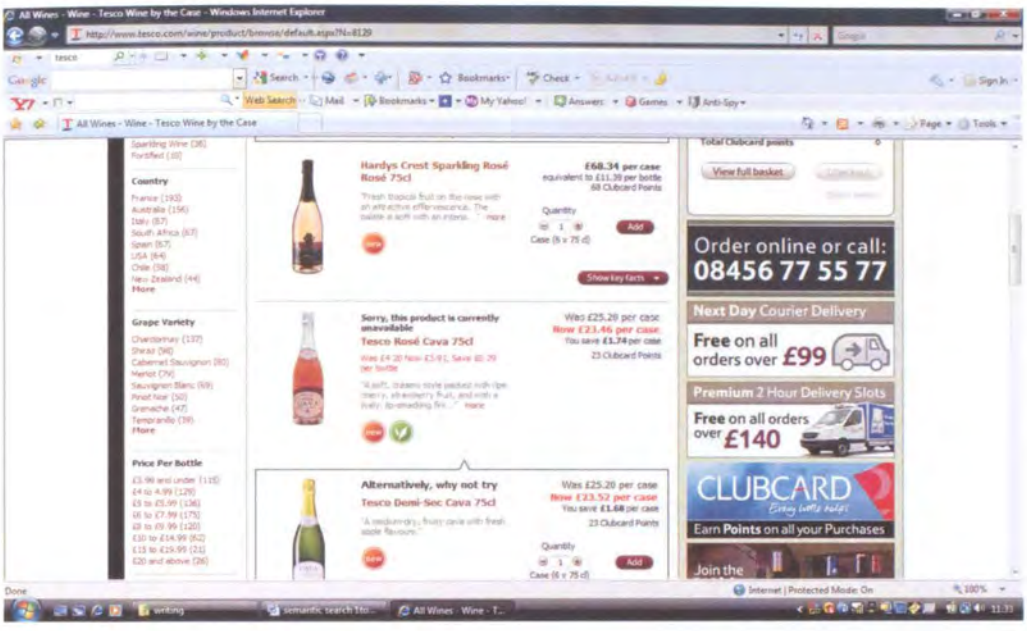


Figure 4.3 Tesco page

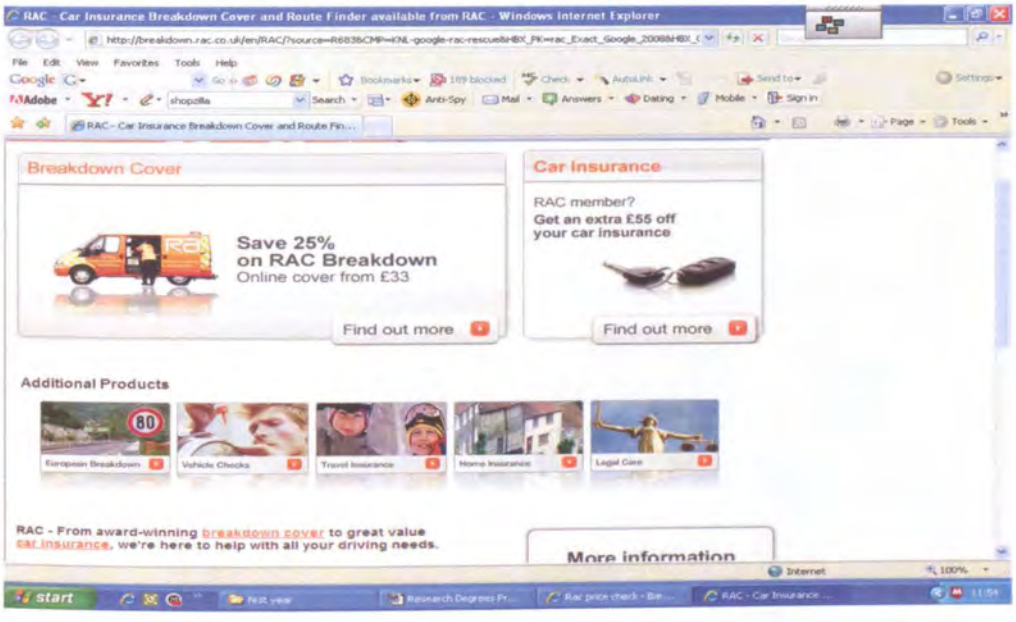


Figure 4.4 RAC page

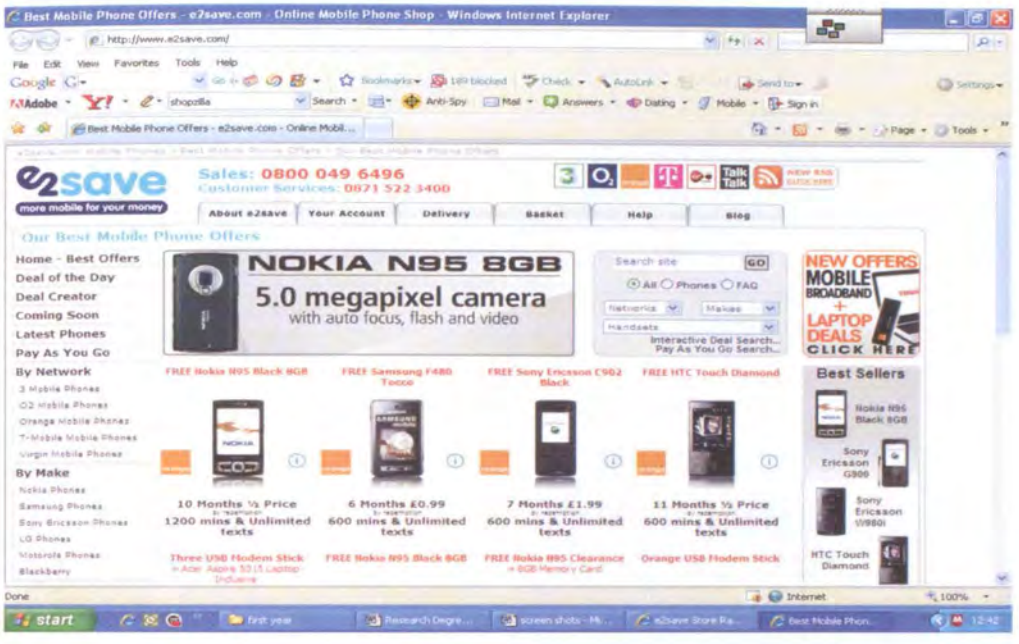


Figure 4.5 e2Save page

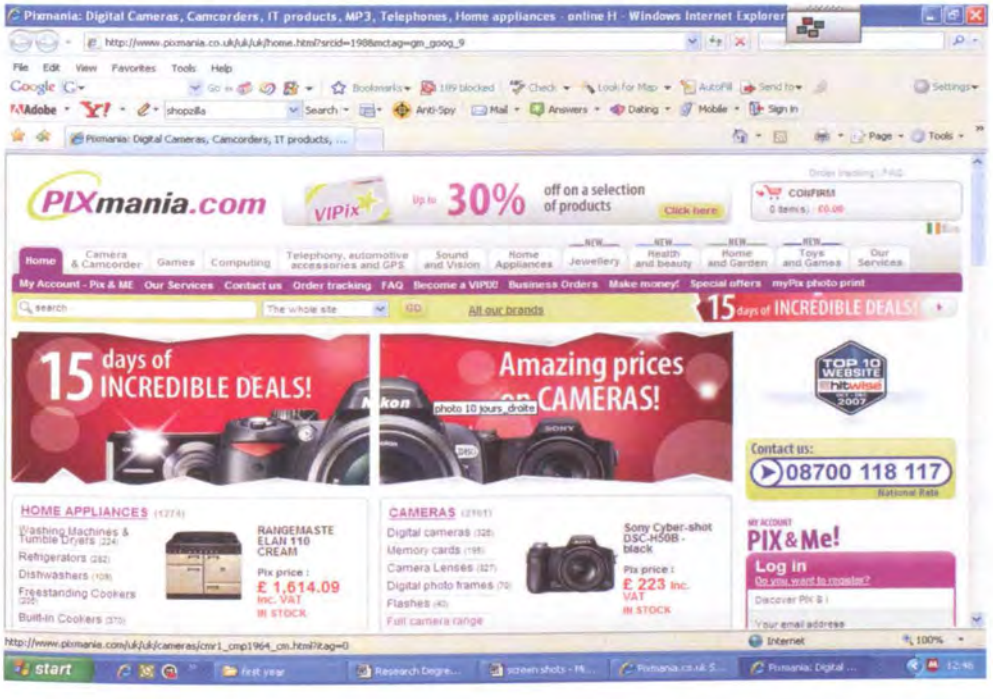


Figure 4.6 Pixmania page

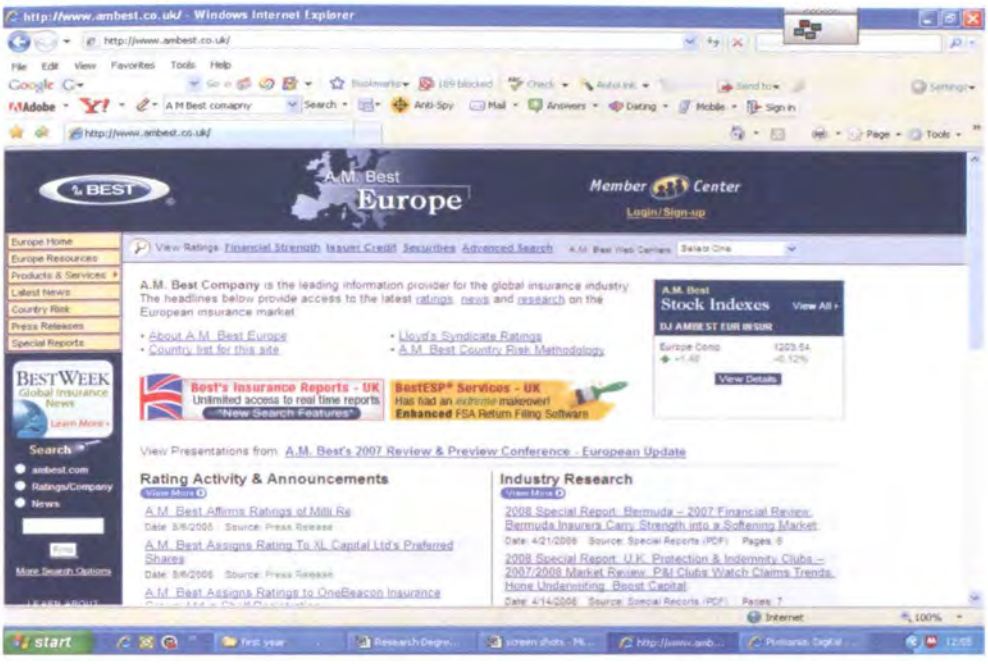


Figure 4.7 Ambest tested page

Likert scale is one of the most commonly used survey system which aims to achieve the correct scaled response from the survey questions. This method was developed by the American educator and psychologist Rensis Likert in 1932 as an attempt to improve the levels of measurement in social and scientific research (Can06). There are a number of different scales, such as 5 point scale and 7 point scale which are used to analyse survey responses. However, Likert's 5 point scale is the most popular and well suited for this study (Can06). In a 5 point Likert scoring system the maximum score is 5 which mean that the parameter is fully considered, for example if usability achieved a score of 5; this means that the site is highly usable with an excellent graphical user interface (GUI). If a parameter is very poorly taken into account, or not even considered score of 1 will be given, taking usability as an example this means that the site is not usable, or has a very poor GUI.

Evaluation data was collected by means of a questionnaire, delivered online using a web-based survey. In total 62 responses were received; the frequencies of responses to scaled questions examples are presented in figure 4.8. A snapshot of survey pages identifying the responses to each question are shown in figure 4.9.

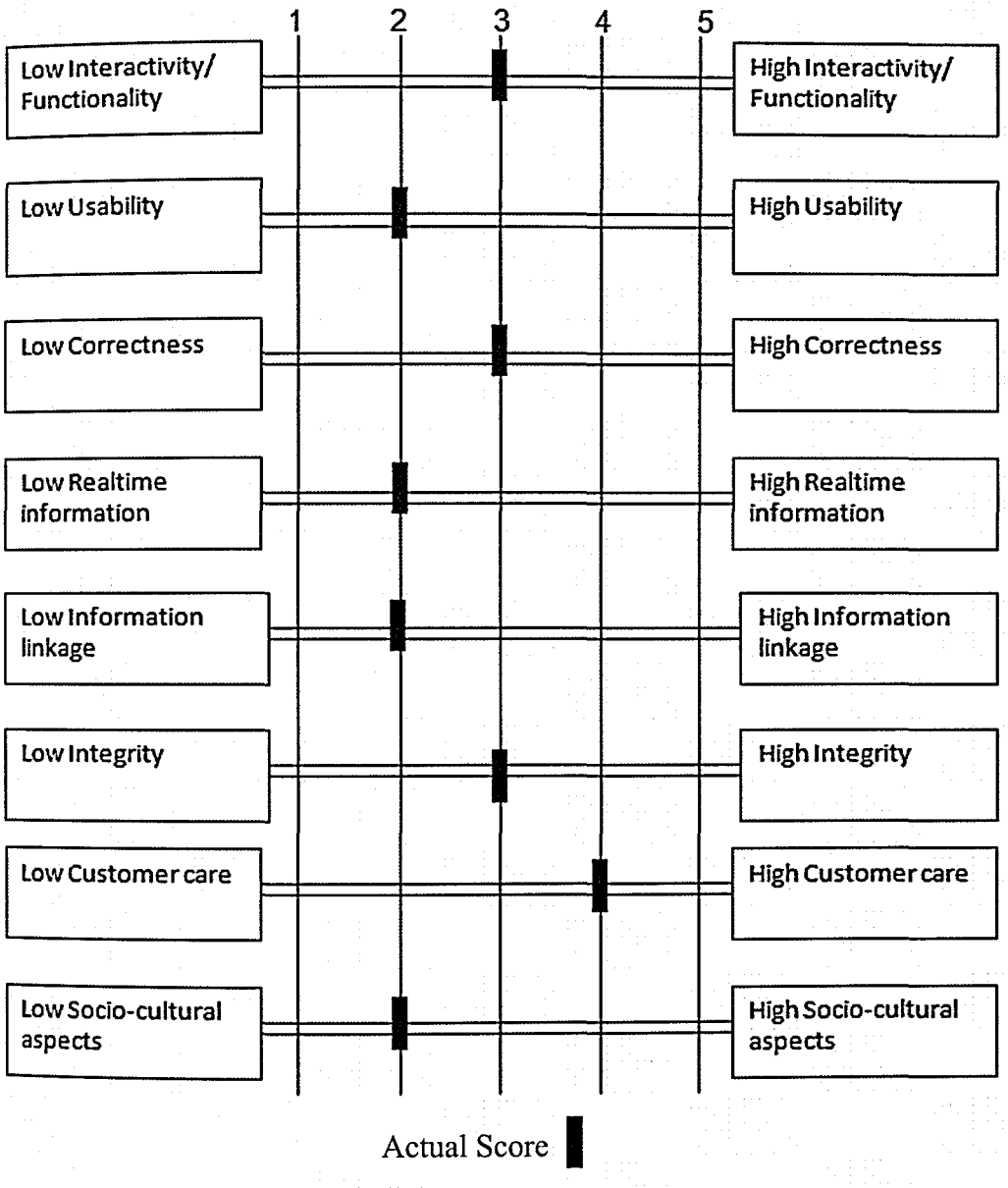
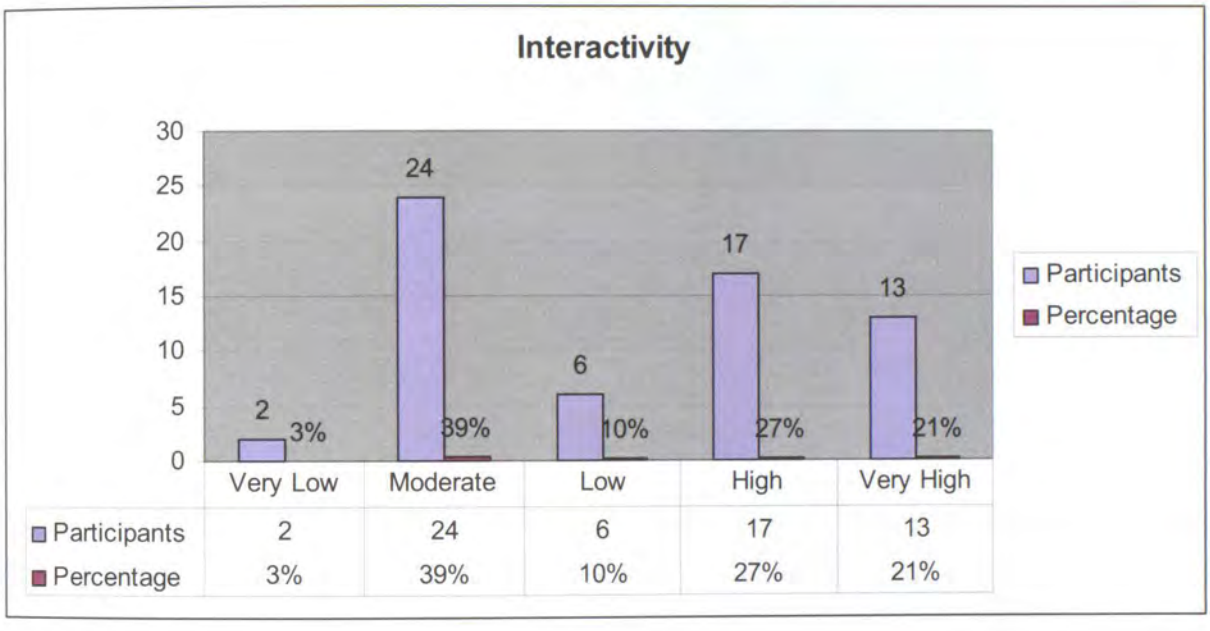


Figure 4.8 an example score profile for TESCO

Figure 4.9 customer reviews page

In order to evaluate the survey responses a statistical analysis on the obtained questionnaires was performed. The analysis was conducted using on-line statistical analysis tools that commonly used in scientific studies (phy09). In figures 4.10-4.17 the scores for individual quality factors as given by the survey participants are shown.



	Very Low	Moderate	Low	High	Very High
Participants	2	24	6	17	13
Percentage	3%	39%	10%	27%	21%

Figure 4.10 Interactivity responses

The results of Interactivity/functionality show that 21% (13) of the participants found the interactivity very high, 27% (17) found it high, 39% (24) found it moderate, 10% (6) found it Low and 3% (2) found it very low with the max score of 39% (14) and the lowest of 3% (2). This gives us a mean of 12.4 with 95% confidence interval for actual mean being between 1.556 and 23.24, and standard deviation of 8.73. The median is 13.0 with average absolute deviation from median being 6.60. Therefore the obtained evaluation confirms the suitability of the used sample for the evaluation of the Interactivity/functionality of the site, TESCO in this case.

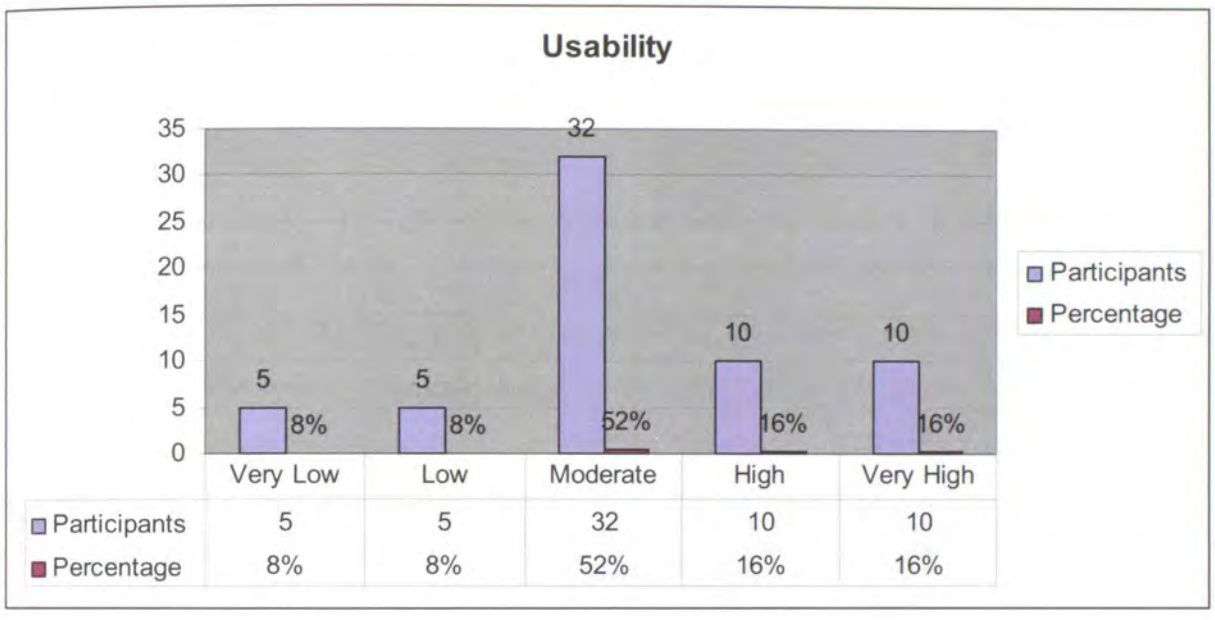


Figure 4.11 Usability responses

The results of usability show that 16% (10) of the participant found the usability very high, 16% (10) is high, 52% (32) is moderate, 8% (5) is Low and 8% (5) is very low which the max score of 52% (32) and the lowest of 8% (5). This shows us a mean of 12.4 with 95% confidence interval for actual Mean being between 1.552 and 26.35, and standard deviation of 11.2. The Median is 10.0 with Average Absolute Deviation from Median = 6.40. Therefore the obtained evaluation confirms the suitability of the used sample for the evaluation of the usability of the site.

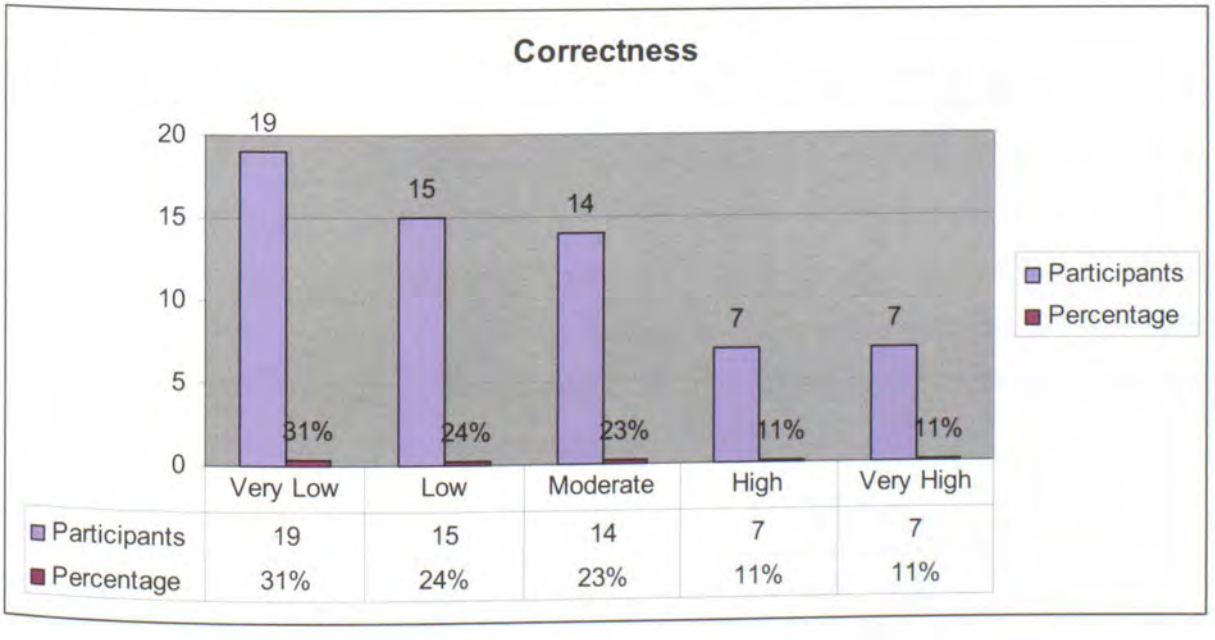


Figure 4.12 Correctness responses

The results of correctness show that 11% (7) of the participant found the correctness very high, 11% (7) is high, 23% (14) is moderate, 24% (15) is Low and 31% (19) is very low which the max score of 31% (19) and the lowest of 11% (7). This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 5.854 and 18.95, and standard deviation of 5.27. The Median is 14.0 with Average Absolute Deviation from Median = 4. Therefore the obtained evaluation confirms the suitability of the used sample for the evaluation of the correctness of the site.

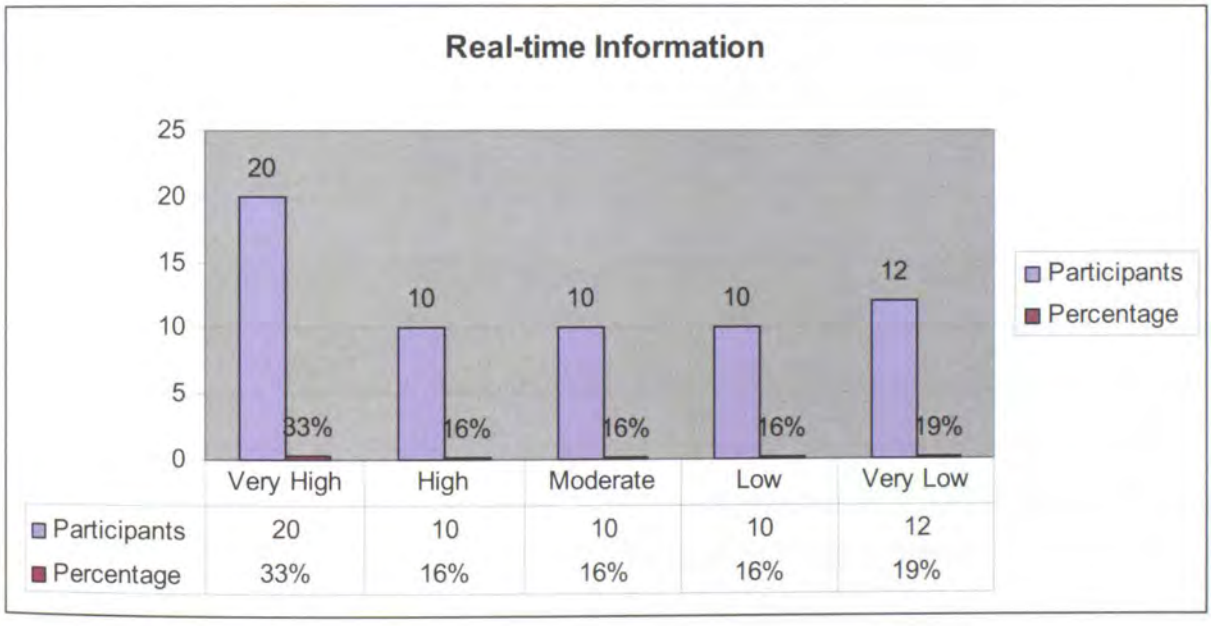


Figure 4.13 Real-time responses

The results of real-time information show that 33% (20) of the participant found the real-time information is very high, 16% (10) is high, 16% (10) is moderate, 16% (10) is Low and 19% (12) is very low which the max score of 33% (20) and the lowest of 16% (10). This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 7.017 and 17.78, and standard deviation of 4.34. The Median is 10.0 with Average Absolute Deviation from Median = 2.40 Therefore the obtained evaluation confirms the suitability of the used sample for the evaluation of the real-time information of the site.

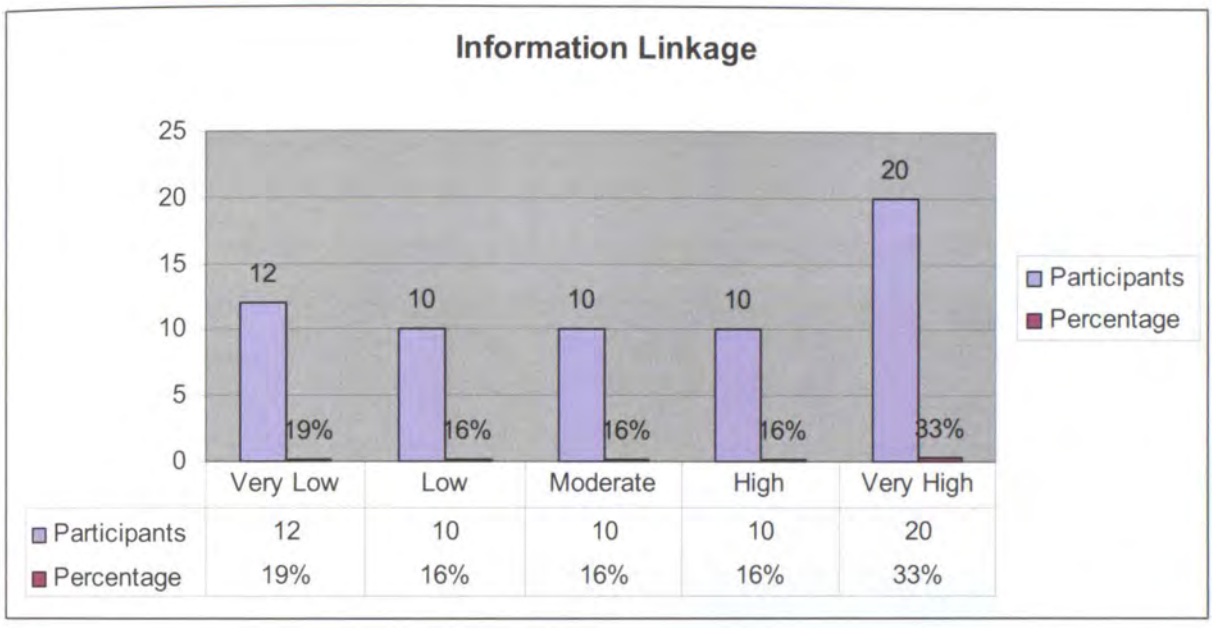


Figure 4.14 Information linkage responses

The results of information linkage show that 33% (20) of the participant found the information linkage is very high, 19% (12) is high, 16% (10) is moderate, 16% (10) is Low and 16% (10) is very low which the max score of 33% (20) and the lowest of 16% (10). This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 7.017 and 17.78, and standard deviation of 4.34. The Median is 10.0 with Average Absolute Deviation from Median = 2.40 Therefore the obtained evaluation confirm the suitability of the used sample for the evaluation of the information linkage of the site.

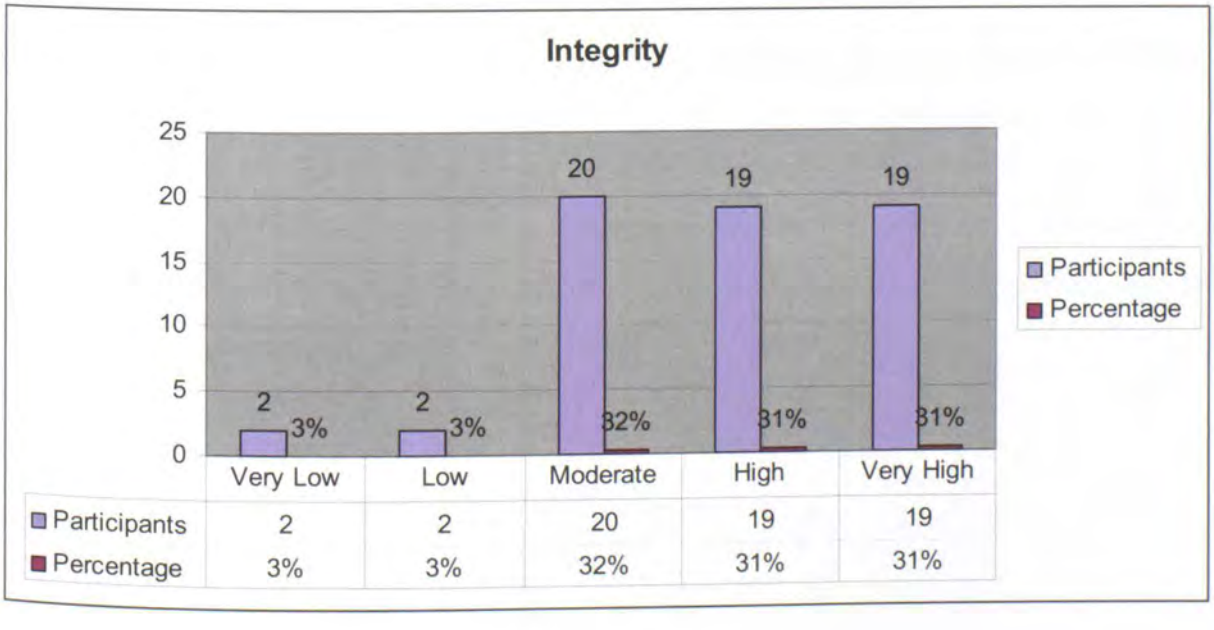


Figure 4.15 Integrity responses

The results of integrity show that 31% (19) of the participant found the integrity is very high, 31% (19) is high, 32% (20) is moderate, 3% (2) is Low and 3% (2) is very low which the max score of 32% (20) and the lowest of 3% (2). This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 0.6028 and 24.20, and standard deviation of 9.50. The Median is 19.0 with Average Absolute Deviation from Median = 7.00 Therefore the obtained evaluation confirms the suitability of the used sample for the evaluation of the integrity of the site.

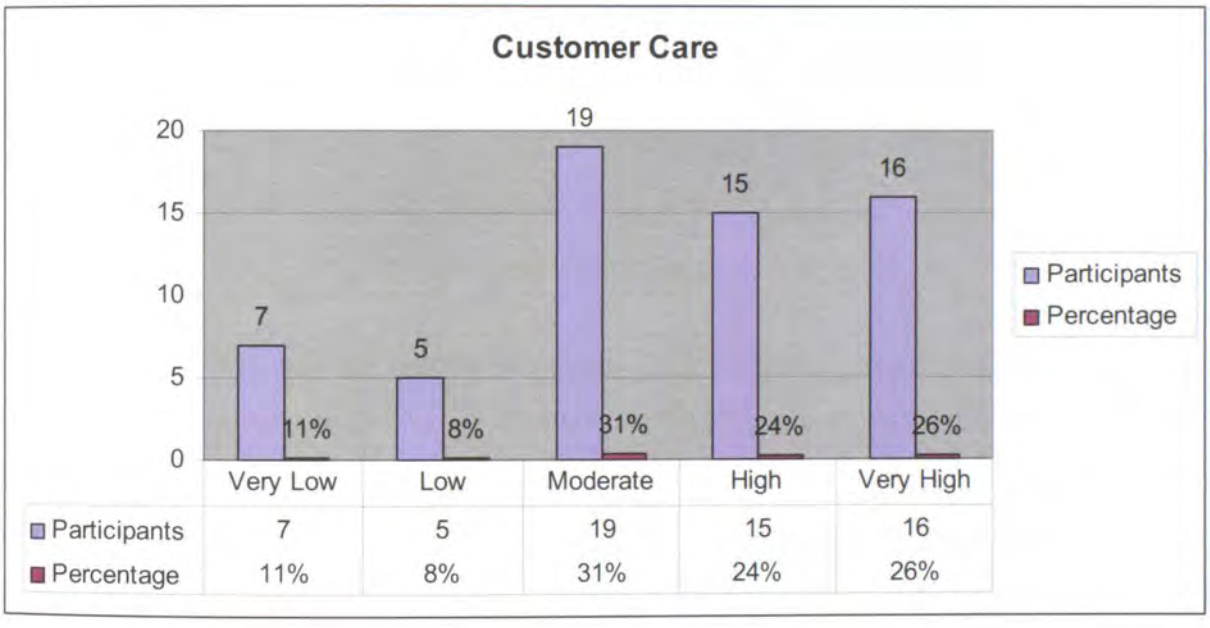


Figure 4.16 Customer care responses

The results of customer care show that 26% (16) of the participant found the customer care is very high, 24% (15) is high, 31% (19) is moderate, 8% (5) is Low and 11% (7) is very low which the max score of 31% (19) and the lowest of 8% (5). This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 4.869 and 19.93, and standard deviation of 6.07. The Median is 15.0 with Average Absolute Deviation from Median = 4.60 Therefore the obtained evaluation confirm the suitability of the used sample for the evaluation of the customer care of the site.

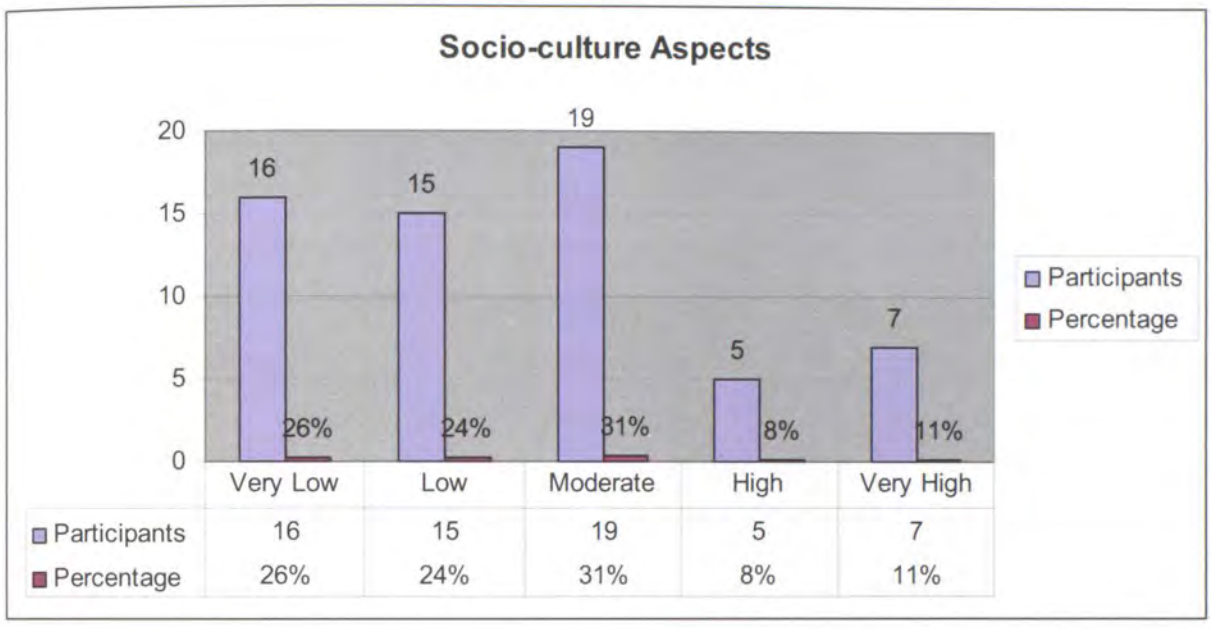


Figure 4.17 Socio-cultural aspects responses

The results of socio-cultural aspects show that 11% (7) of the participant found the socio-cultural aspects is very high, 8% (5) is high, 31% (19) is moderate, 24% (15) is Low and 26% (16) is very low which the max score of 31% (19) and the lowest of 8% (5).This shows that a mean of 12.4 with 95% confidence interval for actual Mean being between 4.869 and 19.93, and standard deviation of 6.07. The Median is 15.0 with Average Absolute Deviation from Median = 4.60 Therefore the obtained evaluation confirm the suitability of the used sample for the evaluation of the socio-cultural aspects of the site.

Additional results from the evaluation are presented in figures 4.18-4.25. Responses to the parameter "Interactivity/functionality" show a reasonable degree of variation in respondents' interaction on the given web pages, in term of interactivity/functionality RAC at 25% and Tesco at 24% have the highest scores and all others shared 17% each (figure 4.18). Usability has been influenced by a number of factors, including visual effects, aspects of GUI, ease of use etc. which have a direct impact on the usability score, with Tesco having the highest score of 30%, RAC and Pixmania scoring 21% and Ambest and e2save each scoring 14% (figure 4.19). The responses on correctness suggest that the customers are relatively concerned with the terminating pages and the ability of the system to correct itself in its aim of enhancing the E-commerce, Tesco, RAC, Ambest, e2save and Pixamnia equally shared 20% (figure 4.20). This shows that correctness is almost equal with all the web pages and it is difficult to evaluate in terms of terminating or corrupted pages. Responses to real time

information show that the participants were more interested in timely and accurate information rather than just information on services and products and they responded 24% to RAC, Ambest and e2save at 23% each, Pixmania and Tesco scored 15% (figure 4.21). For information Linkage Tesco scored highest at 24%, e2save and Pixmania scored 23%, Ambest and RAC shared 15% (figure 4.22). Responses on integrity Tesco scored 31%, Pixmania and RAC 23%, e2save 15% and 8% for Ambest (figure 4.23). This show that Tesco received the highest score and is given 31%. Customer care is one of the priorities of the five online businesses however the participants responded with 29% for RAC and Tesco. Ambest, e2save and Pixmania scored 14% (figure 4.24). Socio-cultural aspects is based on many factors such as colour, symbols, nature etc. and is one of the difficult issues to be explained and grasped however the survey participants were given 24% for Tesco. Ambest, e2save were given 23% and the rest of the two online businesses were given 15% (figure 4.25)

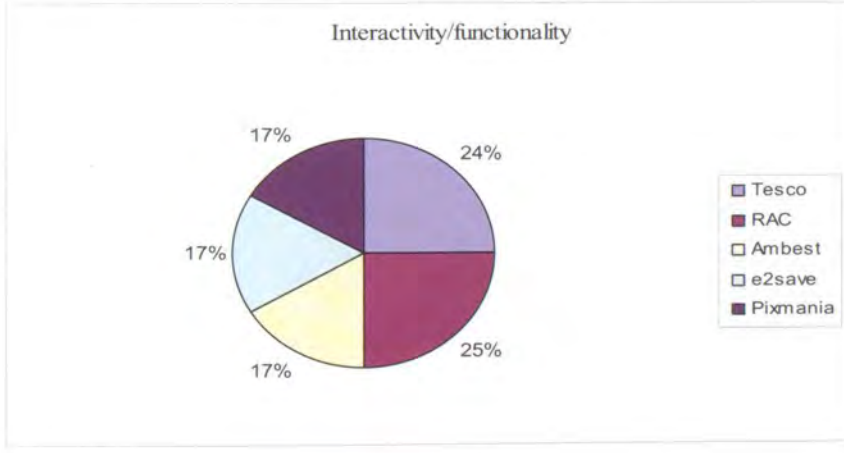


Figure 4.18 Interactivity/functionality

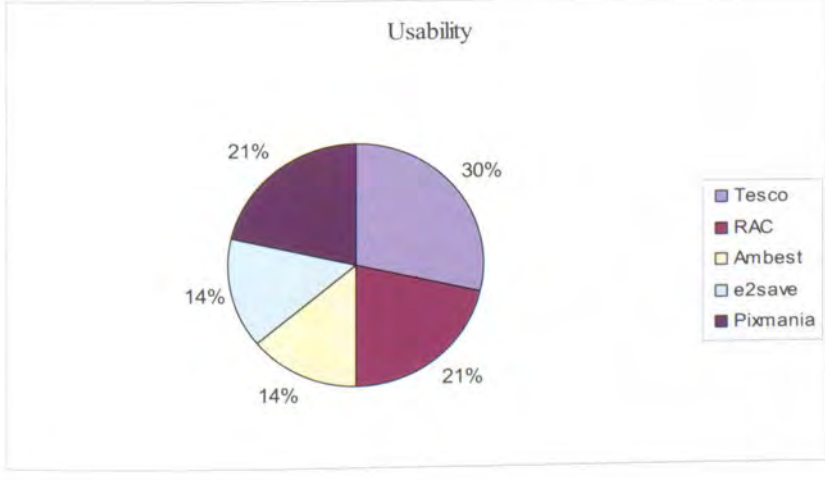


Figure 4.19 Usability

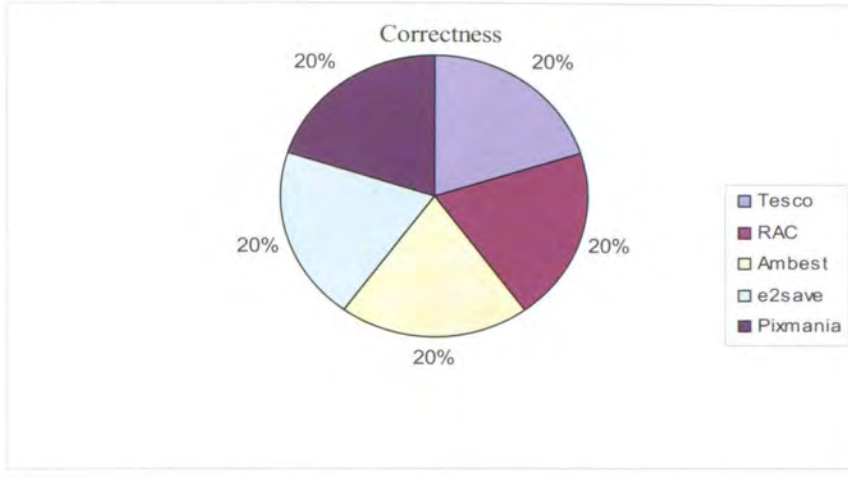


Figure 4.20 Correctness

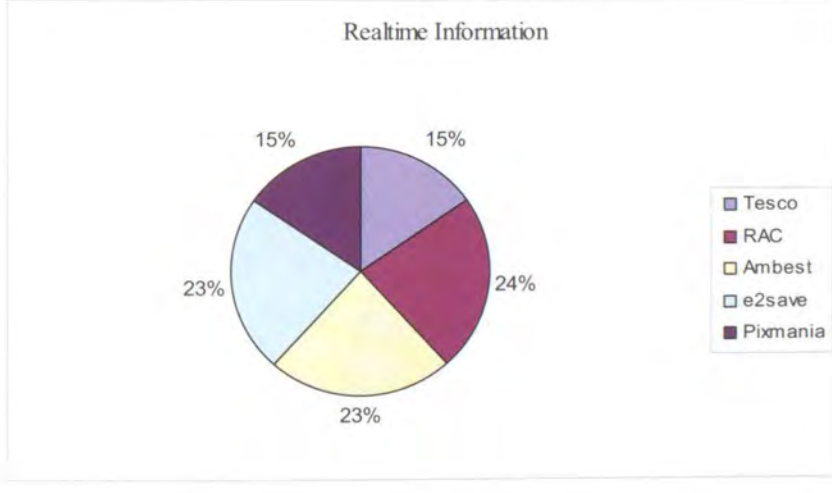


Figure 4.21 Real time information

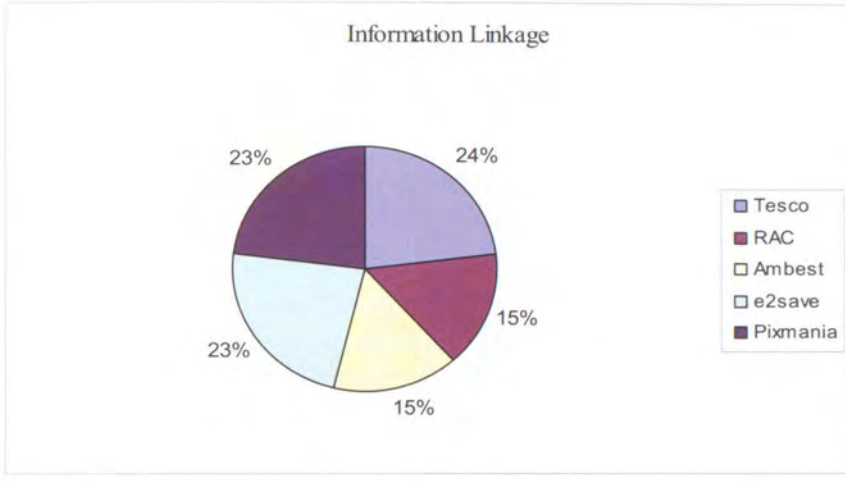


Figure: 4.22 Information Linkage

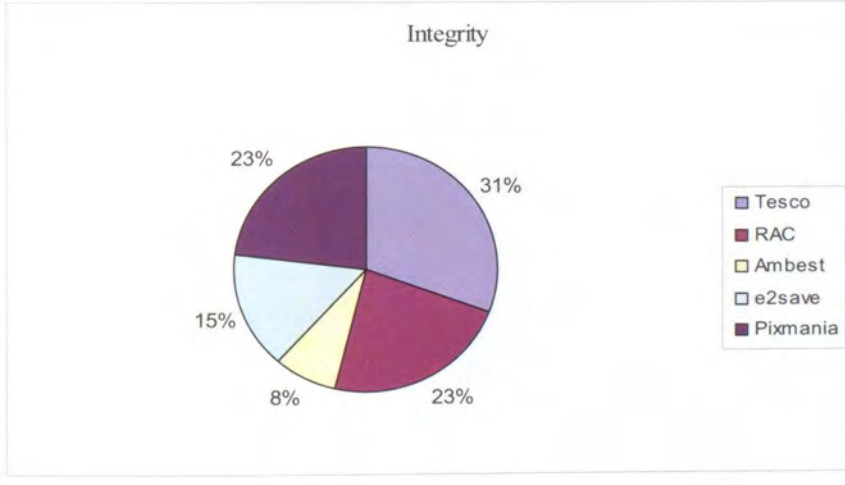


Figure 4.23 Integrity

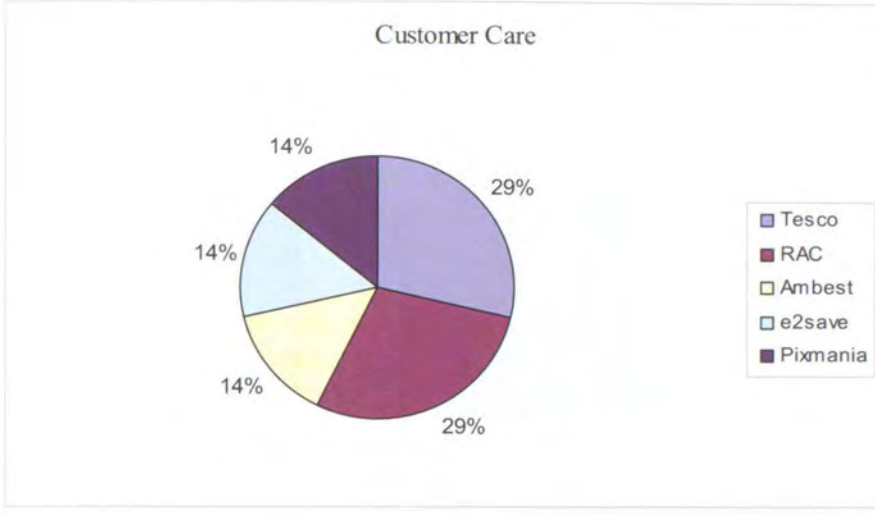


Figure 4.24 Customer care

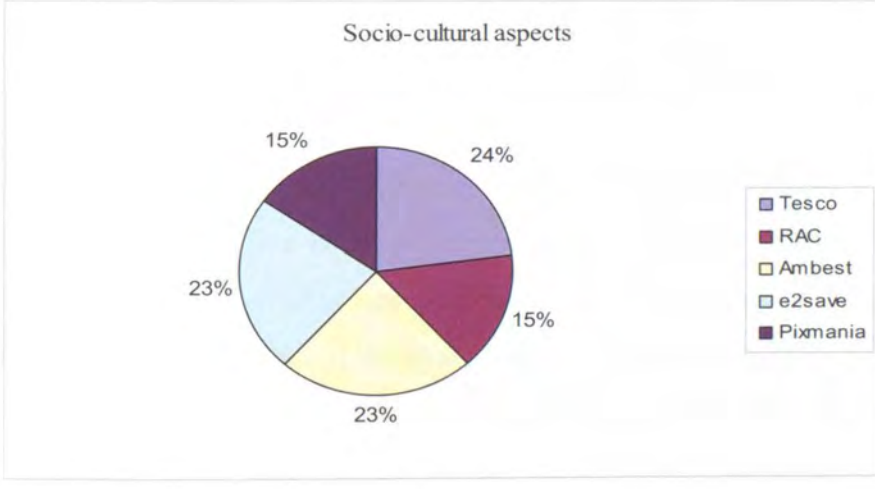


Figure 4.25 Socio cultural aspects

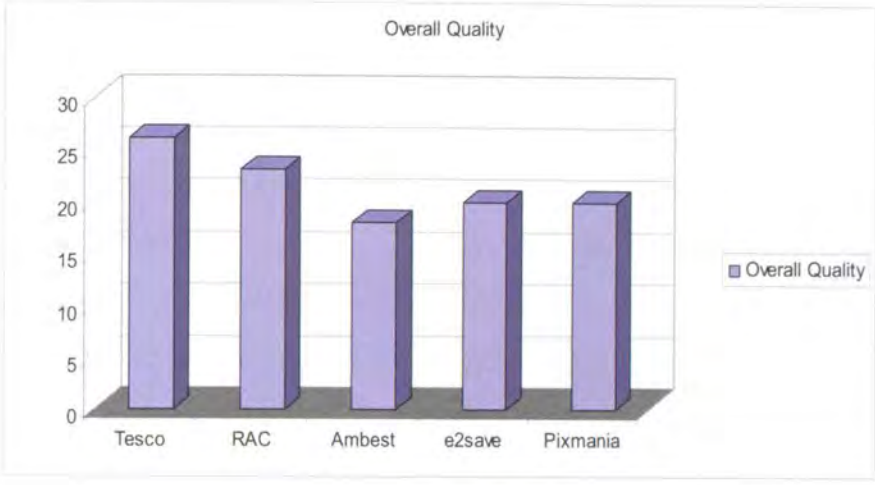


Figure 4.26 Overall Quality

To give an overall view of the quality of considered E-commerce websites, the quality factors have been combined for all sites as shown in figure 4.26 which clearly show that Tesco has achieved the highest overall score.

4.4 Semantic Structure and Semantic Items on E-commerce Sites

Having considered the impact of individual quality factors on the overall quality of websites, this section aims to consider the link between the overall quality of services, particularly in semantic term, and the ways websites are structured. The semantic structure of an E-commerce site may be defined in term of the availability of website's items/products that are meaningfully organised to carry out online business, for a specified time under specified circumstances. It should reflect the dynamic of the market, therefore it should take into account the probability that items/products will change frequently over time due to changes in the market and fashion trend, and the possibility of selling goods in a specified market and in specified circumstances.

Analysing the structure of an E-commerce site is necessary to find if the products in the site are acceptable for qualitative and meaningful business. An E-commerce site has to be in tune with real time data and real time products and provides an estimate of the number of meaningful items that the E-commerce site can offer to the customers. Evaluating the structure therefore can help in defining the quality of E-commerce site.

Earlier work on the E-commerce structure attempted to extend the quality of websites to maximise their business opportunities (chu07, Xia07). Much of the quality improvement of

existing websites however, is focused on to the graphical presentation of the site, in fact website structure is governed the navigational interface (Lyn02). Navigation aims to take the user to different areas of a site and may differ for a number of reasons including the business content, arrangement of the content and the organisation. Many structures are used by designers in order to make an E-commerce site more efficient and reliable. For example, sites that organise information in sequences can enhance the reliability and consistency of the content. The sequence technique is used to structure the site in many ways, for example, managing the content by logical progress for example from the general content to a specific one (Lyn 04).

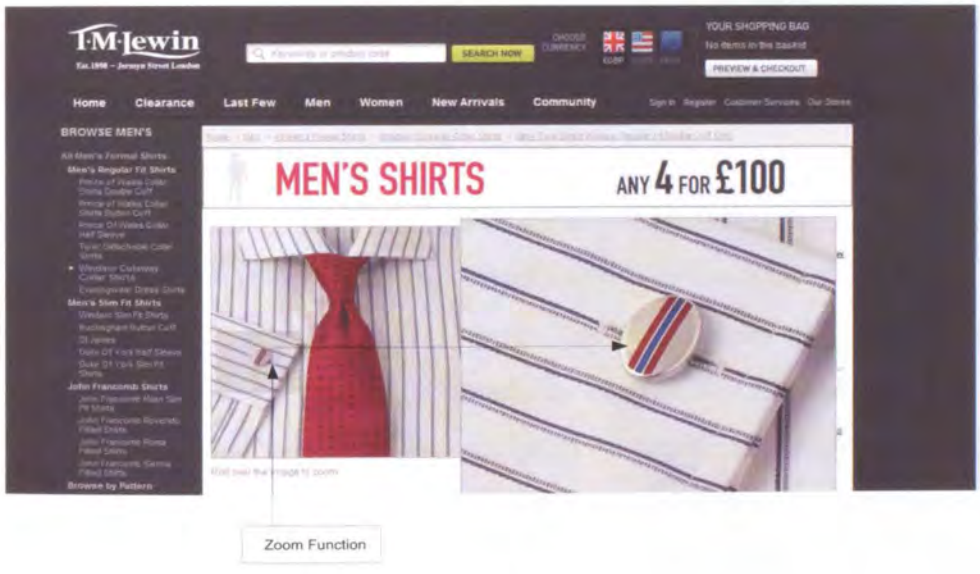


Figure 4.27 Zoom function

If it can be assured that a sequence structure has been effective, and that the reality selling can be achieved, then real-time information growth can make the site more reliable. A number of techniques exist which attempt to make the sequence structure more effective (Jus03). These techniques are used to impress on the customers that they are buying real time products in a real time market. Typically, the techniques are image zooming functions that describe and explain the products (figure 4.27); numerous zooming in and out functions are used in the E-commerce market. They can be used to capture an item from the site and they also enable the customer to move from one general area to the specific area (figure 4.28).

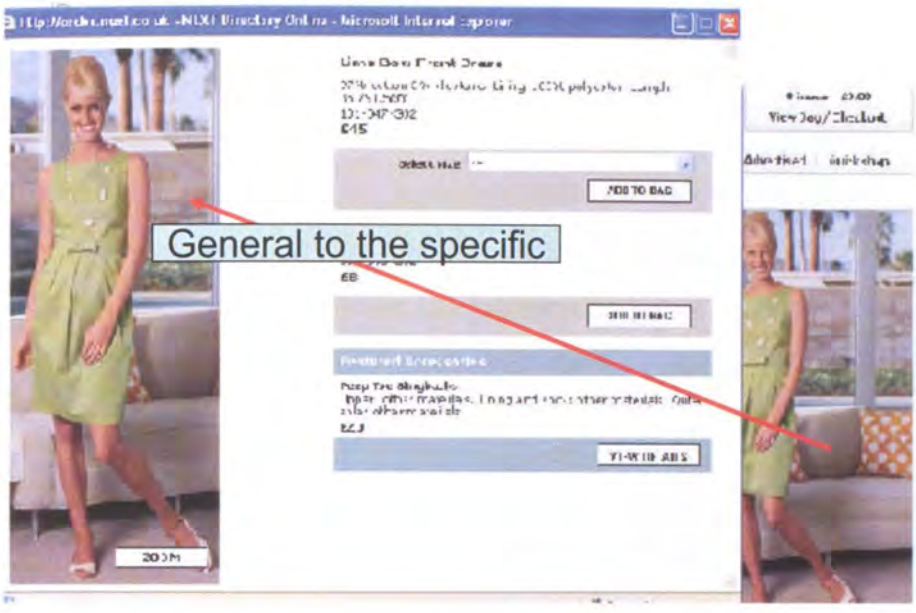


Figure 4.28 General to the specific

In this setting an important area of concern is the use of attributes in E-commerce. The use of attributes on an E-commerce site page determines the communication flow and information accuracy. For example, navigation has many attributes such as top, bottom, left, right or content area navigation. The navigation directs the users from different angles. It depends on the preference of the user as to how the navigation locates and retrieves information. If the user wants to acquire new information, the user has to navigate forward whereas backward navigation will enable the user to regain information obtained previously. The relationship between the attributes and characteristics of an E-commerce site makes the retrieval of information easier.

4.4.1 Linear Structure

Unfortunately, there is no single sequence technique that can be trusted to give accurate and real time product information from the real time business in all circumstances and the fact that the particular technique has been producing real-time information in the past is no assurance that it will do so in the future. As a result, the real-time information is far from reality.

Since websites aim to be unique, designers try their best to achieve that by modelling their structures. A structure represents the method of navigation which decides how the navigator looks at the businesses for example it would be easier for the navigator to find the whole elements of the website in an interdependent and orderly way that means easy navigation from one page to another. An example is the linear structure which means that websites are organized in sequences i.e. to go from one page to another in a particular order. Thus, visitors of a site must follow the web pages in a predefined order (sequence) as shown in figures 4.29 and 4.30.

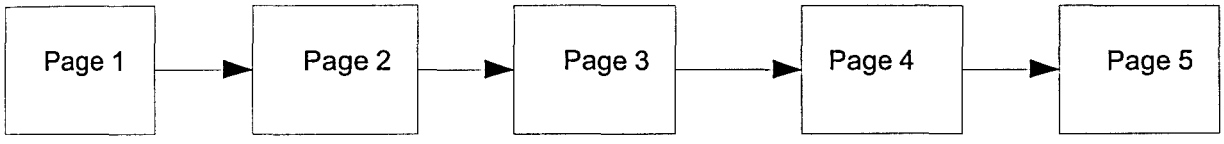


Figure 4:29 Linear structure in sequence

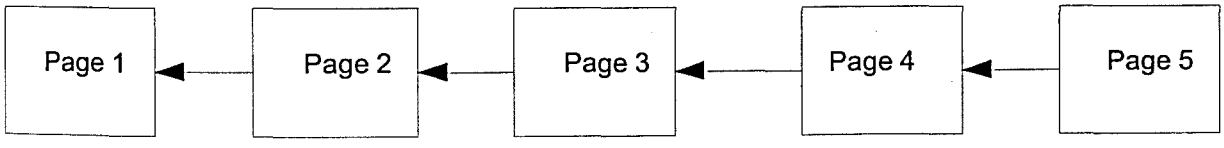


Figure 4:30 Linear structure in a predefined order

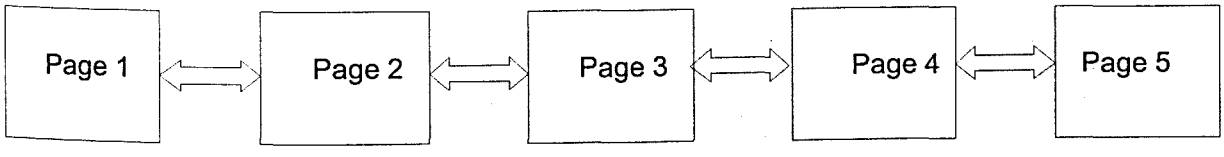


Figure 4:31 Linear reciprocal structure

Moreover, reciprocal navigation as seen in figure 4.31 allows the navigator to move back and forth between the pages for example from a shopping cart check out the customer can go back by using back clicking option. The linear structure is commonly used for photo galleries, shopping check-out etc.

4.4.2 Hierarchical Structure

Another website structure approach is the hierarchy. A hierarchical structure is the most common structure for E-commerce and its purpose is to provide information where a home page exists and other main pages are followed by sub-pages (figure 4.32). Each topic is underlined by a home page and it leads to a headline point, each of which again leads to more detailed information. The detailed information page leads to another level. The top level of the hierarchy explains the main areas of the site. In this hierarchical structure the user enters into the home page and from there goes to the sub-areas. The user may be channelled from the home page to more detailed information (Pau05). The advantages of the hierarchical structure are many. Since it is based on a drop-down site design using categories and subcategories the audience can link to the main areas of interest from within the home page. Therefore this type of design will suit an informational website well. Hierarchical structure is far less limiting than the linear website structure (ado05).

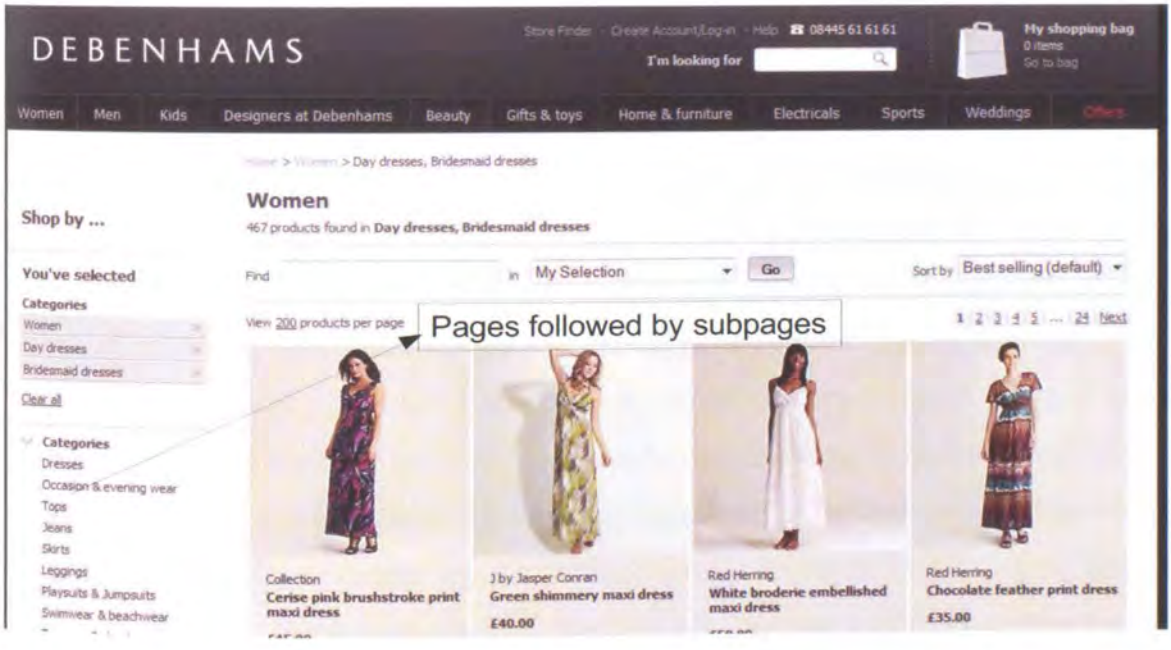


Figure 4.32 Page organisation

It should also be noted that the reliability of a website is often based on the semantic nature of the site and the effectiveness of the structural selection, which allows the information to flow without interruption. The structural formation of the site can assure areas for semantic information which, if it is necessary, may not only improve the real-time information, but

may also reduce the inaccuracy of the real-time product. In fact any structural formations are capable of accepting real-time information.

In summary, the accuracy and reliability of information is dependent, therefore, on the effectiveness of the E-commerce structure chosen. The reliability and the semantic real-time information, however, present a problem in E-commerce. It is always true that sending real-time information on the website can cause problems, but it is not always clear that the accuracy of real-time information on the web is also problematic (Zol05). There exists confusion, too, between sending real-time information as a means of inspiring confidence in the customers, and a means of finding the customers. Furthermore, structural approaches which confuse the customers, in which the websites send information pretending to be real-time and accurate, may be unreliable and inaccurate. It is therefore important to consider structures which can support semantic flow of information that take into account the quality parameters discussed earlier.

4.5 Semantic items evaluation in E-commerce systems

The semantic items/products/services on the web may be defined as meaningful ones. The evaluation an E-commerce site for its semanticity requires a thorough analysis of its products or services (Jon07, Mau04, Car99). For example, to find a meaningful item over the net the customer can send few queries about the product and gather all the necessary information including its origin, make, description etc. However, in order to have a meaning a product must have a relationship to other products, services etc. P. Bouquet et.al (Bou04) describes additional parameters which were considered most important for semantic querying such as ontological distance which encodes the ontological effort that determines the semantic relation between two concepts, lexical distance which determine if two words denote the same concept, type of relation which is the coordination rule which represents a semantic relation between two complex concepts. Therefore, there are many factors associated with the measurement of the semantic items on the net, (Ant05) such as querying information on the item, viewing, and affiliation of the item (Law98).

In the E-commerce system, adding or updating any item might involve altering or modifying the E-commerce website. Typically, the change takes place whenever new services/items come to the market. However, the change might be exploited to give the impression of semantic items on the market. Semantic items may be defined as the meaningful items or

original items of a website which can help in improving the accuracy of information or other meaningful services, (Ore05) or to adapt the items to changed circumstances. Items depend, therefore, on the ease with which a website system can be updated when deficiencies of information occur, and can be improved or to provide real-time information to the customer (Ore05).

However, items/services updates take place when new items/services occur, defects are encountered with the existing items/services, new items are hard to find, or the site cannot communicate with the database. Updating items is often viewed as a difficult task and a costly time consuming event, because it is often more demanding than any other E-commerce work. This is due to the number of questions placed on the E-commerce manager and the site developing team such as: How are the new items/services to be inserted into the existing system? Where are they going to be placed? What kind of modifications may happen? Does the existing code need to be changed or replaced? Even good and well maintained E-commerce site sometimes might not be able to place new items/services in the right place.

4.5.1 Items/Products and Web Structures

In examining the semantic items on E-commerce sites, it is interesting to note the types of items/services typically appearing on each page, where they are located, and whether the visibility of information is accurate. It is consideration of these issues that the web designers must take into account when designing an e-commerce (Ore05). As explained in the previous section, common approach is to arrange items/services in hierarchical fashion, where customers navigate by clicking down, up, across the hierarchy, from a given item in order to locate the real item/s and/or services. For example, in the hierarchy illustrated in figure 4.33, the items dress shirts, casual shirts; short shirts can be described as the children of item 'shirts,' and shirts is the child of clothes. In this figure the first level of visual expression is the 'clothes', the second level is 'shirts, jeans' and the third level of visual expression occurs as dress shirts, casual shirts, short shirts, fashion denim, flare and boot cut.

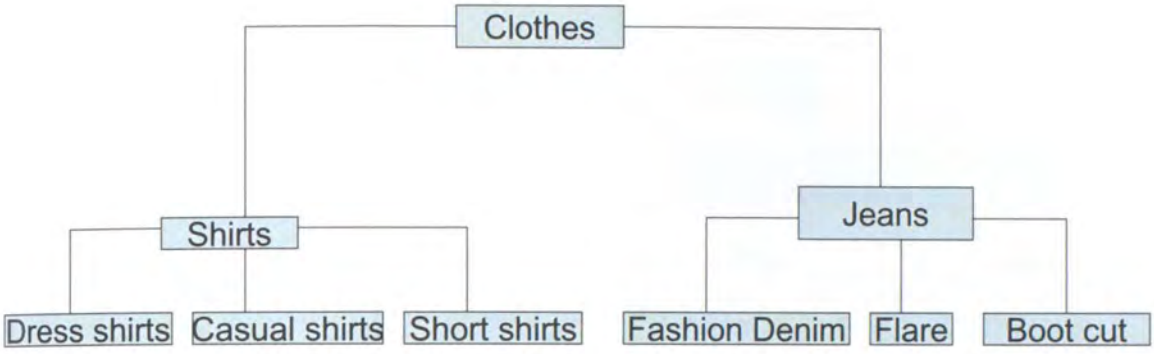


Figure 4.33 A simple hierarchical visual expression

This approach is followed by many E-Commerce websites for example GAP site where from any page customers can reach the first level of 'clothes categories,' such as men, women, gapkids, babygaps, gapmaternity, gapbody (figure 4.34)

Figure 4.34 Reaching the class of items

The sub-class of items can be reached through the first level of categories. This is due to the reason that the first level (class) of items must lead to the second level, where the items are visually represented. The link must take the user to the particular class of items in the same hierarchy. A typical approach is the first level presents the sub-class and the sub class leads down to the children.

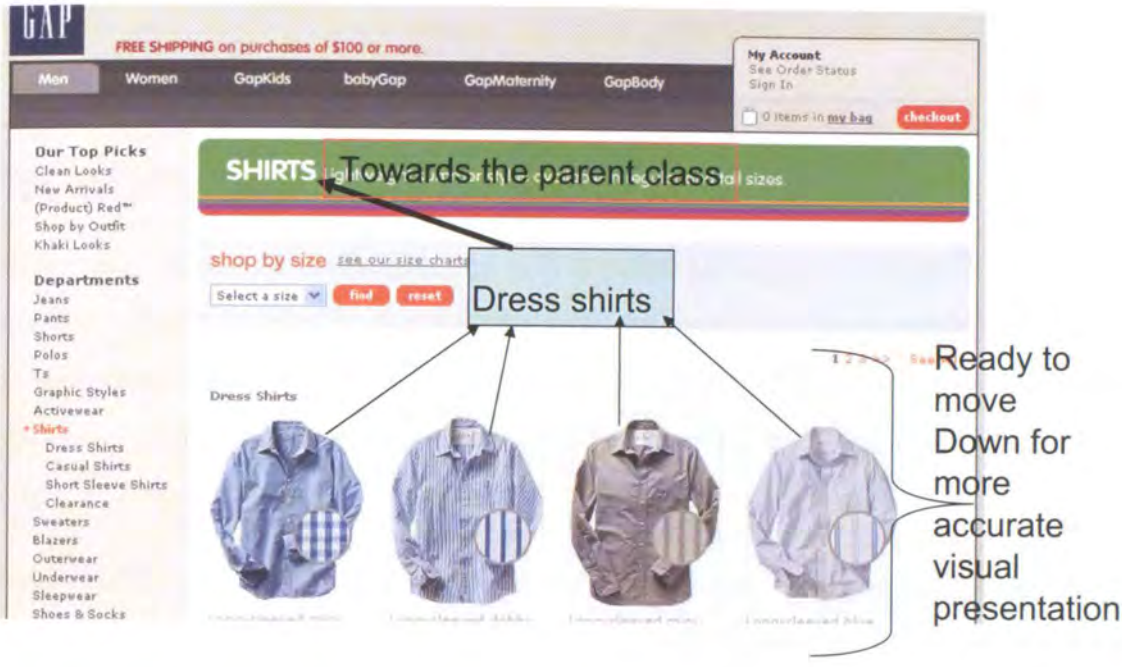


Figure 4.35 Visual representations

When the first level of representation leads to the second level, the second level usually moves down for more accurate visual representation (figure 4.35). The complexity of visual representation of each item tends to increase in each level over time, unless specific attention is taken to maintain the relationship between the levels of visual representation. The larger the levels of representation, the greater will be the probability of items' meaningful representation on the website. The greater the items' meaningful representation, the lesser will be the probability of relationship between the levels of representation. This problem exists because of the inaccurate, outdated, inconsistent item information and the non-related item coding might occur in the second or the third level of the site. Work in updating the items by maintaining the site can also contribute drawbacks in keeping the relationship. This misconception also arises partly due to ignorance of building up and maintaining an E-commerce site. Moreover, there is a misconception that updating the information on the site is easier and therefore the web developers are often tempted to avoid the relationship between each level. This can sometimes affect the structure of the site as well as the consistency of information on the pages, thus making the site more complex and difficult. The type of E-commerce may also strongly influence the levels of representation of items (Xue04, Ore05). For example, an E-commerce site in the clothing business will be tempted to capture the market by working closely with customers to express their items completely and correctly by defining the products through various levels. When there is tight competition in the market,

the E-commerce site developers get tempted to build more complex site pages with more unique features. The sites are often checked to add new features in each level.

4.5.2 Items Evaluation

The quality model presented earlier (section 4.2) contains some generic factors, for each of these factors a number of specific characteristics can be identified, for example usability or real time information can be identified with characteristics such as visual expression. Other factors such as item interactivity/functionality can be identified with item linkage and item behaviour.

The semantic expression of the class of items- visual expression, locating the product, item linkage, item behaviour are easily described in the same way at either the first or second level of representation. In order to assess the presentation levels for their visual expression, item linkage and item behaviour it is important to consider a number companies in the same sector, so an appropriate comparison can be made. Thus, a number of experiments were carried out on five major companies involved in E-commerce and working in the same sector, retail clothing, such as Next, Gap, Debenhams, M&S and T M Lewin. The companies have been evaluated using the quality characteristics discussed above and the findings, which are discussed below, are summarized in figures 4.27, 4.28, 4.29, 4.30.

Visual expression: At the first level of presentation of items, sites provided low visual expression, which was illustrated through the common use of labels to help the customers reach the 'hierarchical categories'. At the second level, most of the sites provided links to the third level again with moderate visual expression at this level. At the third level, most of the sites provided a very high visual expression, and most of these sites reached the final stage of visual expression at this level. However, the remaining sites were led again to very high expression at the fourth level with most of the sites terminating at this level.

Item Linkage: Most of the sites show no item linkage (very low) at the first level, and at the later levels the linkage is shown through images. At the second level, a number of sites are given item linkage but it is considerably low; this means that most of the sites are either bypassing the second level to the next level (third or fourth) or item linkage is not shown at this level. At the third and fourth levels item linkage is relatively moderate. The example of item linkage is shown in figure 4.40.

Item behaviour: Sites show very low item behaviour at the first and second levels. At the third level item behaviour is high and it is very high at the fourth or final level as shown in figure 4.36.

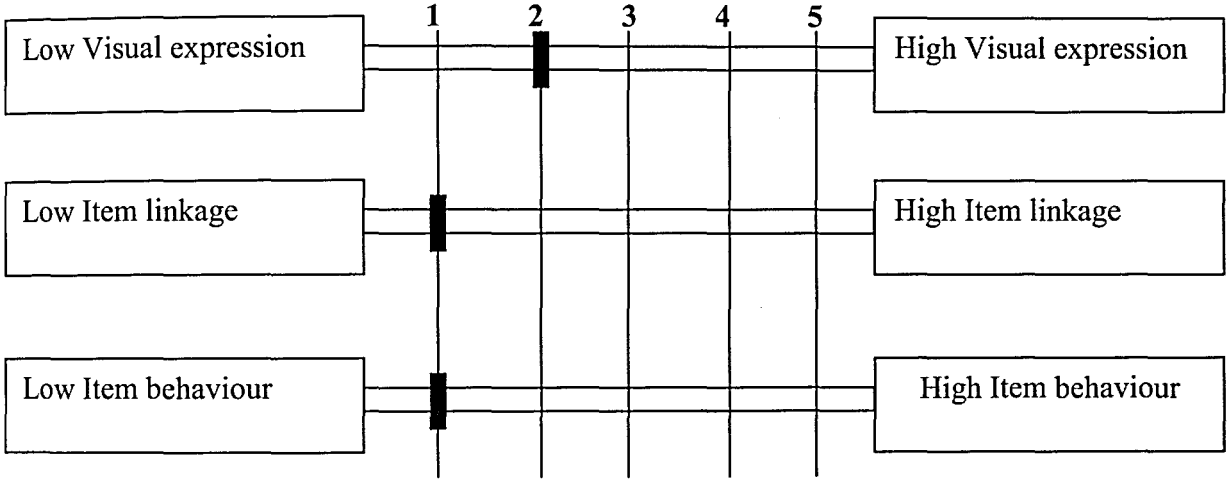


Figure 4.36 Score profile for Level '1'

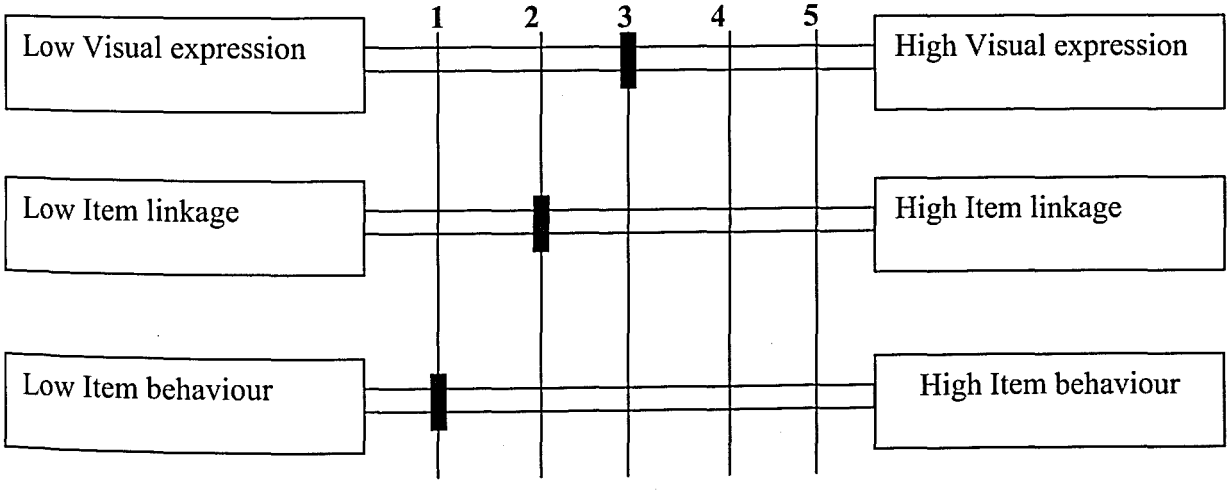


Figure 4.37 Score profile for Level '2'

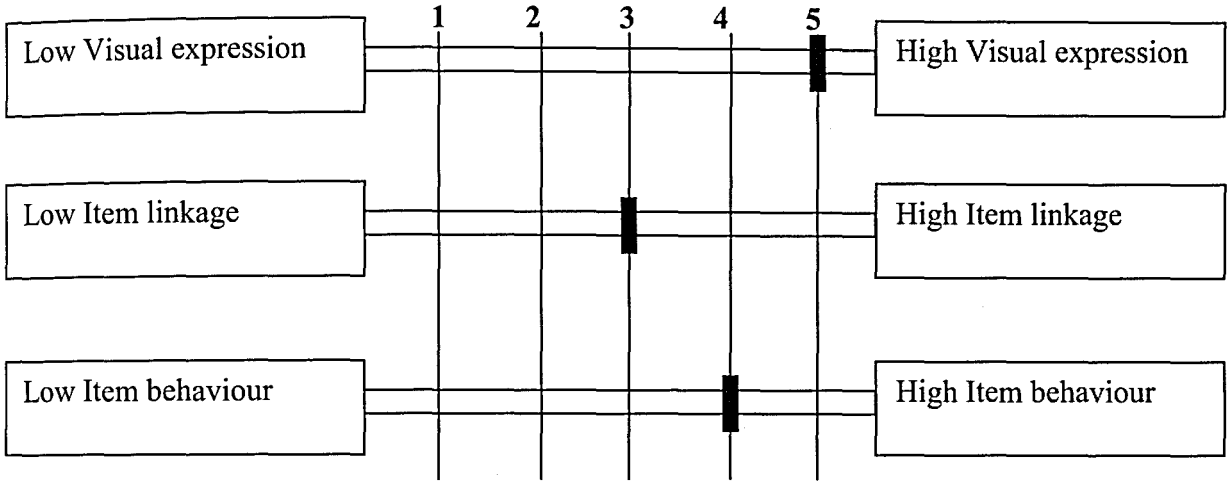


Figure 4.38 Score profile for Level '3'

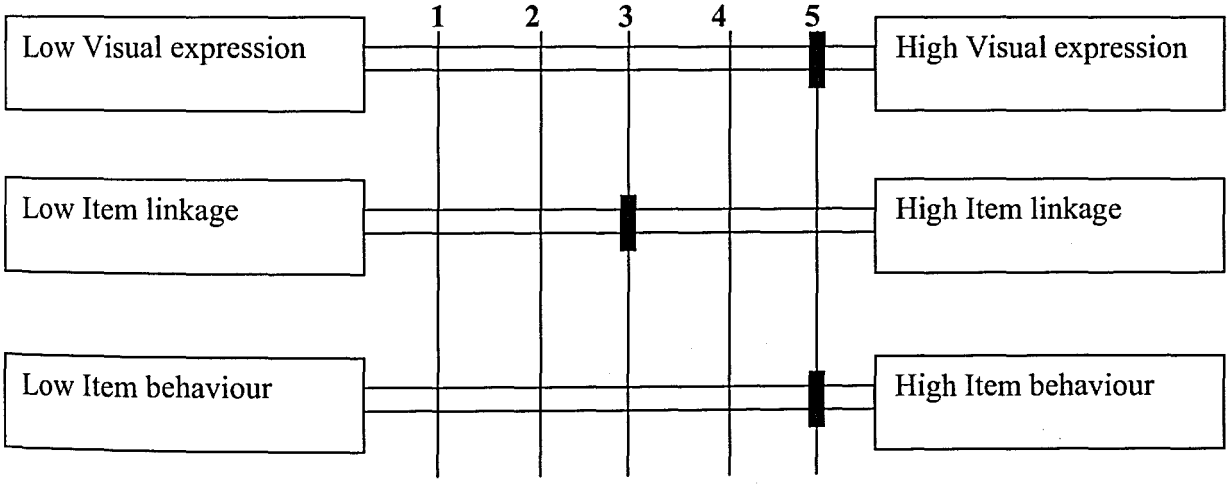


Figure 4.39 Score profile for Level '4'

It should be realised that the items, visual representation is gradually revealed at each level. Most of the E-commerce sites terminate at the third level, therefore denying an ample chance to show item behaviour in the fourth level. Visual expression and the item linkage in the fourth level are closely related to each other, and the item behaviour therefore shows too little variation from the others.

Figure 4.40 item linkage

Figure 4.41 item behaviour

The fact is that the E-commerce sites, which have third, fourth or above levels are capable of showing some sort of item behavior with colour, size etc. as shown in figure 4:41. This is due to the high level of competition in the market. This technique of item behaviour is a new trend in the E-commerce market (Ben05). In fact designing E-commerce sites should

incorporate item behaviour, good design practices and real-time information flow. This should also target and remove the defects of visual expression of items on the site, remove the deficiency of real time information, while maximising customer satisfaction.

The use of techniques and site modelling approaches in E-commerce also forms an important part of any good E-commerce site designing methodology. A recently used technique is the 'item behaviour', which attempts to create real-time items over the net as shown in figure 4.41. Therefore, item behavior can be influenced by many factors, say of a shirt this will be colour, size, price, quantity etc. (Tor06).

The stability of visual expression and the item behaviour can only be effectively achieved where the item really exists which may not always be the case. For sites that need specifications of items that no longer exist, a designing technique should be used to overcome the problem of items specifications. The approach involves placing the images of the item, coding the item with a roll-over function and a fill colour function, selecting the most suitable option set and showing the item with expression from the item chosen. There is no guarantee, however, that the selected item will match the original. Even where a duplicate of the original exists, problems can arise in the selection. Some approaches, therefore, focus on reducing the differences between the original and the duplicate by placing one of the original item's picture on the site and thus creating other functions over the picture. As a result, it is argued that mimics on the E-commerce site are not required. Moreover, keeping the originality of an item may involve substantial re-organisation of an E-commerce site.

4.6 Conclusion

Having considered the quality aspects of common websites, particularly E-commerce sites, and the limitation of some of the existing approaches, this chapter presented a quality model, with a number of quality factors, which was used to evaluate a number of businesses with active E-commerce systems. Customer satisfaction greatly depends on the identified factors, which are expected to perform well when the data is semantic. The model defines some fundamental factors that influence and maintain the quality of websites, therefore impacting their users/customers satisfaction. Indeed, the model evaluation has produced a variety of responses, which showed satisfactions in some areas but not in others; however some overall satisfaction/un-satisfaction trends were identified.

After considering the quality factors at the highest level, the specific characteristics of the some of the factors together with the potential sites structures, which have impact on the quality of websites, were considered. An evaluation using the identified characteristics were conducted which has shown the impact of the structure and layers of websites on some of the quality characteristics. In fact, the semantic nature of the E-commerce site can be improved by configuring a structure, in which the customer can easily navigate from one point to the other (e.g. from the general to the unique information), which can improve site usability. To enhance the real-time information it should be noted that the flow of uninterrupted information is necessary in E-commerce and that there is no reason to separate semantic information from real-time information. Semantic information is meaningful information and it is difficult to retrieve without appropriate information linkage. A good design approach, effective communication model, integration of real time item information and real time visual expression, can all contribute towards creating and modeling an effective semantic E-commerce system which can help in addressing many of the quality issues identified in this chapter.

Chapter 5 - Defining a New Semantic Framework

5.1 Introduction

The aim of semantic real time information searching and discovery is to solve the information matching problem by defining real-time information search and discovery semantically and thereby providing a potential solution of the interoperability problems faced by many websites and web services. Therefore, it provides an opportunity to develop and maintain real-time semantic information websites, real-time semantic information search, self-organising and hyper linking architectures. This new approach uses semantic web as its theoretical model and defines its own frame work and architecture for real-time semantic information search and automatically self-organising semantic web.

The semantic approach will enable the automatic integration of information from the semantic web and Internet. Thus it will form a self-organising system with hyper linking information network. The goal of the new approach is to provide significant advantages to information searchers and in the broader sense to semantic web. This is achieved by way of providing greater, efficient and real-time access to semantic information, and the development and incorporation of more semantic information over the web and the Internet. Moreover, the realisation of a new semantic approach should take into account incorporation the the following components:

Self-organising: The effective and efficient integration of semantic information in the new approach, requires that the links and information on the web pages (content) such as things, people, documents, objects and concepts and their relationships are to be individually identified. The self-organising mechanism of the new approach will be able to do this by identifying the individuality (uniqueness) of a 'product or thing' and its relationships.

Hyper linking: The self organising part of the new approach is to be able to automatically process or integrate semantic information, which means it has to be 'hyper linked' properly. The key features or concepts of the things, must be represented in a simplified way so that machines can process the information easily. Thus it will become 'data' linked with 'defined

data' (meta data). It will explain the interconnection between data, for example the way a piece of information is related to others, this can be done with combined ontologies, in order to link 'data' more accurately and precisely.

Information Network: The linked real-time semantic information will form an 'information network'.

This chapter starts with a discussion of the meaning of the semantic information together with semantic encoding and indexing, which explain and define, using examples, the approaches used in web application techniques particularly in terms of indexing web based objects and information according to the user relationship and the semantic information search. The semantic architecture which can be used for the implementation of the proposed approach is then presented, together with some descriptions of semantic items relationships and processes. This followed by a discussion of a self organising hyper linking system. Finally, the chapter's findings are summarised in the conclusion section.

5.2 Semantic Information

The semantic information can be defined as any information which carries 'Semantic' which means "meaning of meaning or pertaining to meaning", and it has to relate to 'something,' (as explained in chapter 3). The Information is represented by an abstraction or noun which means 'it is about something.' The data must carry news, facts or details of entities, objects etc therefore information must be attached to 'something.' Thus, for the purpose of this study we can define 'semantic information' as 'meaningful information.'

In this respect, the semantic information should present an efficient way of defining, managing and integrating the meaningful data on to the web, based on the current world experiences of things and objects. Therefore, it could be applied to the web by the meaningful integration of 'data' which is linked with meta data (data about data). This semantic information will be able to form the basis of semantic web. However, such an integration of data into the web requires the designing of the web page or web documents with semantic information (meaningful information linking pages). This can be achieved by incorporating semantic languages, such as RDF/RDFS, XML or XML schema (Bro02, sta04, Gom04) into web pages.This will help in encoding semantic information which is extracted from real life objects (Figure 5.1), therefore generating meaningful data over the web that is semantic by nature.

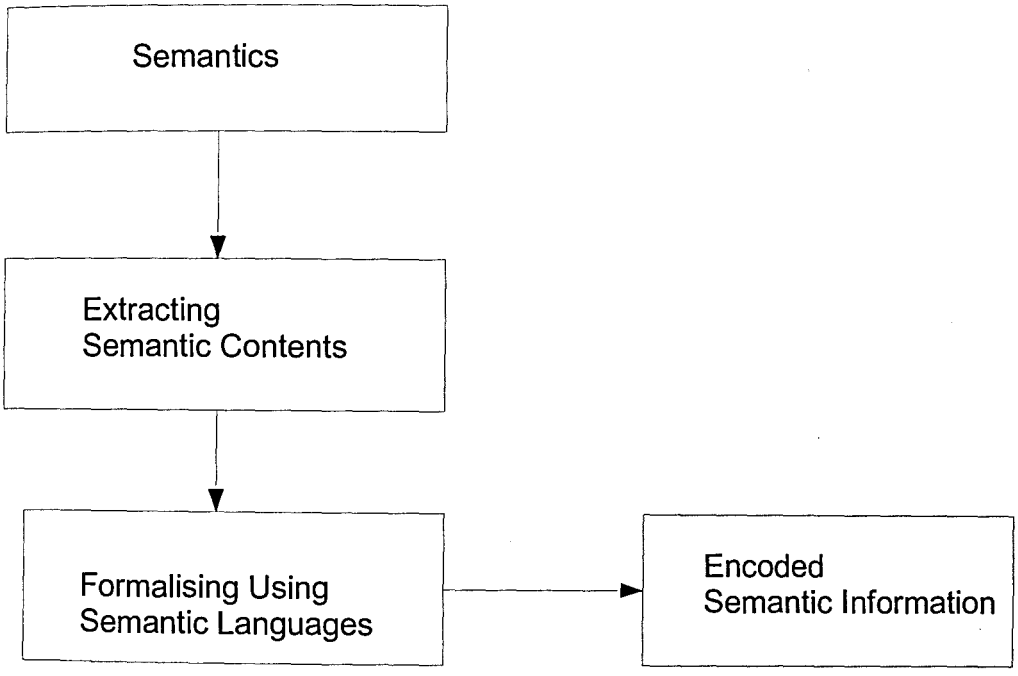


Figure 5.1 Encoding Semantic Information

Semantic web is formed using semantic content, which is described by semantic languages, and ontologies. Semantic information is formed by the semantic content, for example things, people or concepts and their relationships. The meaningful content that can be generated on to the web can form a semantic web.

In fact, semantic information in the real world always exists, for example 'an object' and its semantic (meaningful) content with well established relationships with other objects. However, when it comes to the web, it is not always possible to establish such relationships, because of object representation and the absence of proper use of semantic content on the web and the Internet, therefore W3 consortium has taken the lead to establish semantic (meaningful) information searches over the web (Bal05).

5.2.1 Semantic Content Encoding

In recent years, the amount of information on the web seems to be exponentially growing, but most of the semantic information being discovered or retrieved is ambiguous. This is due to the employed retrieving methods and techniques, which sometimes fail to discover the

semantic information relationships and the process complicated semantic information handling.

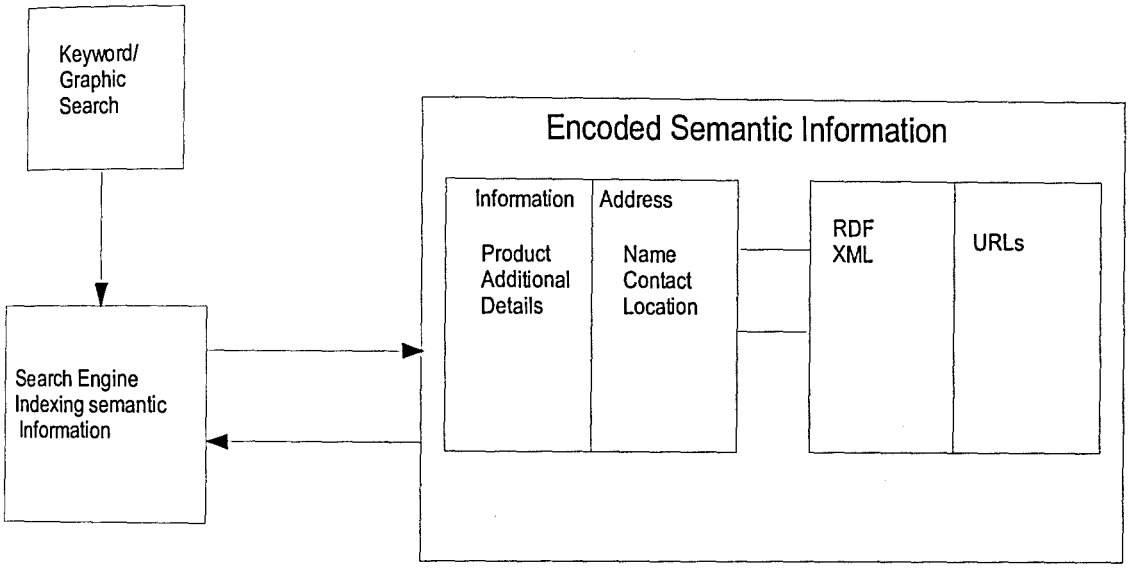


Figure 5.2 Semantic content encoding

The process of encoding semantic information on web pages in meaningful manner requires a search engine with efficient semantic indexing (Figure 5.2). The search engine contains additional information which is relevant to each page, its content and its semantic relationships. Moreover, semantic languages such as RDF and XML can be used to create semantic web content. In the semantic web content there is 'semantic information' which, for example, consists of the product and its additional details, for example 'dietary' requirements or a person's shopping history. This relationship again semantically relates to its more other semantic information/relationships, such as address, name or contact details of the company or person. 'Semantic content' is directed towards the 'URLs' which point towards the 'semantic content' generated by XML, RDF. The search engine can bring up 'semantic information which is contained in each URLs. The client or user can make use of 'keywords' or 'graphic search' to facilitate the search to find semantic information over the semantic web.

Although some of the existing markup languages sometimes failed in helping to retrieve the semantic information, using XML and RDF can help in overcoming the information encoding problems. XML and RDF encoded semantic information can be easily made available for human processing and therefore the computer can easily display the semantic information in an appropriate way. XML documents normally use a simple syntax as shown in table 5.1 which is self-explanatory.

<?xml version="1.0" encoding="ISO-8859-1"?>
<...>
<>.....</>
<>.....</>
<>.....</>
<>.....</.....>
</.....>

Table 5.1 XML document tree

The first line in this encoding document shows XML declaration- the XML version and the character encoding used in this document. For example this document identifies the 1.0 specification of XML and it uses ISO-8859-1 character set. The next line shows the root element of the document and the following four lines describe the child elements of the root. For example:

<fanta>
<to>supplier</to>
<from>fanta</from>
<heading>new arrival</heading>
<body>this is to let you know about our new products</body>
</fanta>

Table 5.2 XML elements

And the last line describes the end of the root element that is the </fanta> (table 5.2). From this example it is easy to assume that the XML document contains information 'from fanta to supplier'. XML encoding must be properly nested and each element must have a closing tag as shown in table 5.3.

```
<root>
 <child>
  <subchild>
  </subchild>
 </child>
</root>
```

Table: 5.3 XML element with a closing tag

XML documents should have a single tag pair to define the root element. All other elements should be nested within the root element. All elements can have sub (children) elements. Sub elements should be in pairs and correctly nested within their parent element (table 5.4)

```
<fanta>

<name> orange </name>

<ingredients>

<substance> orange </substance>

<place> S.America </place>

</ingredients>

</fanta>
```

Table 5.4 Fanta.XML

The following examples show the XML document and the DTD (Document Type Definition):

```
<!ELEMENT fanta (name, ingredients*)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT ingredients (substance, place)>
<!ELEMENT substance (#PCDATA)>
<!ELEMENT place (#PCDATA)>
```

Table 5.5 fanta.dtd

DTD files must contain a list of declarations enclosed with the characters <! and >. There can be four types of declarations, which are distinguished by the following keywords: ELEMENT (element type declaration), ATTLIST (attribute-list declaration), ENTITY (entity declaration), NOTATION (notation declaration). An example of a document structure is presented below:

```
<!ELEMENT fanta (supplier+, ingredients, details?)>
<!ELEMENT supplier (#PCDATA)>
<!ELEMENT ingredients (#PCDATA)>
<!ELEMENT details EMPTY>
<!ATTLIST details lang CDATA #IMPLIED
contents (name | address) "contact">
```

Table 5.6 Fanta element declaration

In the element type declarations, the structure of the individual elements in the document must be described. For example: <!ELEMENT element name (content model)>. Elements in Table 5.6 can be illustrated using a simple Entity Relationship (ER) Diagram as shown in Figure 5.3.

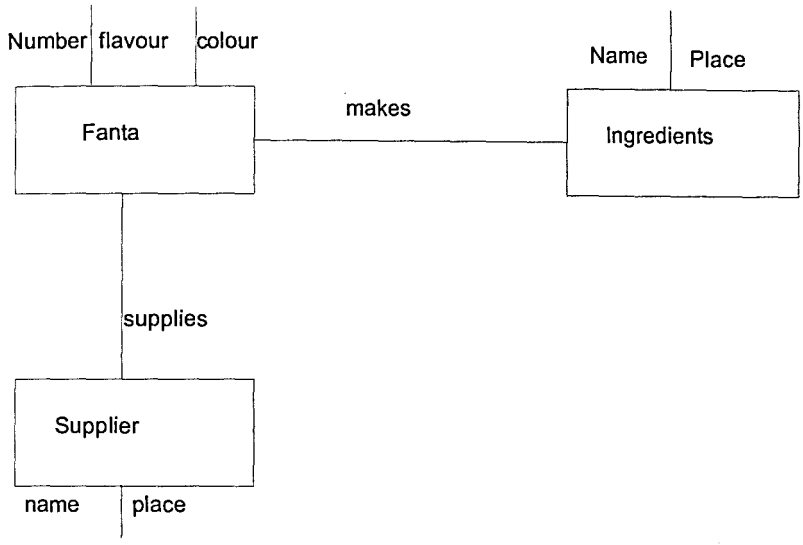


Figure 5.3 Diagram for 'fanta'

The ER diagram shows the 'fanta' relationships with entities, which relates fanta, ingredients and supplier. These three entities are linked by the makes and supplies relationships which correspond with to the following:

<!..The DTD..>

<! ELEMENT FANTA (fanta*, ingredients*, supplier*, makes*, supplies*)

Fanta is the root element of the DTD. The five entities are the sub-elements of the root element. In this example each entity has particular properties (attributes) in the Entity Relationship model (ER). For example ingredients entity is described by ingredient's name and place. Entities with attributes are represented in an XML DTD as:

<! ELEMENT INGREDIENTS (Name, Place)>

The following hierarchy diagram of the FANTA DTD (figure 5.4) illustrates the concept more clearly.

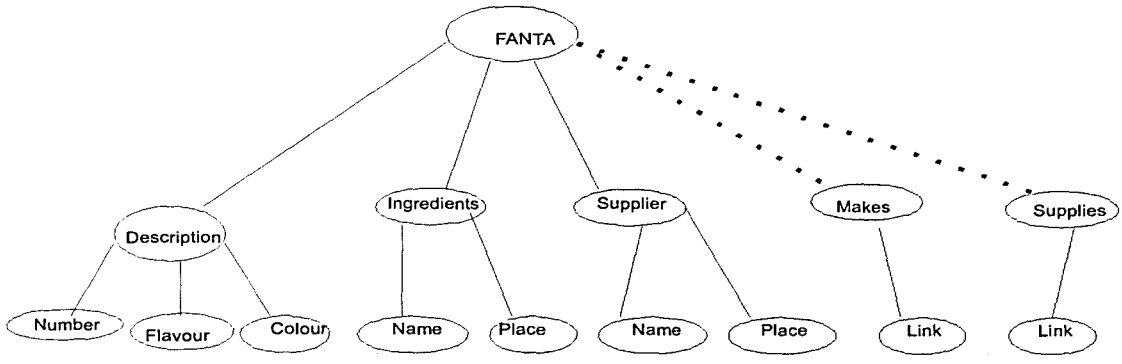


Figure 5.4 The Hierarchy diagram

Figure 5.4 show 'FANTA' as the root element with three nodes or children which is seen in layer 2. The nodes "Description", "Ingredients" and "Supplier" are the entities. The relationships can be presented using 'makes' and 'supplies' are the relationships. The DTD of layer 2 of the 'fanta' can be seen as follows:

```
<!ELEMENT DESCRIPTION (NUMBER, FLAVOUR, COLOUR)>
<!ELEMENT INGRDIENTS (NAME, PLACE)>
<!ELEMENT SUPPLIER (NAME, PLACE)>
<!ELEMENT MAKES (LINK)*>
<!ELEMENT SUPPLIES (LINK)*>
<!ATLIST FANTA type CDATA # IMPLIED>
<!ATLIST INGREDIENTS type CDATA # IMPLIED>
<!ATLIST SUPPLIER type CDATA # IMPLIED>
<!ATLIST MAKES type CDATA # IMPLIED>
<!ATLIST SUPPLIES type CDATA # IMPLIED>
```

Table: 5.7 The DTD of FANTA layer 2

The ER links are represented in table 5.7, such as "makes" and "supplies" which specifies how nodes are linked. The cardinality attribute of the link element records the cardinality ratio of the ER model. The declaration would be: <!ATLIST CARDINALITY CDATA#REQUIRED>

The third layer has different types of nodes, which corresponds to the three entities in the ER diagram. The element declaration of the different types of nodes in layer three can be seen as:

```
<!ELEMENT NUMBER (#PCDATA)>
<!ELEMENT COLOUR (#PCDATA)>
<!ELEMENT FLAVOUR (#PCDATA)>
<!ELEMENT NAME (#PCDATA)>
<!ELEMENT PLACE (#PCDATA)>
```

Table 5.8 Third layer with different types of nodes

5.2.2 Indexing

As the semantic contents and their relationships are rapidly growing, the demand for semantic indexing is also an increasing phenomena. Encoding semantic information, require the development of efficient semantic indexing and searching techniques. This can be achieved by using, the semantic information, to index 'product, additional details and contact information etc.' The semantic content of an identified entity will serve as its unique features.

```
<!DOCTYPE supplier [
 <!ELEMENT supplier EMPTY>
 <!ATTLIST supplier
  supplierno ID #REQUIRED
  Location ID CDATA #REQUIRED
  Location Address CDATA #REQUIRED >
 <!ELEMENT fanta (supplier*)>
]>
<fanta>
 <fanta supplierno="id1001" Location ID="101" LocationAddress="London 101"/>
 <fanta supplierno="id11002" Location ID="102" LocationAddress="Kingston  102"/>
</fanta>
```

Table: 5.9 XML Indexing with element IDs

In table 5.9 every supplier element has a unique ID assigned by the supplierno attribute value. It is an assumption that an XML parser would lift this information to build an index table, which consists of supplier elements and their ID values.

Index Table	Key Value	Node
Supplierno=	"id1001"	<supplier> <supplier supplierno="id1001"./.>
Supplierno=	"id1002"	<supplier supplierno="id11002".. />
.....	.....	..... <supplier/>

Table: 5.10 Index table for IDs in an XML document

This conceptual index table can be accessed by various APIs such as: XmlDocument.getElementbyId() and XPathNavigator.MoveToId() methods, or the XPath id() function.

One of the advantages of this, is that there is no need to define the complete content model for the document in DTD—In order to define the IDs it is only necessary to declare an element and its ID attribute. For example, the above document can be defined in the DTD as:

```
<?xml version="1.0"?>
<!DOCTYPE fanta [
 <!ELEMENT supplier ANY>
 <!ATTLIST supplier
  supplierno ID #REQUIRED
 >
]>
<fanta>
 <supplier supplierno="id1001" LocationID="1001" supplierAddress="London. 101"/>
 <supplier supplierno="id1002" LocationID="1002" supplierAddress="Kingston  102"/>
</fanta>
```

Table 5.11 Document Defined DTD

The path expressions are able to select nodes from the XML document. The following example shows the path expression and their effects.

fantastore: selects all the child nodes of the fantastore element
/fantastore : selectsroot element fantastore (If the path starts with a slash (/) it always represents an absolute path to an element.)
fantastore/fanta :selects all fanta elements' children from fantastore
//fanta: selects all fanta elements from the whole document
fantastore//fanta: selects all fanta elements that are descendant of the fantastore element, from the fantatore element
//@london : selects all attributes that are named london

Table 5.12 Path expression with effects

5.2.3 Indexing Relationships

As in any other document indexing, the approach presented in this work requires the use of one of the indexing methods for indexing and retrieving semantic information (Luy05, Ram06). For example XML Queries are used to navigate data by regular path expressions such as XPath /fanta//supplier[@location="London"]).

The approach is based on hash tree access (Sri04), using per node or per document identifiers, descendant/ancestor search (graphical search), element search by using keyword, incremental update and looking after the size of indexing such as entry number and entry size. The hash tree search is suitable since cross searching as well as parent to children based search is needed. In this approach every path and entity and its relationships are selected for indexing, so that the data guide has to be done with same sequence of labels. Every path in the data base should have only one corresponding path in the data guide. For example:

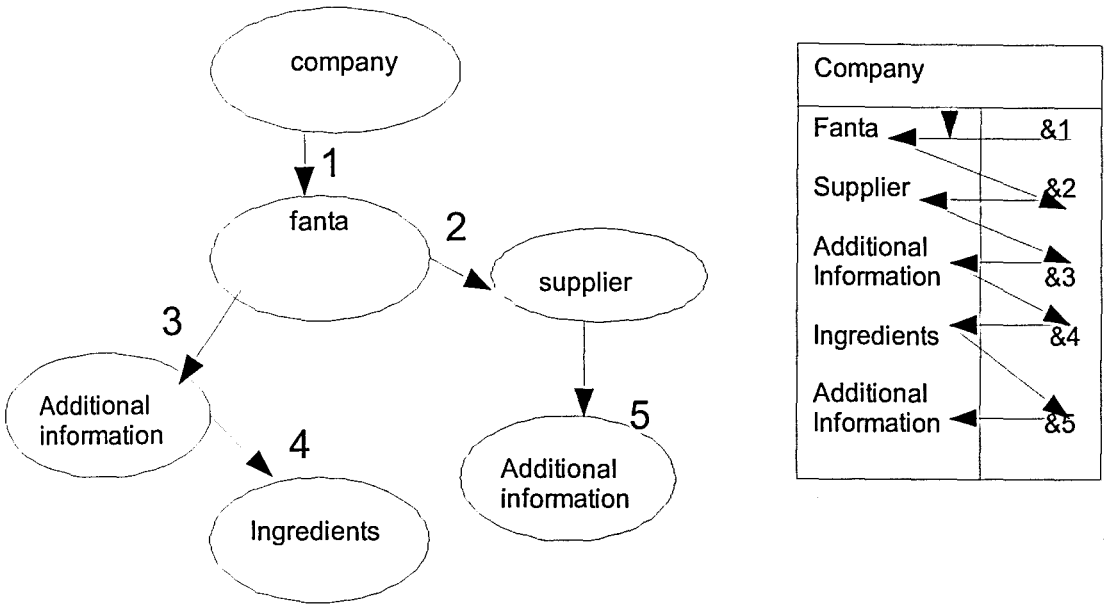


Figure 5.5 Hash Tree Access

Figure 5.5 shows the indexing and search using hash tree access, also shows that each node is identified by an identifying number, and it goes from ancestor to descendant and also allows cross search descendant to ancestor (graphical search). However, keyword search is not necessary at this point and each time the node looks for updates and the insertion is incremental.

5.2.4 Local and Global Indexing Relationships

In order to integrate and establish relationships between entities within a local or a global (on to the web) environment, the first step will be to identify entities and relationships of products services etc. Then, the identified entities and their relationships are classified and labeled. The itemised entities and their relationships can be then stored on a database. In order to facilitate the correspondence between entities, tags can be used. The tagging system can notify of the available products and their relationships. Therefore, the non-related products or services can be easily identified or updated.

The entities relationships can be expressed using Entity Relationship Diagrams and coded using languages such as XML as shown in figures 5.6 and 5.7 for a "Fanta" and its relations/associations.

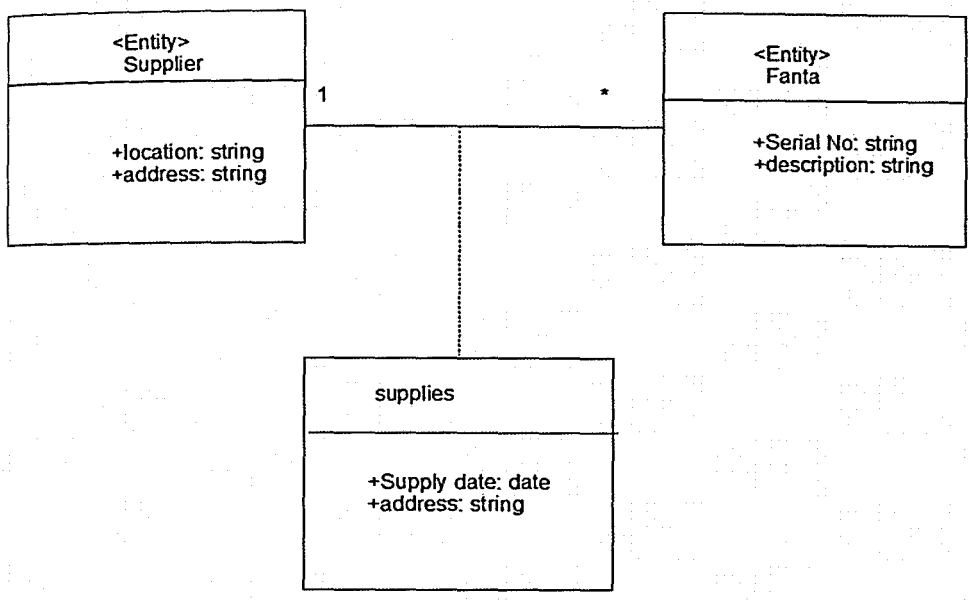


Figure 5.6 Relationship for "Fanta" and "Supplier" Entities

Figure 5.6 shows a simple relationship between entity 'Fanta' and the entity 'supplier' together with the relationship type "supplies" as an association entity, which can be reflected in the database, making it easier to establish entity relationships. Whereas, figure 5.7 shows that the ancestor ('Fanta') is relating and establishing relationships with other children (nodes) such as 'ingredients', 'supplier' and 'additional information'.

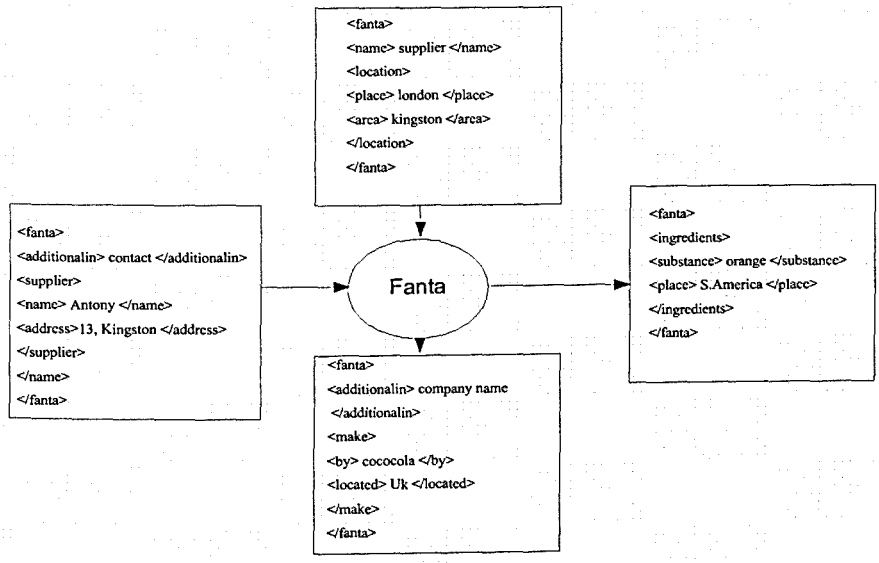


Figure 5.7 Relationships between the ancestor and other nodes

However, when entities and their information are on the web (global environment) encoding of information from web pages will be required, so that the system will be ready to index the pages. This indexing will take place according to the entity relationships. After the indexing

of pages, the system will be capable of automatically discovering new information semantically on the web. Discovering, encoding and integrating the content according to the relationship of information provide the real-time semantic which should be stored in a database. These processes allow data to be defined clearly from the web (semantic information), so that the sources of information can be understood easily.

The new information and content builds up the relationships between objects and things, the discovery of new information from the web is performed by the meta data, which includes meaningful information with its relationships. This can be done using XML schema or RDF or RDFS as basis of semantic webs.

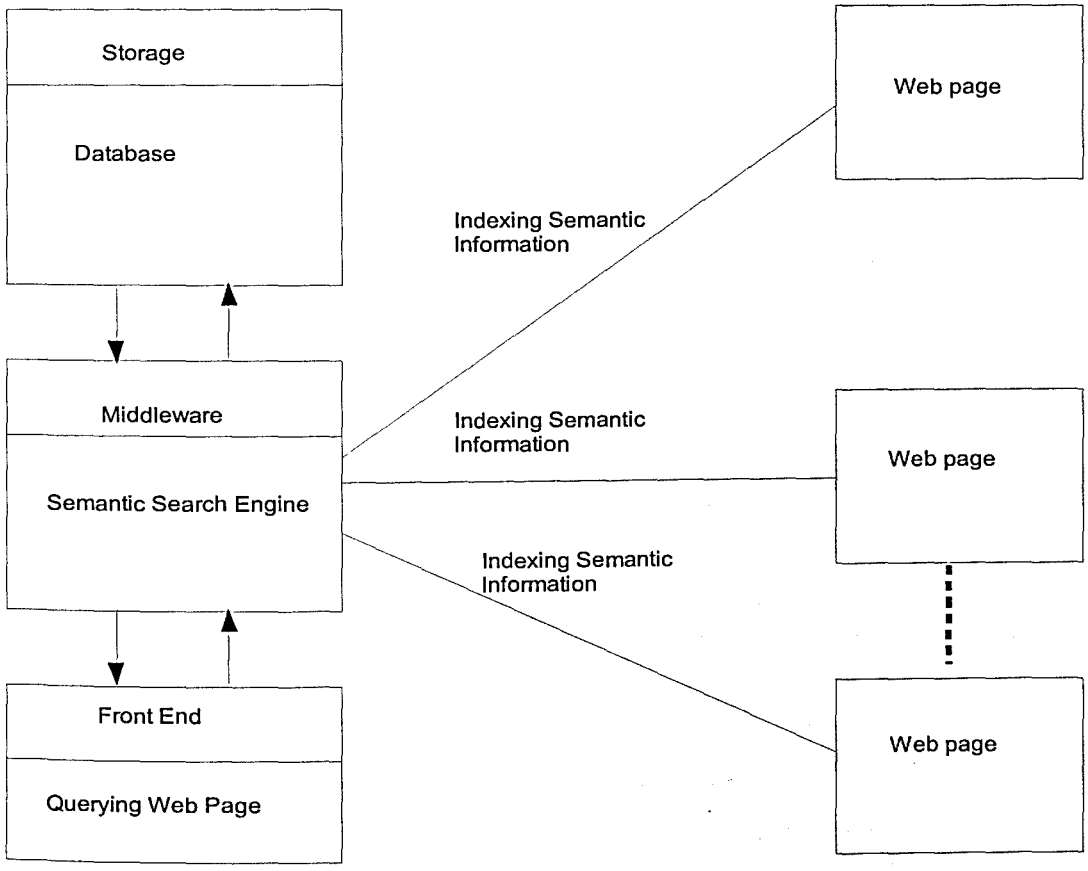


Figure 5.8 Querying and indexing semantic information

Figure 5.8 shows a web page querying the database where the database index content constituted from various web pages according to context and meaning and the links to web page. Figure 5.8 also illustrates the semantic search engine behaviour, which is activated whenever it receives the input from the web page creating/updating the database using

content indexing from various web pages. The database is updated by the semantic engine, which constantly queries data from web pages using 'keywords' phrases or relationships of objects or persons. In fact, the semantic search engine is expected, after many iterations, to create a semantic cross layer, which denoted as Semantic Layer, of semantic indexed information and semantic link network that are stored in the database (figure 5.9). This covers all the discovered semantic information relationships, object relationships, grouping and classification etc.

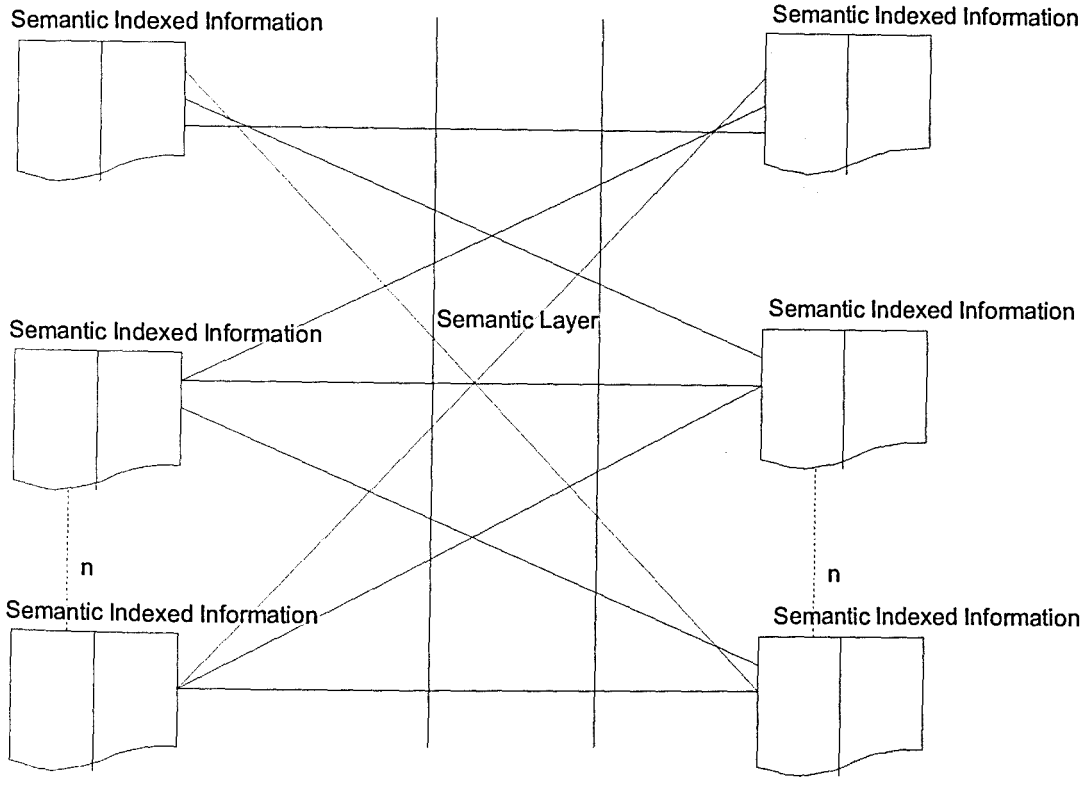


Figure 5.9 Semantic Layer Cross Linking Encoded Semantic Information

In summary the new approach is based on providing information on various objects and things, thus it can overcome the problem of 'information mishandling and misleading' and ensure that even the layman is able to use this new system of search, without knowing about the 'product or thing' or to master a special 'keyword' for searching. The approach is also based on 'real-time,' therefore it can handle complex situations and queries, and it is able to 'gather' new information which will be able to explain 'objects and things' for the end user, for example it allows information to come (attached) with the 'objects, things or products'.

5.3 Semantic Architecture

The implementation of the approach discussed earlier the construction of new semantic information architecture as shown in figure 5.10. This semantic real-time information architecture has a number of major components. The first component is the user input or query, which is looking for some relevant information. In the second component of the system is the semantic search and indexing engine. The third component is the intelligent engine looks for semantic information from the stored semantic data (depositories) but also it can find information from active real time web page contents and perform ontologies updates. The fourth component, is the semantic layer which consists of semantic data (meta data) and the 'semantic information.' The fifth component, is the data depositary (database) where the updated semantic information is stored. The sixth component, the semantic display where the systems answers the initial search questions according to the ranking of information 'semantic relation.'

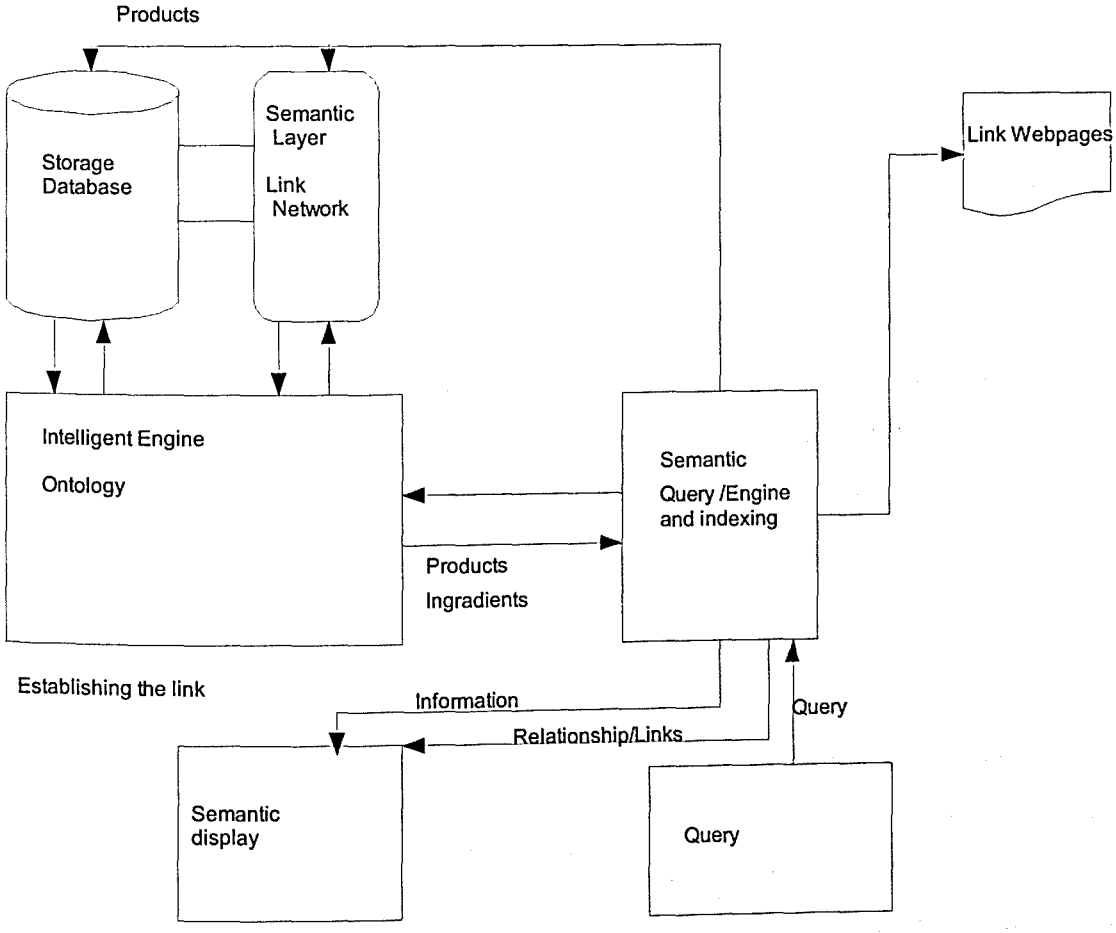


Figure 5.10 Semantic information architecture

When the information search begins, the query engine seeks new information from the semantic layer. The semantic layer includes the semantic database, which contains the pre-discovered information and the database for establishing new links. The semantic layer sends the search results to the query engine and then the query engine connects to the intelligent engine. The intelligent engine establishes a relationship between the semantic content and links to the query. If the intelligent engine is unable to find the semantic information from local depositary, it looks for new information and its relationship from the web contents, once it is found, then it will connect with the semantic layer and the semantic layer updates the information in the database. Thus, a number of processes are running in the background, using either local depositary or global web pages, which ensure that the nearest response to the initial query is semantically displayed i.e. with additional semantic information and/or semantic ranking.

For example a semantic product search, for say'diabetics' can be readily specified as product + diabetics, where the term diabetics is the query term and product is the class, which are required factors for rearranging shopping for diabetic products. Similarly, in other complex queries, the semantic product search is expected to be very efficient. For example, when asking for low calorie products in a shop in which multiple products are involved, the query can be specified as product + low calorie. It will bring up all the products with specified calories. In order to get maximum satisfaction from queries, each of the required product + keywords must be chosen, which means that the query results are semantically related to each other. This includes ranking of information such as dietary products for instance diabetic. When the user looks for the product, the system itself searches for information which is carried within the semantic web application layer.

5.3.1 Examples of Semantic Item Relationships and Processes

Search engines are the most popular applications on the web but, they still have plenty of room for improvement. The approach proposed earlier is intended to augment and improve traditional search result by clearly identifying semantic relationships between products, services etc. As explained in the previous section, the user can explicitly make a query to the search engine for more information on a product/service. For example, when the user specifies the keyword 'soft drink' for a product, the user/customer expects the query results

to be examples of soft drinks as well as additional related information that might be relevant to the user.

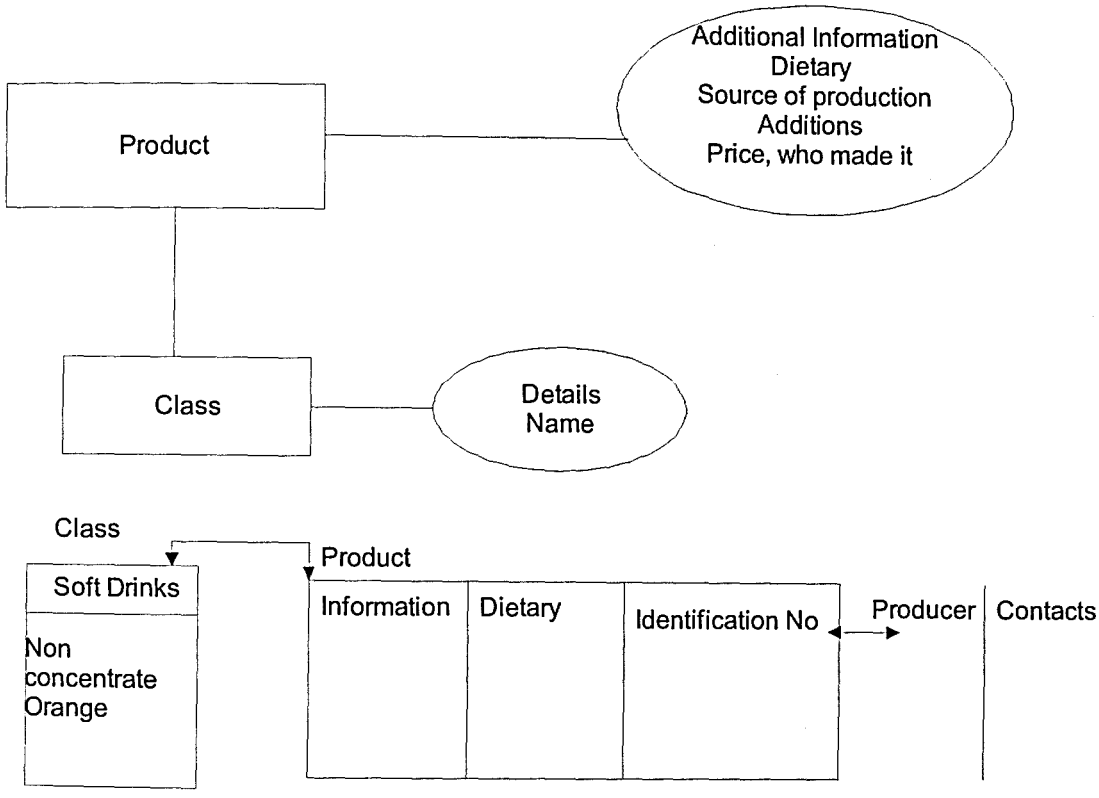


Figure 5.11 An Example of Product Information Relationship Diagram

An example of a semantic relationship is shown in figure 5.11 with detailed information on the products in term of their semantic information relationships. This can be extended into a meta data model, with a product which is unique together with additional information such as dietary , source of production, ingredients, price, by whom and when it was made. In order to facilitate the semantic search, the data storage has to be semantically arranged as shown in figure 5.12.

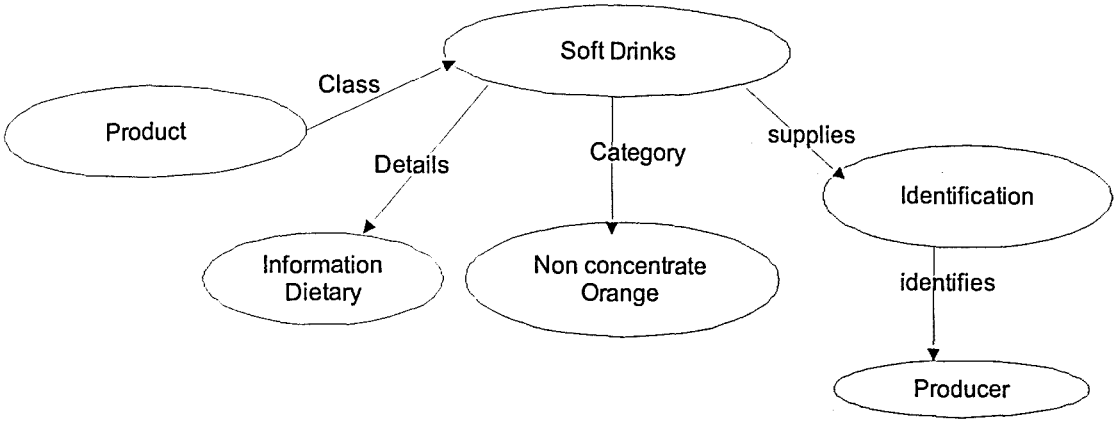


Figure 5.12 Diagram for semantic arrangement of product information

The implementation model will be based on the meta data which aims to enhance the semantic information search for particular 'product/object,' since it is produced out of semantic information (meta data). In order to search for a 'product/object,' the user needs to interact with the semantic real-time information search architecture, through the process "Searching", is shown in figure 5.13. The whole process starts with the user input or query, which is looking for some specific information, using keywords such as "Orange soft drink". The 'searching' process starts, using the architecture described earlier and finally the answers to the initial search are displayed semantically, according to the ranking of information together with any additional semantic information.

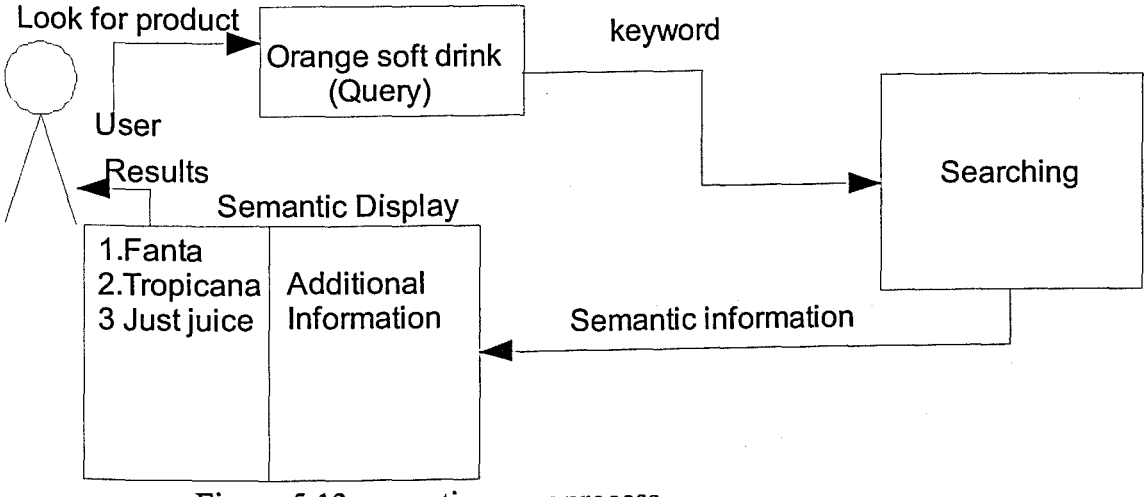


Figure 5.13 semantic query process

An example of semantic display where the search comes up with the current semantic information of a product is shown in figure 5.14 which show the actual product presentation according to user requirements-search.

Figure 5.14 an example of semantic display

5.4 Self Organising Hyper Linking Semantic System

The semantic information architecture presented in section 5.3 relies mainly on user search in order to semantically organise relationships between various products/services. This can however be extended into a semantically self organising global system, which is automatically and continuously trying to establish new links using past stored information and semantics search. In fact real-time retrieval of 'contents' requires self organising hyper linking semantic system which will function as described in the previous sections. Thus, it will enable the semantic retrieval of meaningful content of the web pages, based on objects, products or things and their associations and relationships and links the relevant information with meaningful relationships of the content of a web page with the formal query, to derive a more efficient conclusion and ranking of the results.

As before the information provided by the user, in the form of query, is an important factor for discovering hyperlinks i.e. in discovering content and establishing the relationships semantically. However, the self organising system, using existing information stored in a database, will automatically and continuously identify semantic content and establishes relationships and then group the links accordingly. With this automatic linking association, the 'semantic network' is created, thus the intelligent engine, is able to find out the relevant semantic information which is then stored accordingly in the database. This will result in the ranking of 'hyperlinking content' that is based on real-time semantic information and its relevance according to the context of the search.

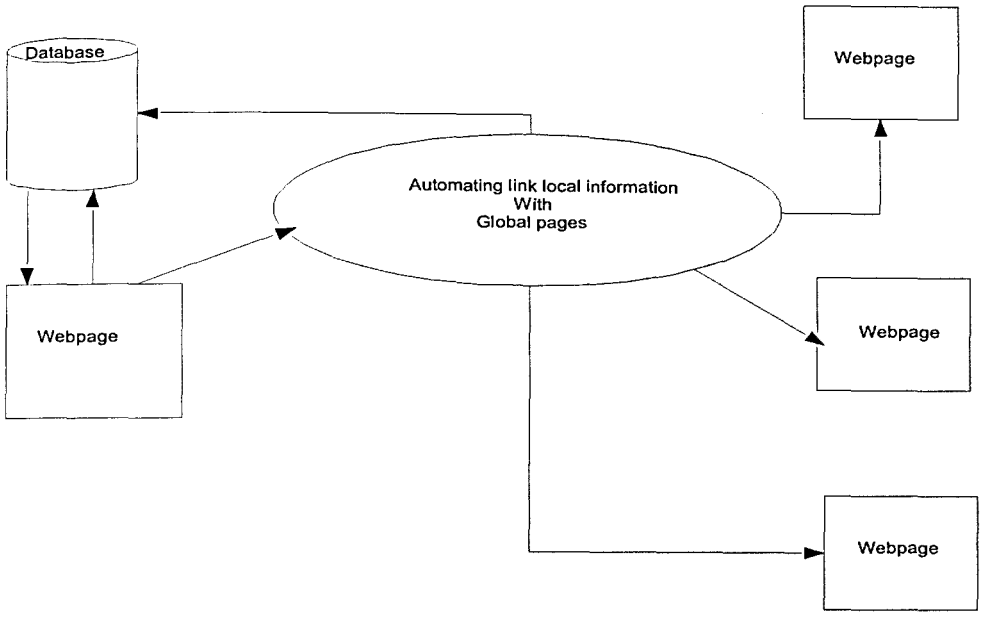


Figure 5.15 Self Organising Hyperlinking Information System

Figure 5.15 shows the real-time hyperlinking of semantic information, and the semantic database with its semantic contents as well as the automating link. The system automatically scans the existing semantic data and establishes links and relationships with the global pages. It will then create three types of linking structures:

- Semantic information, that index the meaningful information provided from the user query.
- Hyperlinking, automatically linking local information with global pages, according to the relevance of the search.
- Ranking of information results, that stores the 'semantic content' of each web page.

The hyperlinking system can retrieve semantic content from the global pages with links being grouped only by semantic information. Searching the 'semantic contents' and linking by the 'intelligent engine' can be defined as a set of links called $L=\{L1,L2,L3,.....Ln\}$ where each 'L' refers to any stored link of 'object, thing or person' and is noted by a 'class' that are extracted by the meta data from the page (content within a link). In the 'intelligent engine' (see section 5.3) each semantic item refers to a semantic content and URLs which encodes and establishes the relationship between global pages. For example a link 'L1' expresses a semantic relationship between semantic content of a class 'Class1' and another class 'Class2' and is associated with a connection number or weight, for identifying the strength of their

relationship . This will also help to identify and differentiate one class content from another one (chapter six explains the weighted value for identifying each item from the class).

5.5 Conclusion

In this chapter, the different aspects of semantic real-time information search were considered, and an indexing architecture that discovers and integrates information using self organising hyperlinks was presented, which can help in locating semantic information more accurately. The semantic real-time information search approach is based on a powerful semantic frame work and architecture that can automatically self organise and establish relationships according to the semantic search, among semantic content, and thus creates a semantic network of hyperlinks. On account of this new frame work and architecture, we have given an example of a semantic display, with semantic information (as explained in section 5.1) for examining the feasibility of this new approach, and particularly focusing on semantic real-time information search, and identifying uniquely the information attached to a product.

Using the proposed framework the search starts with keywords, representing say products, as input and the system generates results which are closely associated with the semantic product in terms of semantic metadata. The semantic product, incorporating the user requirements, searches for the semantic information. The intelligent engine looks for matching information and products, and presents the results to the query engine from the data layer. There are a number of other features associated with this new approach. A key feature is that a customer/user can easily search for information, just by clicking the object (product) itself. Each product has additional information associated with it which can be ranked such as dietary information, source of product, ingredients, manufacturer, when was it produced, price etc. Moreover, the approach will enable complex queries, thus allowing the search for complex answers on each product. This information could be used to rearrange the site according to user requirements (represented by semantic meta data), which will enhance the user experience.

Chapter 6 - Semantic Layer with an Intelligent Engine

6.1 Introduction

The current state of Internet and websites were reviewed in chapter 2 and 3, with the evaluation of some existing sites conducted in chapter 4 where it was observed that existing approaches fall short of satisfying the need of the Internet and website user. Therefore a number of semantic approaches were proposed to bring structure to the meaningful content of the web pages.

This chapter describes a new approach to semantic real-time information search which seeks to combine some existing approaches whilst resolving many problems of semantic real-time information search. This is done by introducing semantic real-time (entities) items/products, so that semantic web can be both semantic to an E-business, and can offer fully fledged E-commerce architecture for any businesses and organisations for real-time information search. This approach enables an overall semantic nature of E-commerce entity/product to be determined, which can then be identified on the web. Unlike existing attempts, techniques and approaches, this semantic real-time information search is derived from real-time life experience and items/products. Indeed, the approach provides a solution to one of the basic problems in evolving semantic E-business or semantic information searches by automatically establishing semantic links between entities and attributes and, these relationships.

This chapter starts by presenting a more suitable definition of semanticity, in the context of artificial intelligence on which the new approach has evolved. Then, the intelligent engine in the semantic layer is described together with all the required processes. Neural Networks are then briefly described and the choice of a network based intelligent engine is rationalised. The probabilistic network of the engine together with the ways how the connection weights are calculated are then presented. followed by examples on how the approach may be used in an E-commerce system for semantic information search and discovery. Finally, the chapter's findings are summarised in the conclusion section.

6.2 Semantic Web and Artificial Intelligence

The semantic web definitions and aims reviewed in the previous chapters indicate a potential future transformation of the web. However, there have been little suggestions in the wider literature of how this is going to be achieved or what is meant by "future web would be machine understandable information over the net." (Dan 05, Ber01) Having considered the existing definitions of semantic web together with the suggested quality factors for the evaluation of semantic web (chapter 4), and the semantic architecture developed in Chapter 5, a more appropriate definition of semantic web in the context of this work is "Semantic web is a meaningful integration of semantic items/products/services and their attributes that together constitute semantic information and which can be interpreted by machines for humans". Thus, it is essential to consider the logical reasoning aspects of the web, so that machines can easily deal with the semantic information for human understanding as included in the definition. Therefore, we need to add some forms of intelligence into the new approach which has lead into the consideration of artificial intelligence (AI).

Artificial intelligence, also known as synthetic intelligence, is a branch of science and engineering which studies and designs intelligent machines, in particular, intelligent computer programmes (Jos05, Luc00). It is also often used to describe the property or characteristic of machines or computer systems that demonstrate intelligence through learning, reasoning, planning, communication, knowledge and perception, as well as the physical ability to move and manipulate objects for example in robotics (Int09). There are many branches of artificial intelligence such as, Expert System, Machine Learning, Neural Network, Data Mining, Knowledge discovery, Natural Language Processing, Robotics, Computer Vision, etc. (Ari09, Tho08, Jin01, Mir98, Kor90). Semantic web applications will require many of the characteristics of AI, i.e. learning, reasoning etc for their effective implementation. Therefore, having considered semantic web and artificial intelligence we proceed to propose the architecture of the intelligent engine in the new approach.

6.3 Intelligent Engine in the Semantic Layer

Although existing semantic techniques show that different opinions of semantic web may exist, they lack comprehensive guidelines as to how to handle the semantic information by machines (Sam10). Only in handling this semantic information, however can semanticity truly be achieved. This ability to handle such information requires a new approach, and thus

constitutes the basis for the semantic real-time information search and discovery. The new approach is based on the semantic definition presented above and incorporates an intelligent engine which is expected to:

1. Link the semantic information from the local databases and the online websites (globally).

2. Identify the essential semantic information (semantic items/products/services) required for a semantic web.

3. Link the individual semantic items/products/services in order to achieve semantic information search.

4. Intelligently create new relationships between the semantic information searches.

5. Meet overall quality factors which can be used as a standard of excellence evaluation for the semantic web and information search.

The intelligent engine is incorporated into the semantic layer to form the core of the architecture of the system, as discussed in Chapter 5, with logical reasoning and learning abilities. It can also potentially be incorporated above the RDF and XML layers in existing semantic web architectures (Ber06) in order to make an intelligent and smooth processing of semantic information between layers (figure 6.1).

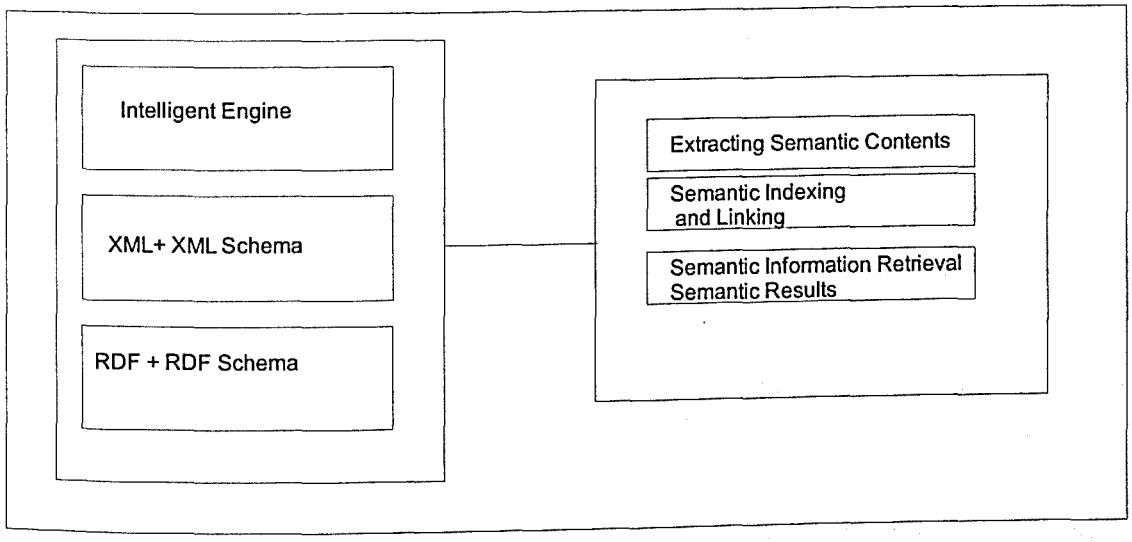


Figure 6.1 The Semantic web layer

The intelligent engine is there to extract semantic information and its contents. It is also responsible for automatic linking and storing of semantic information, semantic indexing, producing the semantic results and semantic information retrieval from various source such as websites, databases, search engines.

The lower layers of the semantic web (consisting of XML, XMLSchema) are much standardised, whereas the middle layers (RDF, Ontology) are still at the developing stage (Kat10, Ber04, Hor05). The upper layers, consisting of ontology vocabulary, provide logic that needs to be much more developed and these layers are still being researched (Jam09, Tor09, Pat04). The new approach can contribute well to the layers of the semantic layer architecture, since only the RDF and XML layers are heavily used. As discussed earlier it also will contribute to the reasoning part of the semantic web and can play a major role in indexing and querying of data by asserting semantics.

6.3.1 Ontology for the New Approach

Ontology describes the relationships between existing entities groupings such as relationships within a hierarchy etc. and their meaning in terms of concepts and relationships within a domain. The new approach uses the semantic database to store pre-discovered information (semantic data) and any kind of new data determined dynamically, as results of a search in the semantic system, are recorded by a logging process. The data within the semantic database can be used for establishing historical suggestions and establishing entity relationships and learning for creating ontologies. Establishing links with Metadata, is carried out by recognising the data related to a specific entity (specific image) from the semantic database.

The intelligent engine stores the resulting links inside the entity header for dynamic data sets. The goal of storing these resulting links inside the entity header is to give the user the opportunity to view the semantic links; therefore providing an easy way of sharing and publishing data or maintaining entity relationships by mapping the ontology with the previously discovered ones stored in the database.

As mentioned earlier, an ontology describes meaning of a domain including a finite set of concepts and their relationships or relationship between the concepts. This can be illustrated in a semantic metadata describing the content of a website, e.g. an E-commerce web site with customer, product, and E-selection are typical concepts, and which uses "keywords" attached

to a document in the website to identify the concepts. The concepts and their relationships are graphically presented in Figure 6.2 as an "E-selection Ontology" where ellipses are classes and rectangles are properties. Using this ontology an E-selection is made from a company by a customer. Each E-selection includes a product, with its price, which belongs to a product group and relates to other products. Each customer transaction is recorded in the company system.

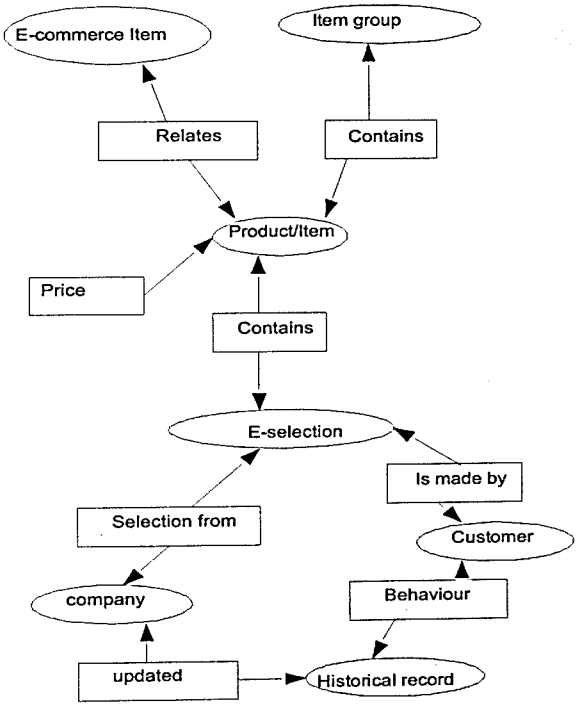


Figure 6.2 An E-selection Ontology

6.3.2 The Learning Process

As explained in chapter 5, the major activity of the new system is performed by various indexing and recognition methods. Pre-discovered data from the semantic layer could be used to determine the historical background of the user and entities. These dataset will also serve the system in the decision making process. The pre-discovered data also helps to retrieve information semantically established in the semantic layer. This can be achieved in two ways, the first one is to retrieve information according to the query that the user has entered and by logging the query in to the system, which will also help the system to determine the historical activity of the user. The second will retrieve information regarding the previously established entity relationships that are semantically related to each other. Thus, the learning process

relies on the use of historical data and requires the use and the application of artificial intelligence techniques in order to produce meaningful information.

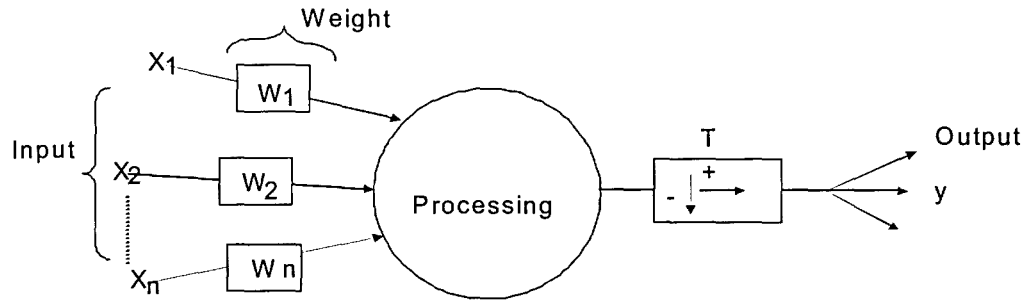
6.3.3 The Links

The semantic layer (figures 5.9 & 5.10) implies a connection link network for establishing the semantic relationships between items as well as ranking the semantic strength or weight. This means that some sort of links exist between a large numbers of items at least between those that belong to the same class. This can be easily represented by a network link model where each link has a weight representing the strength of semantic relationship, with weight say equal '0' implies no link whatsoever and weight equal '1' implies a very strong semantic link/connection. In fact the intelligent engine function is to manage the link network and its dynamic configuration. The intelligent engine can be constructed using some of the AI techniques mentioned in section 6.2, however due to the nature of problem at hand neural networks are well suited for this type of applications, thus some form of neural network seems to be an appropriate approach for the implementation of the intelligent engine of the semantic architecture described in chapter 5.

6.4 Neural Network

An artificial neural network (ANN), also called a simulated neural network (SNN), or commonly just neural network (NN), is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation (Hir08, Ham90, Mor90) (figure 6.3). Neural networks consist of non-linear adaptive computational processing elements called neurons. They are capable of changing their structures according to their internal or external information. The output (y) of the neural network is connected with its input (x), based on its neurons with the synaptic weight (w) connecting each neuron within a neural network, and threshold (T). The synaptic weight is very important since the activation of the network depends on it. The neural network is an interconnected group of nodes akin to the vast network of neurons in the human

brain (Bel03). A neural network is known by its neurons, network paradigms and learning



$$y = \sum_{i=1}^{n} (x_i w_i) + T$$

(Fil10).

Figure 6.3 The Artificial neural network

6.4.1 Adapting Forward Network for Semantic Relationships

For the purpose of this study we adopt a network with various weights for the construction of the intelligent engine with the responsibility of making connections and links between entities and potentially re-organises the network accordingly. As shown in figure 6.4 the neural network with input, output, neuron body, connection points and weighted values.

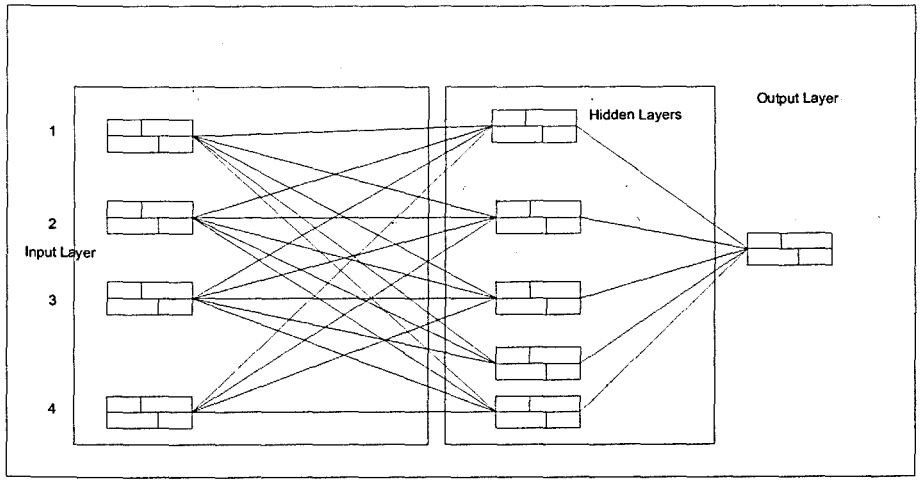


Figure 6.4 An Artificial Neural Network

An example showing how a simple neural network can be used for the purpose of establishing a relationship between entities and their output is shown in figure 6.5.

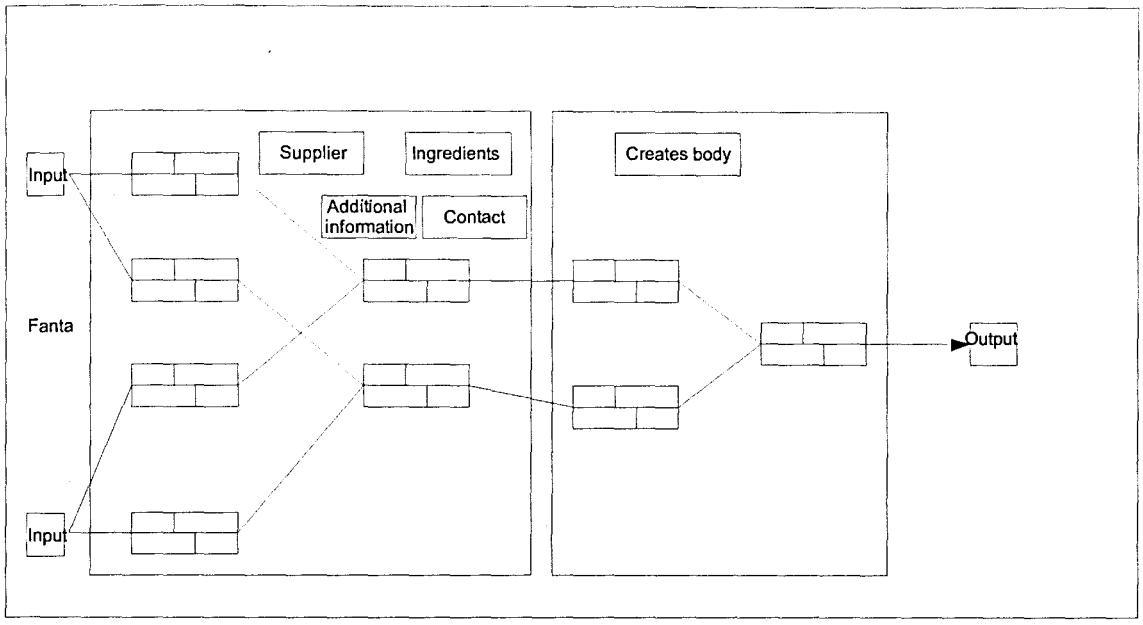


Figure 6.5 Relationships Network

This network illustrates the relationships between entities and creating neuron body and supplies the output. In the first layer two neuron bodies are formed from the given input and passes on to the second layer to form one neuron body that produces the output. Figure 6.6 shows simulation of a neural network, input values propagate through connection weights and neuron bodies and produces output which is based on the activation functions, built-in bias as well as the input accumulators.

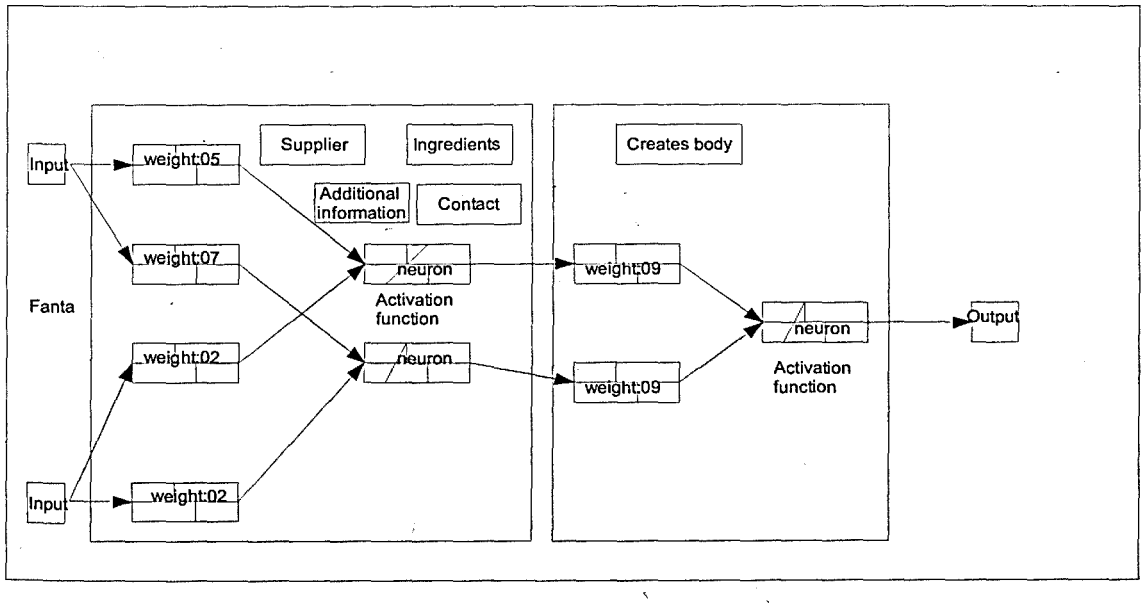


Figure 6.6 Propagation of the input values

During training network error values are given in the outputs and these errors are propagated backwards through the network resulting in the modification of weights and biases in neuron bodies, connections and links (figure 6.7). The error calculation can be presented as the difference between the actual output and desired output which can be accumulative (Cha06, Abu85, Rum86).

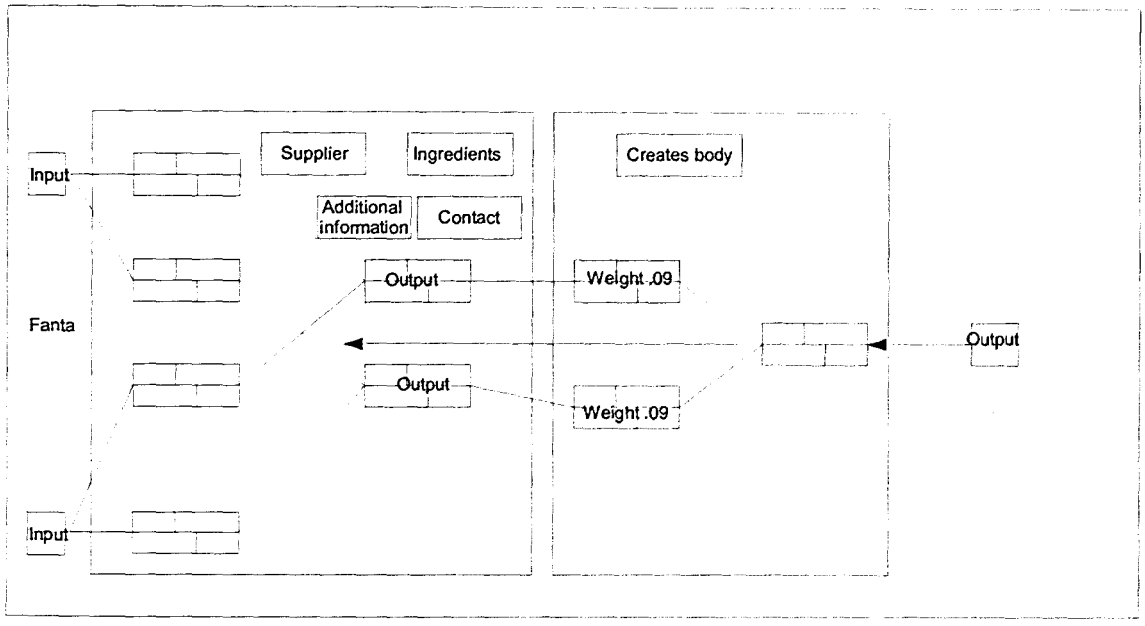


Figure 6.7 Backward error propagation

6.5 The Intelligent Engine

This section defines the real-time semantic information search with incorporated neural network forming the core of the intelligent engine, of the semantic architecture, which connects the query engine, database for semantic information storage (pre discovered) and database for establishing new links (figure 5.10). When the information search begins, the query engine begins searching for new information from the semantic layer. The semantic layer includes the semantic database which contains the pre-discovered information and the database for establishing new links. The semantic layer sends the search results to the query engine which then connects to the network in order to establish a relationship between semantic content and linking. If the network finds new information or its relationship from the semantic web contents, then it will connect with the semantic layer which updates the information in the database.

6.5.1 Semantic Link Layer

The intelligent engine architecture is based on a probabilistic network for establishing entity relationships. It is composed of layers and contains different sets of nodes as shown in figure 6.8. The layers of the network are composed of the items/products which are capable of accepting the inputs. Its main function is to calculate the weight of items/products links and connects them accordingly. Once, the connection starts with accepting the input nodes the system itself establishes semantic connections and ranks them according to the semantic strength. Thus, the semantic connection and ranking is unique, and each time the user searches for information the strength of the links increases and they become more semantic. The probabilistic network is established on account of semantic information search for example when searching using different 'keywords' it establishes the semantic connections and links between them.

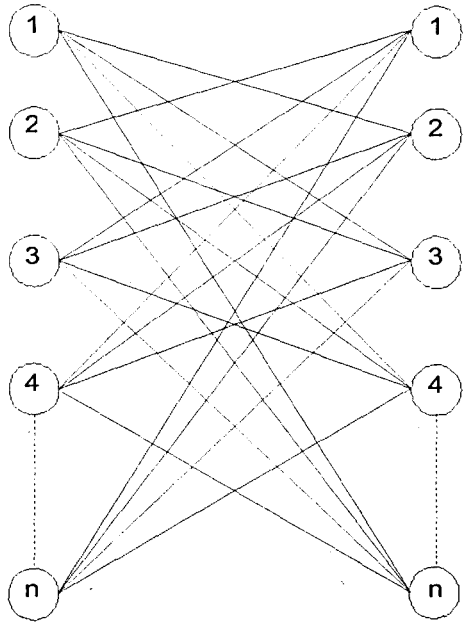


Figure 6.8 The probabilistic network topology

The connection weight is assigned according to the strength of semantic relationships which is established by the user's query and is associated with the semanticity of the 'input words.' The output is produced according to the strength of the meaning (semantic) associated with the input. This implies that the keyword in the query corresponds to the connection weight with the highest semantic and therefore in this way the output and the input nodes are semantically connected. A visual example the strength of connections is represented in figure

6.9 which clearly shows that there a strong connection between "Fanta" and "Orange", a weaker link between "Fanta" and "Sprite" and a very weak link between "Fanta" and "Shirt".

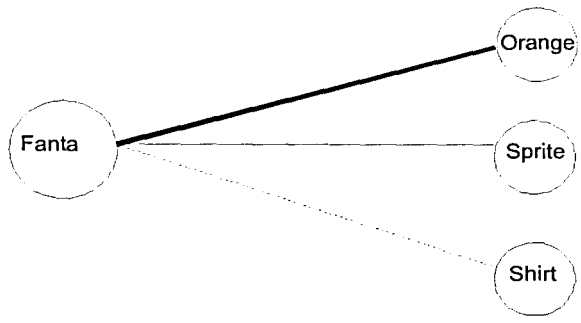


Figure 6.9 Strength of the Connection

The calculated weight therefore represents the strength of connections between two or more entities, keywords, images etc. For example, when a user is searching for an entity, a list of semantically connected entities is found and ranked according to their weights. However, considering that the proposed approach is aimed at web applications particularly in E-commerce it is important to give the administrators or business managers the ability to emphasis certain relationships between available entities/products. This will give the business the ability to promote certain products when customers are searching for related ones. Thus, the new approach will take into account not only the link weight that is established as a result of historical usage i.e. occurrences but also the manager/administrator defined weight.

6.5.2 The Criteria for Establishing Semantic Relationships

The network topology represents the semantic relationships between the entities and their semantic connection points. The semantic connections and links between entities vary according to the weight of each connection. In fact, the weight of the connection determines the semantic nature of each connection with each entity attempting to semantically connect to some other entities. The weight value, however, depends on the frequency of the occurrences of entities and potentially a pre-defined weight which establishes/estimates at the initial stage the strength of the relationship of two or more entities. The occurrences and their contribution to the total weight are done dynamically by the system, while the pre-defined weight contribution is defined by the user, i.e. is subjective. Both weights can make contributions to assigning to the total weighted value as given in equation (6.1)

$$W = R \times W_{pre} + (1 - R) \times O_n \qquad \text{equation (6.1)}$$

Where W is the total weight that defines the strength of the link between two entities,

W_{pre} is the pre-determined weight of the link which can be defined by the users based on their experience with the considered entities,

O_n is the weight defined by the system using the normalised frequency of occurrences

$$O_n = \frac{O}{max\ O(x)} \qquad \text{equation (6.2)}$$

O is number of occurrences, max O(x) is the maximum value of occurrences

$R \epsilon [0,1]$ is the ratio of the pre-detfined weight and the normalised occurrence.

Thus, when:

$R = 0$, the total weight is defined by occurrences only

$R=1$, the total weight is defined by the pre-defined weight only

Any other values of R imply that both occurrences and pre-defined weight contribute toward total weight.

The pre-defined weight is given according to the connection established by the user. If the user makes connections and links semantically, a connection weight (pre-defined weight) will be assigned according to the user's view of the strength of the semantic connection. The highest semantic element gets the highest weight and this will define the semantic connection between two or more entities.

For the purpose of this study we adopted a network with various weights for the construction of the intelligent engine with the responsibility of making connections and links between entities and potentially re-organises the network accordingly. As the approach is based on forward networks for semantic relationships there is a clear need to define the connection weight which, in line with other neural or probabilistic networks, is based on history and learning strategy. While this still applies in the proposed intelligent engine by representing the weight using the normalised occurrence *On* in equation 6.1, an additional weight *Wpre* (pre-determined weight) has been introduced in this work to reflect the users' needs and to

allow the users to modify the weight according to their business requirement for example marketing a new product by giving its connection high weight (strong relationships to all main products). Moreover, a ratio of the pre-determined weight and the normalised occurrence (R) has also been introduced again to give the user the flexibility in defining the total weight W which defines the strength of items/products relationships.

When implementing the approach a weighted value index is considered which help in storing the weighted value which is created from occurrences, ratio and the predefined value of products and their properties in the product list in the value field of the index. Thus, if a user is searching for women's 'dress', the ID path leading to the specific relationships is retrieved from the weighted value index.

In the product weight index, the unique product weight of products is stored in the product list. The occurrences, ratio and predefined value of each product element are stored in the pre-defined and weighted value fields of the product. Thus, when a query id made regarding a product, all related indexed products are displayed according to the strength (weight) of their relationships.

6.6 Experimental Evaluation

As discussed earlier there is a need to identify the semantic strength (weight) of individual links between various entities. For this purpose a number of experiments based on "google search" results are used, where the search for specific entities is exploited in order to establish their semantic relationships. This is done by considering a variety of entities and defining the weights of their relationships. In the first experiment the weights of the relationships are dynamically determined i.e. are based on occurrences only. Then, the weights are calculated using a pre-defined factors as well as dynamically defined ones i.e. a combination of the two weights. Finally, some of the entities are considered to have strong relationships, therefore their weights are pre-defined only.

The experimental evaluation starts by considering one-to-many relationships i.e. relationships between one item say "shirt" and many other items such as "trousers", "suit", "tie" etc as shown in figure 6.10, with the aim to establish the strength of their relationships, thus establishing a level of semanticity. The strength of the relationships is based on their weights which is calculated using equation (6.1) that take into account the occurrences, the pre-defined weight and their ratio. In the experiments the ratio of the pre-defined and

dynamically calculated weights (based on occurrence) is varied in order to show the impact of the subjectively pre-defined weight and the system determined one, on the overall weight i.e. on the strength of semantic relationships.

The results of the evaluation are shown in figures 6.11 – 6.15 starting with the pre-defined weight *Wpre* set to 0.5, implying a neutral view of the user (manager, administrator etc.) on the relationships while varying the value of R, starting with $R = 0$, i.e. the weight varies according to occurrence, next $R = 0.2$, 0.5 and 0.8, implying that both occurrences and pre-defined weights are taken into account in various degrees. Finally, $R = 1$ i.e. the total weight depends only on the pre-defined weight.

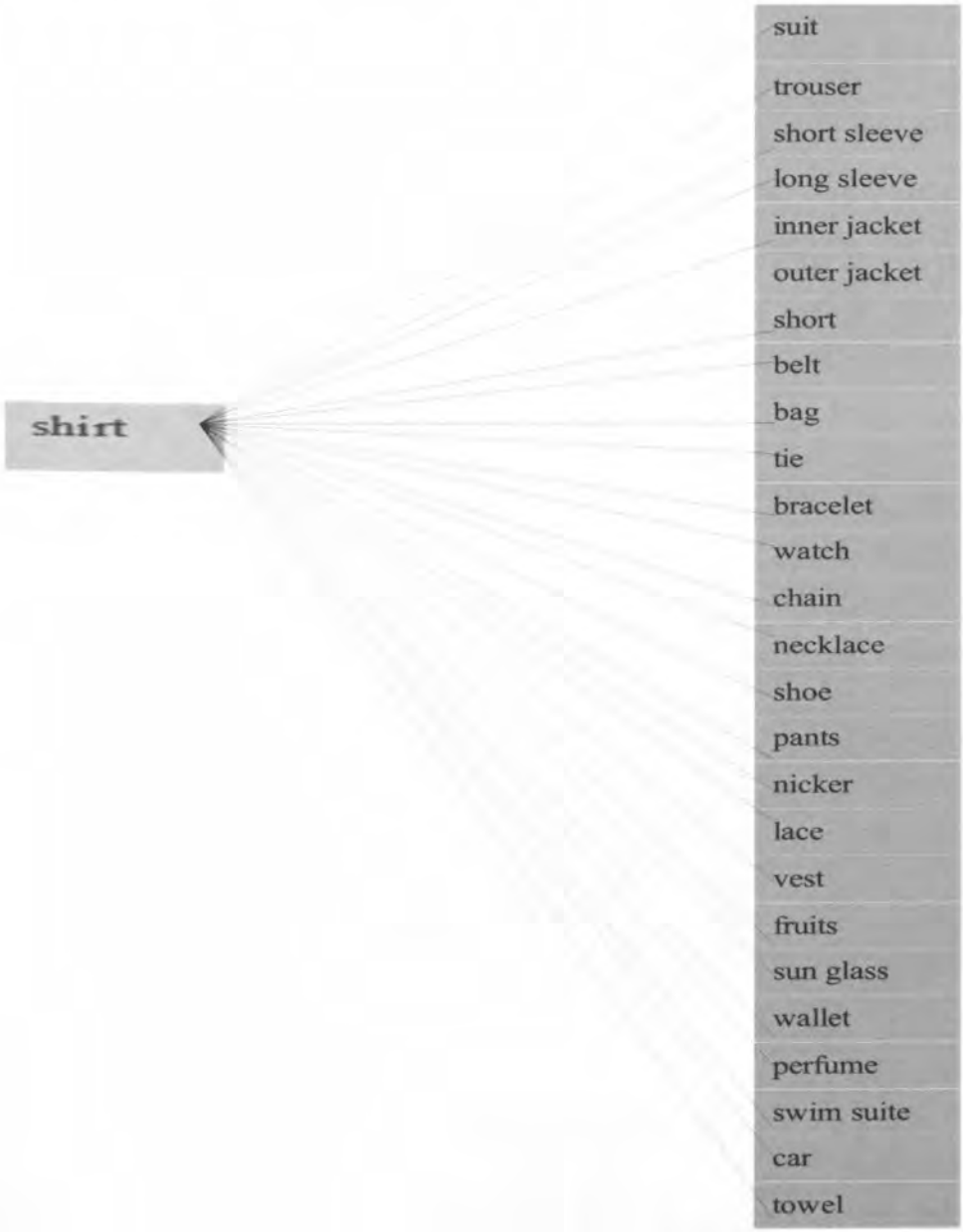


Figure 6.10 List of the used entities

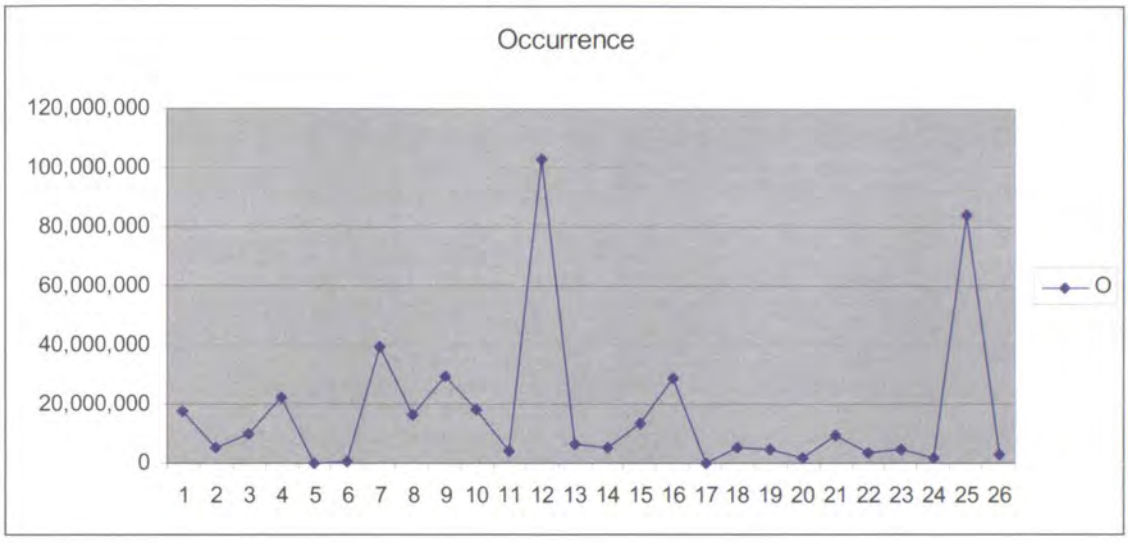


Figure 6.11 Occurrence of the entities/keywords

Figure 6.11 shows the working of the new approach for assigning weighted value. This experiment is based only on the occurrence of entities/keywords, i.e. pre-defined weight is not taken in account with ratio $R = 0$. The occurrences are very high at two points 103,000,000 and 84,300,000, thus, the weighted value reflects the highest semanticity at those levels, which is based on the previous frequency of search. Thus, the semantic connection is made according to the occurrences only.

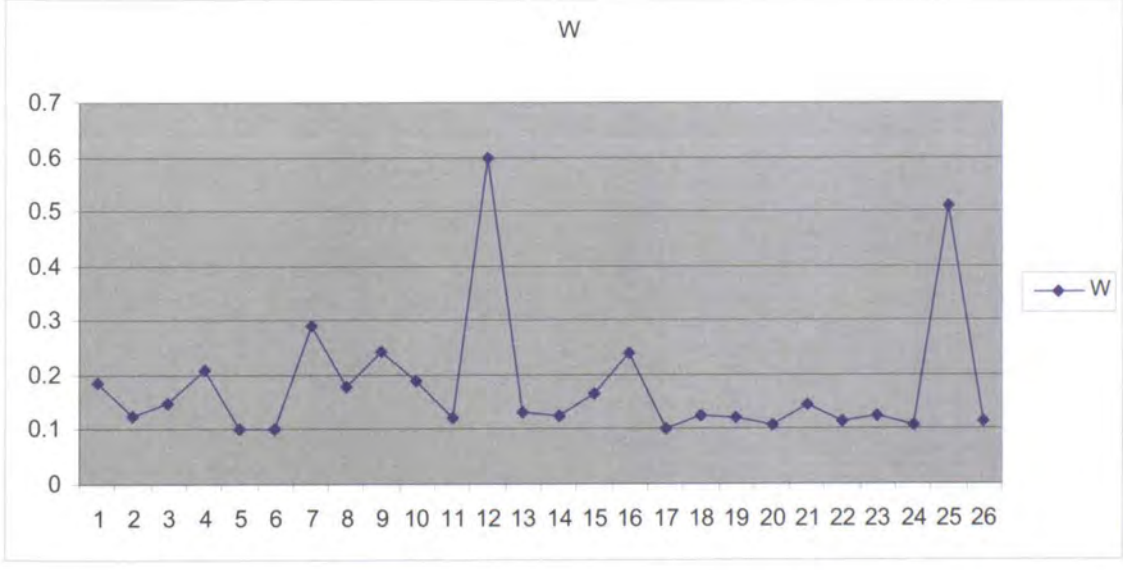


Figure 6.12 Pre-defined value *Wpre = 0.5* and ratio *R = 0.2*

In figure 6.12 the *Wpre* value set to 0.5 and *R* is set to 0.2. Thus, the weighted value depends on both the occurrences and the pre-defined weight. It is clear from figure 6.12 that the graph

shows the same pattern as figure 6.11 but it takes into account the contribution of the pre-defined weight and its impact. This can be seen more clearly in figure 6.13 where *Wpre* value is still 0.5 and *R* is increased to 0.5 therefore increasing the impact of the pre-defined weight on the total weight. The same pattern is repeated in figure 6.14 where *R* is set to 0.8 increasing further the impact of the pre-defined weight.

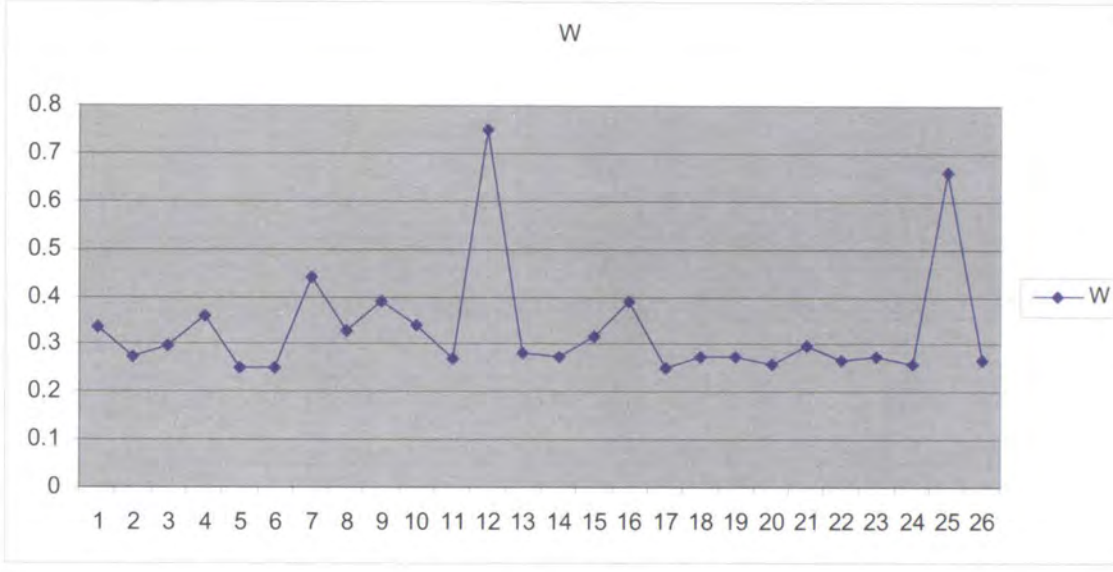


Figure 6.13 Pre-defined value *Wpre = 0.5* and ratio *R = 0.5*

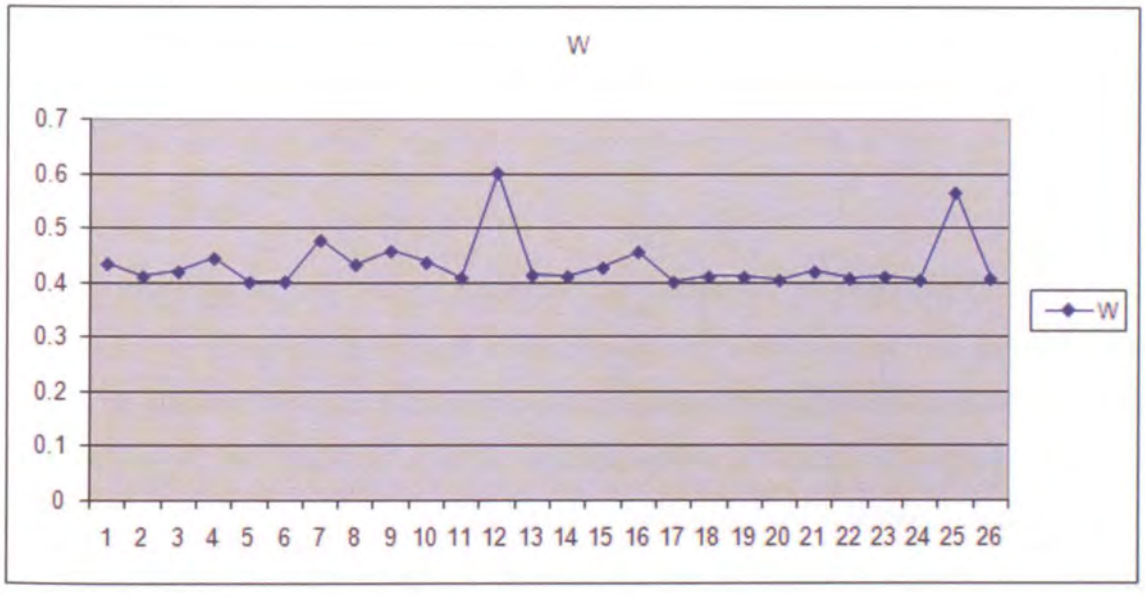


Figure 6.14 Pre-defined value *Wpre = 0.5* and ratio *R = 0.8*

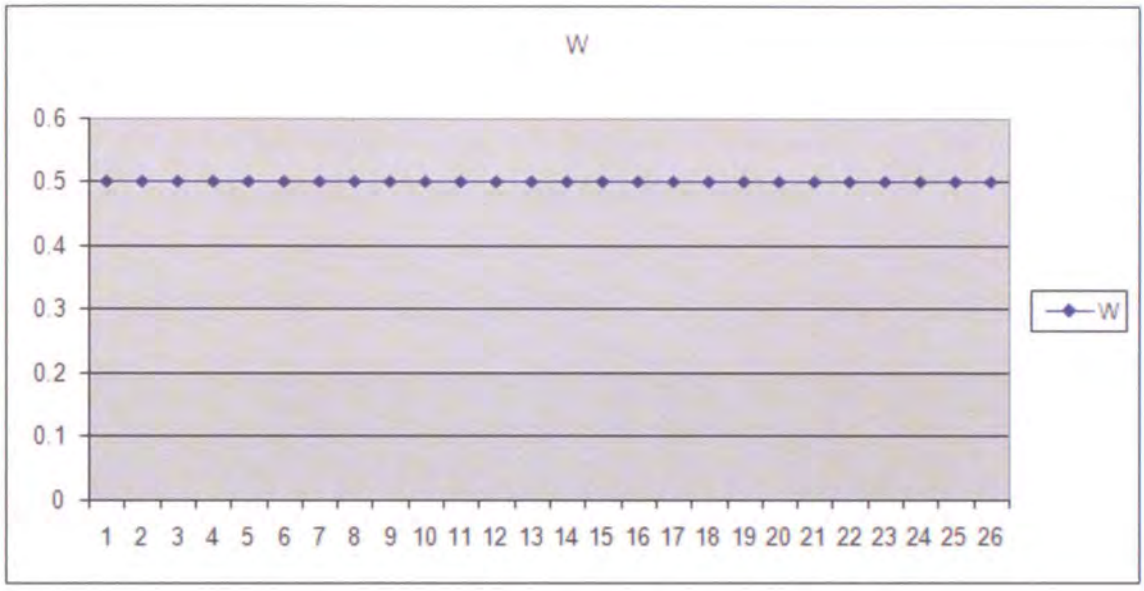


Figure 6.15 Pre-defined value *Wpre = 0.5* and ratio *R = 1*

In figure 6.15 the *Wpre* value set to 0.5 and *R* is set to 1. By setting *R* to 1, the impact of the occurrences is totally removed, the total weighted value is equal to the fixed pre-defined weight (*Wpre = 0.5*) which is it maintained for all entities as shown in figure 6.15. The experiment was carried out to show the impact of the pre-defined weight i.e. pre-defined strength. However, in real life this scenario is not expected, since different entities are very unlikely to have the same weights and strengths of semantic relationships even when only defined by the user. Thus, the next experiments take into account the subjective view of the user on individual entity relationships i.e. *Wpre* is varied according the subjective semantic strength which is driven by the user's experience with individual entities and their relationships.

This starts by again by setting *R =1*, therefore the impact of the occurrences is totally removed, the total weighted value is equal to the pre-defined weight which is it varied by the user and is done subjectively, i.e. the user is deciding on the weight and strength of every relationship (figure 6.16). This is a more realistic scenario since different entities are expected to have the different weights and strengths of semantic relationships. There are clear differences between the subjective view of the user (figure 6.16) and the system determined view, which based on occurrences (6.11). The first one reflect real life experience of say the business, the second is based on collective view, usage mainly number of occurrences/hits. The first one can reflect the expert knowledge in a particularly field or specific interest in promoting certain relationships (marketing, advertising etc.) therefore it is accurate reflections in a number of situations while the collective view provides an accurate overall

picture, but more mechanical one, and perhaps showing a wider view which might in some cases create unexpected connections. Therefore, the approach takes a unified view of user defined weight and system defined weight in combination which can only improve the search and make the semantic relationships more realistic and more accurate.

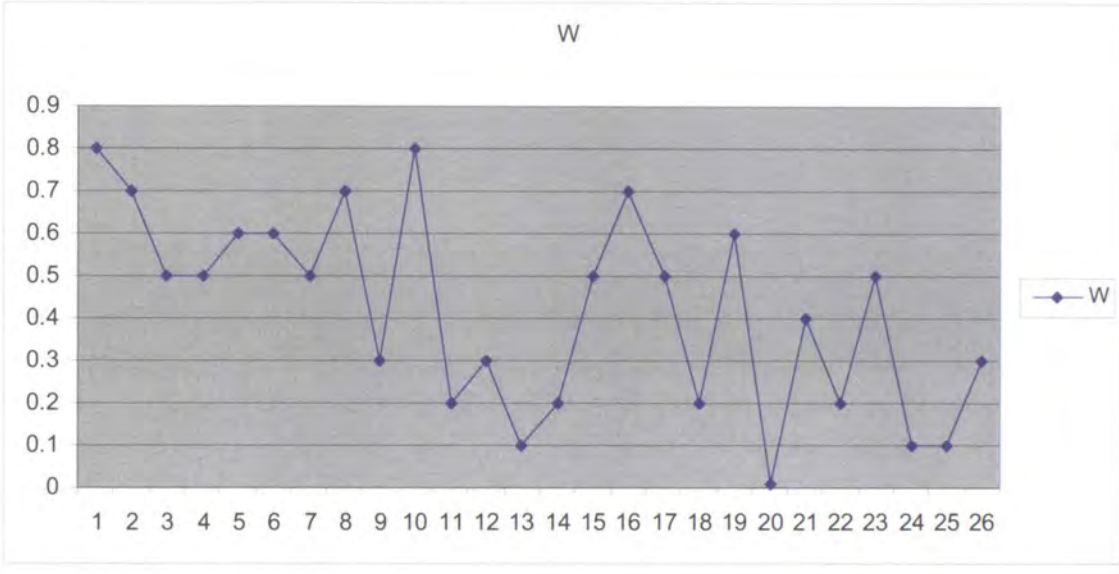


Figure 6.16 The impact of the user selected values With $R = 1$

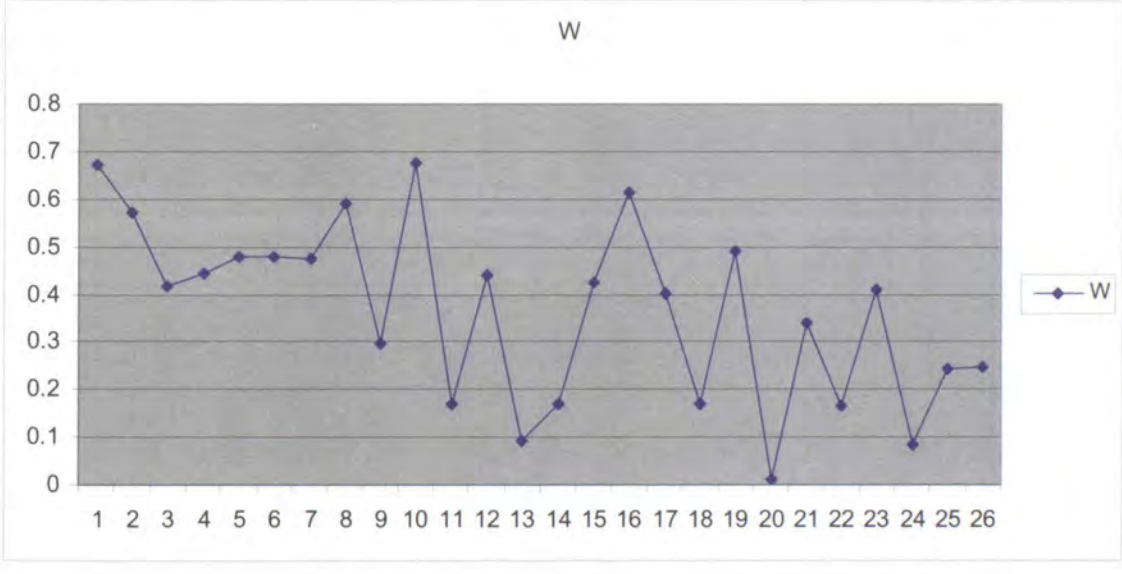


Figure 6.17 The impact of the user selected values With $R = 0.8$

Further, the selected (pre-defined) values of the weight are kept while varying the ratio R. This can be first seen in figure 6.17 where R is set to 0.8, with the impact of the user defined weight is still dominant, and a similar to figure 6.16 pattern is emerging. As the value of R is decreased to 0.5 the impact of the user defined weight is reduced (figure 6.18). Finally, as R

is set to 0.2 is impact of occurrences is higher (figure 6.19) and a pattern similar to that of figure 6.11 is re-emerging.

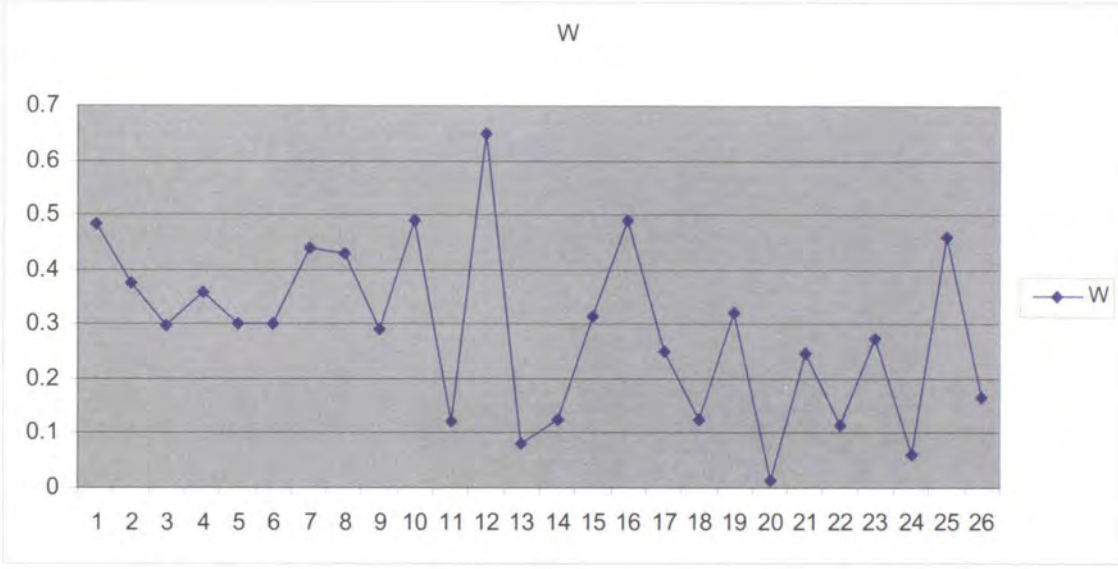


Figure 6.18 The impact of the user selected values With $R = 0.5$

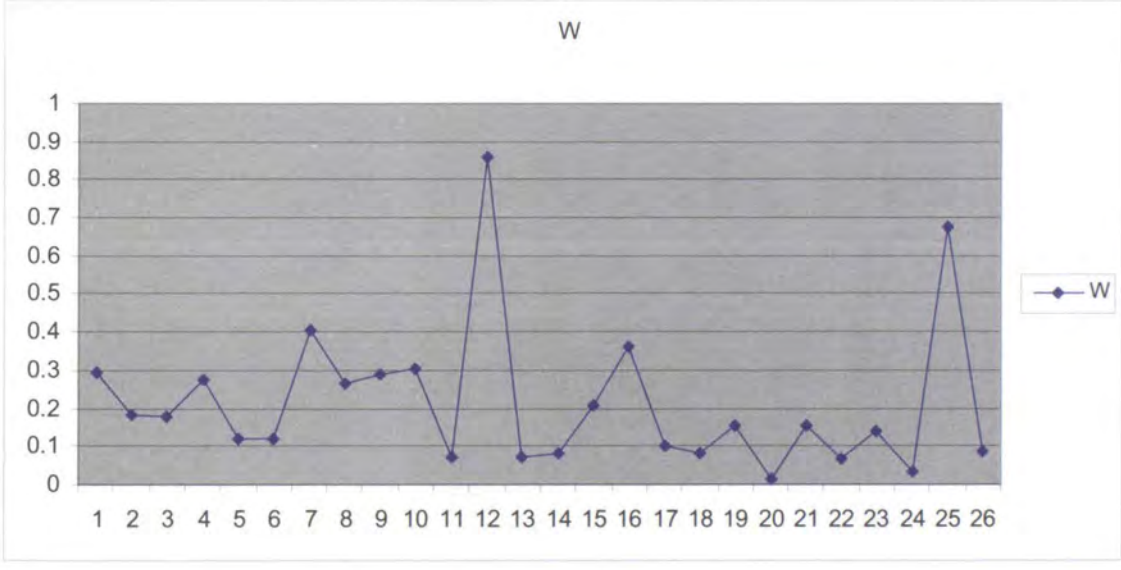


Figure 6.19 The impact of the user selected values With $R = 0.2$

The previous experiments have considered the evaluation one-to-many relationships i.e. relationships between one entity and many other entities. In the following experiments two-to-many relationships are considered i.e. separate two entities having relationships to the same entities as shown in figure 6.20. Thus, a comparison can be made of the strength of

semantic relationships. The weights are again calculated using equation (6.1) that takes into account the occurrences, the pre-defined weight and their ratio.

The particular chosen scenario in figure 6.20 is used to view the differences of semantic connections namely "Italian" and "German" keywords and the displayed entities. Firstly twenty one items are selected for Italian and German classes (figure 6.20) with the same pre-defined value (*Wpre = 0.5)* for making semantic connections. However, the same entities for 'Italian and German class' showed variation in their semantic connections.

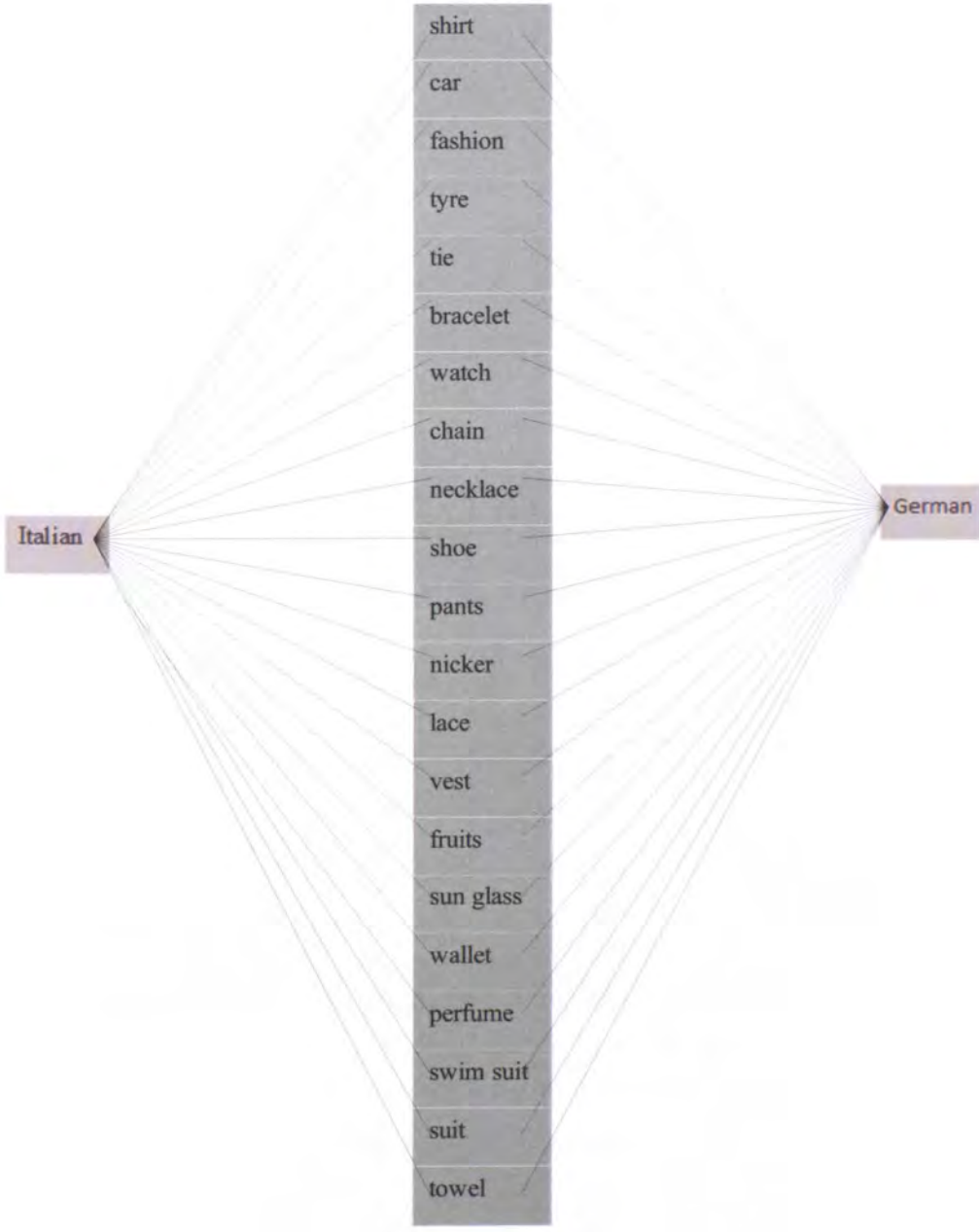


Figure 6.20 Italian and German entities

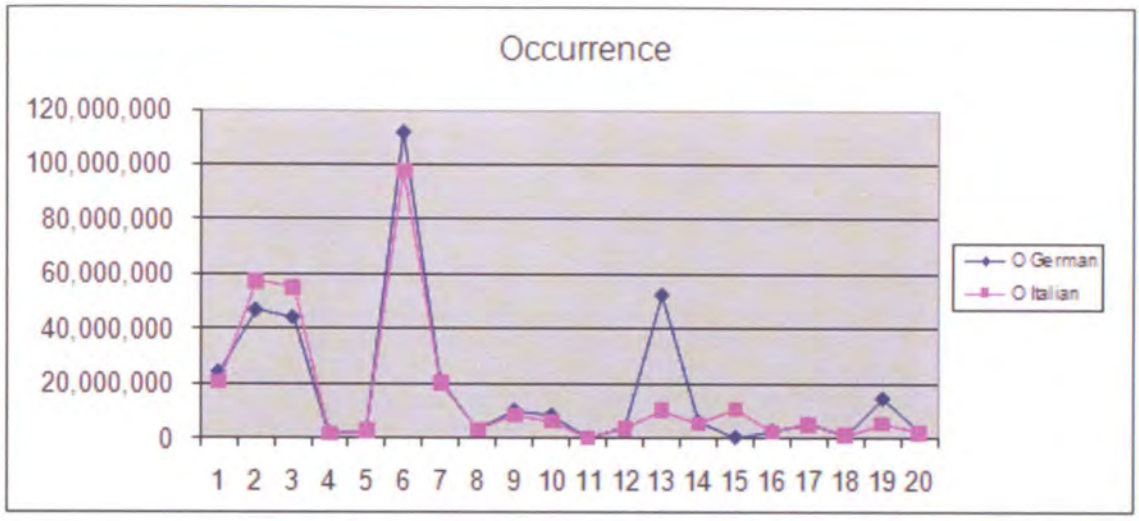


Figure 6.21 Occurrence of the entities/keywords

Figure 6.21 shows the variations of occurrences that exist between two semantic connections within two classes. The occurrences are shown for German and Italian classes using two-to-many comparison. This experiment is based only on the occurrence of entities/keywords, i.e. pre-defined weight is not taken into account with the ratio $R = 0$. Although the graph show similar pattern for both "German" and "Italian" for the chosen entities, the differences are significant at certain points of the graph. The differences become smaller as we increase the value of R from 0.2 to 0.5 which are shown in figures 6.22 and 6.23.

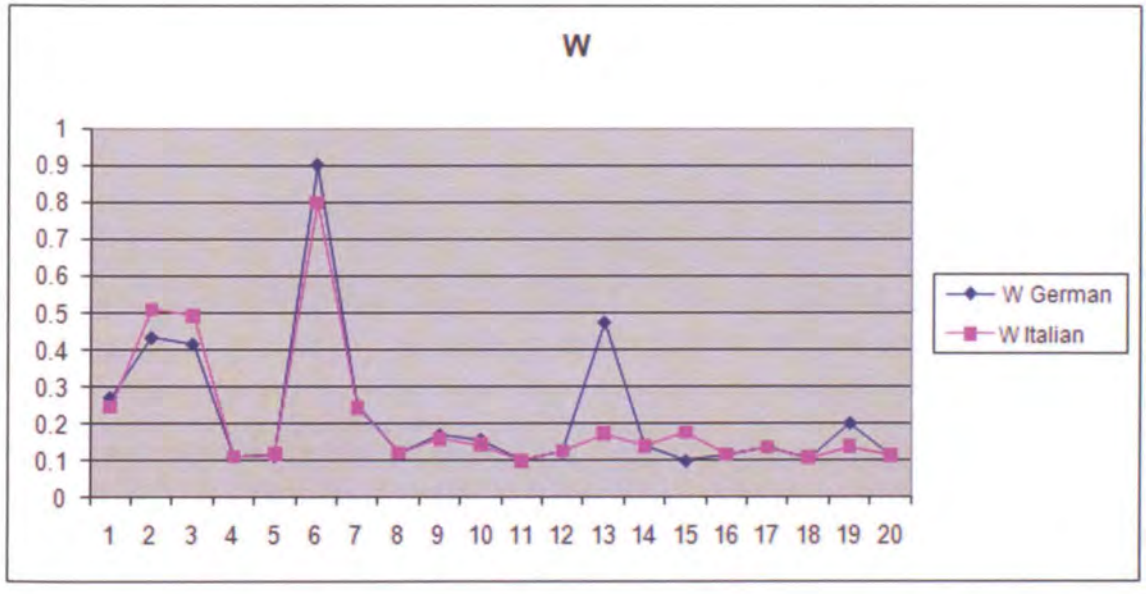


Figure 6.22 Pre-defined value $Wpre = 0.5$ and ratio $R = 0.2$

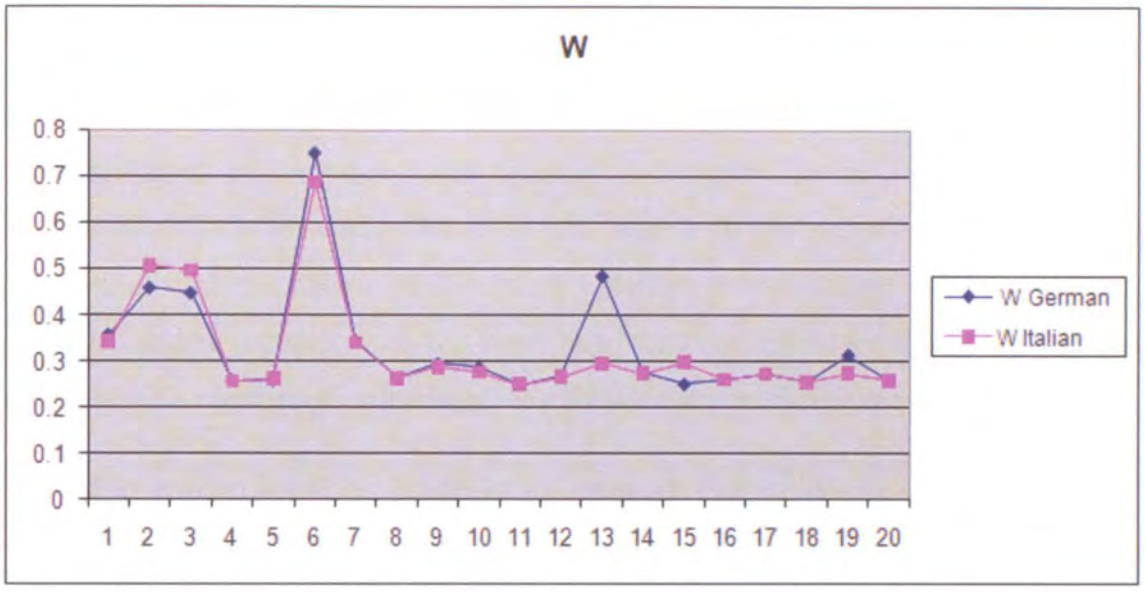


Figure 6.23 Pre-defined value *Wpre = 0.5* and ratio *R = 0.5*

The same pattern is repeated in figure 6.24 where *R* is set to 0.8 increasing further the impact of the pre-defined weight, which shows that the total weight is moving toward the pre-defined one for both entities as the ratio *R* increases.

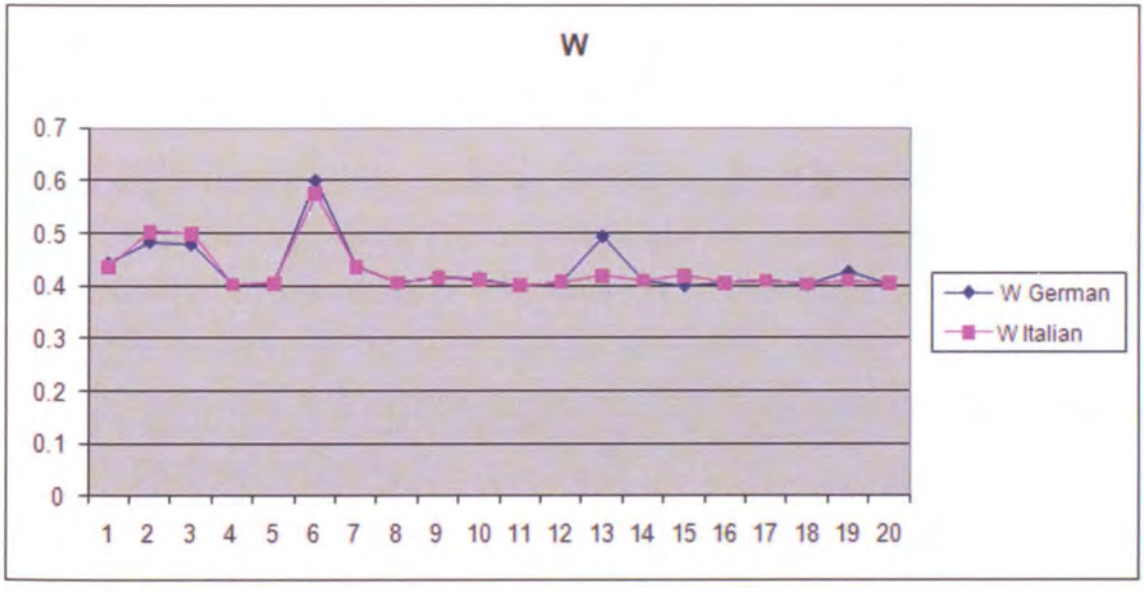


Figure 6.24 Pre-defined value *Wpre = 0.5* and ratio *R = 0.8*

6.7 Conclusion

This chapter started by discussing the applicability of some artificial intelligence techniques for the establishment of semantic relationships of different entities, with particularly emphasis on probabilistic networks. An intelligent engine in the semantic layer was then presented together with the choice of a network topology for its design. The probabilistic networks of the engine together with the ways how the connection weights are calculated were illustrated and a number of experiments to evaluate the approach were conducted.

The investigation into probabilistic networks for semantic real-time information search and discovery was focused on the weighting factors that define the strength of the semantic relationships between different entities. The presented approach takes into account both system determined and user defined weights, as well as the ratio between the two weights in its determination of the total weight that defines the strength of the semantic relationships. The user defined weight can be applied when a local E-commerce business, with its products and services and its own data storage, needs to define the relationships between its products/services according to its business logic and the potential profits, which can be done by setting the relationships at input stage (or even get them modified at a later stage), but which also can be modified automatically according to usage.

The conducted experiments show the impact on the total weight of the individual weights, it also show that results from system only determined weight and user defined weight can vary considerably, therefore a combination of both would be recommended, however, the final decision should be driven by the business case and the potential benefits. The business can use its local database, and well as entities available globally on the web, which is considered in the case study presented in Chapter 7.

Chapter 7- Integration and Evaluation

7.1 Introduction

Although the massive growth of web technologies and E-commerce has provided continuous access to information which is used by millions of people across the world, E-commerce is still facing many issues and challenges that require research, developments and improvements, especially from the semantic E-commerce point of view. A number of theses issues were discussed and considered in the previous chapters and a semantic architecture framework was proposed in chapter five, some aspects of which particularly the intelligent engine were presented in chapter six with some initial evaluation of the proposed network based engine.

This chapter presents the integration of the proposed framework and architecture together with the indexing, semantic storage, semantic search and semantic display concepts. The integration of the new approach is carried out using a number of technologies and concepts that were discussed and reviewed in chapters two and three. In order to evaluate the potential of the approach a discussion on E-commerce as a background is considered and then a case study is constructed to support and evaluate the new approach, together with results from applying the approach particularly in a local environment.

The chapter starts with a discussion of E-commerce systems and the requirement for product E-catalogues/lists, then it carries on with a description of the selected business and E-commerce application. This is followed by a description of the components of the employed system which comprises the query interface, semantic engine and semantic display and storage. Then, the semantic data store where the product lists are stored and the advantages of storing and retrieving the lists using the new approach are highlighted. The application of the approach is then described particularly in terms of the products semantic relationships and semantic display which can be applied in local and global environments. Then, an experimental study is conducted with the aim to evaluate the system using a case study based on a retail E-commerce business application. The results and analysis of the experimental evaluation are then presented. The chapter concludes with a summary highlighting the results of the experimental study and showing the significant benefits of the new approach.

7.2 E-commerce and the Web

Nowadays online business is becoming very common and is supported by the availability and accessibility of the web. Websites involved in online business are usually known as E-commerce or online shopping sites. As Chris Ramsey (Ram05) stated these websites are there to advertise goods and services and thereby promote corporate identity by selling the products, services and other goods. Thus, E-commerce websites have in the short and long run to be able to satisfy their customers/users. This requires the sites to have many functions such as maintaining products/services lists/catalogues, pricing information and shopping cart function etc. as well as secure transaction mechanisms. E-commerce systems should also have the ability to provide specific information to the customer by creating the data electronically before the actual purchase which allows the customer to look for and evaluate the product and print/download the information. E-commerce systems should also enable storage of customer information and supply the customer/s with a reference number etc.

However, it is hard for a customer to relate to many of an E-commerce site's pages if they are represented by poor structure and poor relationships between available products/services. A solution is to define E-commerce features in terms that are more understandable to the customer (see chapter four). One such approach is the Merchant Server Architecture (MSA) (Nic02). The essence of the MSA architecture approach is based on the function of product presentation which is done through using the product information from an electronically produced catalogue and/or order submission etc. This technique involves listing and presenting the product information. Another technique is the SET (Secure Electronic Transaction) architecture (Set97, Nic02) which involves adding an electronic payment function to the merchant server model. Open Market commerce architecture is another technique (Rob00) which includes functions for product presentation, listing of product information from an electronically produced catalogue, order entry and submission, electronic payment and customer service. Another architecture which is designed for B2B E-commerce is known as Open Buying on the Internet (OBI) (Ste00, Tru03, Sin09). This approach provides an opportunity for the buying organisation to select a supplier, browse the supplier's catalogue and place an order confirming the product entry and payment.

It is clear however, that the use of product lists/catalogues form the core of all approaches as well as most online businesses. Therefore, in this work a case study based on the E-commerce product list is constructed, which will incorporate some of the above discussed

features, to further demonstrate the semantic information search and retrieval approach and architecture presented in the previous chapters. The product list is created and stored in a database and is displayed on a site for various products and services.

7.3 Case Study: Product List in E-commerce Systems

The product list in E-commerce is one of the most important concerns of the E-business community as it is the main link between the customer and the entrepreneur. An entrepreneur is the one who sells products on the web. Customers need to browse the products before making any purchase which is true in both real retail environment or online, on the other hand entrepreneurs need to keep track of the products including quantity (available and sold),variety, types etc. In all cases a product list must exist and all products must be stored.

However, as the Internet community grows, the growths of E-commerce problems also increase and therefore E-commerce systems require continuous improvements and technological developments especially from the search engine point of view and information storage. Entrepreneurs and online customers are the two main players of E-commerce, and E-commerce is expanding because of these two major players. In every E-commerce application and technological usage, one of the major topics of discussion and research is that of the storage of data and the safe as well as the efficient retrieval of the data. The stored data provide information about the business (products/services) and the customer including customer information, payment, card information, ID, delivery and invoice information etc.

The major concern of this thesis is the 'information search and retrieval'; therefore product lists are significantly important since as mentioned above they are crucial for the entrepreneur and for the online customer. E-commerce community is very much dependant on product lists which have to be reflected in both the database as well as the user interface. The database stores the information and the customer retrieves the information from the user interface. In order to enhance the entrepreneur and customer satisfaction the efficiency of information search and retrieval over the product lists is very important. Therefore in this work we deal with product lists and indexing in semantic language format such as RDF, XML and other ontology formats so that information can be stored and retrieved semantically. The product lists can be indexed and stored in the semantic database in the format of text, images, audios, videos etc. The product lists are to be identified, classified and indexed in such a way so that

semantic queries can be easily performed. The semantic query should be able to retrieve all the semantic information which is necessary for the user and the entrepreneur.

Semantic product lists are the main concern and most important components of semantic E-commerce systems, as they are the main connecting links between the customer and entrepreneur. The customer searches the product list for specific products and is the potential purchaser of products. The traditional and current way of storing data is through a database management system, where the data is stored and retrieved later by way of a keyword search. In chapter five we have explained the various types of indexing and product listing. A semantically stored product list has also to be retrieved semantically, indeed a product might naturally and semantically belong to a group (base class) of same/similar products and/or be semantically related to completely different products (figure 7.1). This process will help the data to be semantically sorted for the end user.

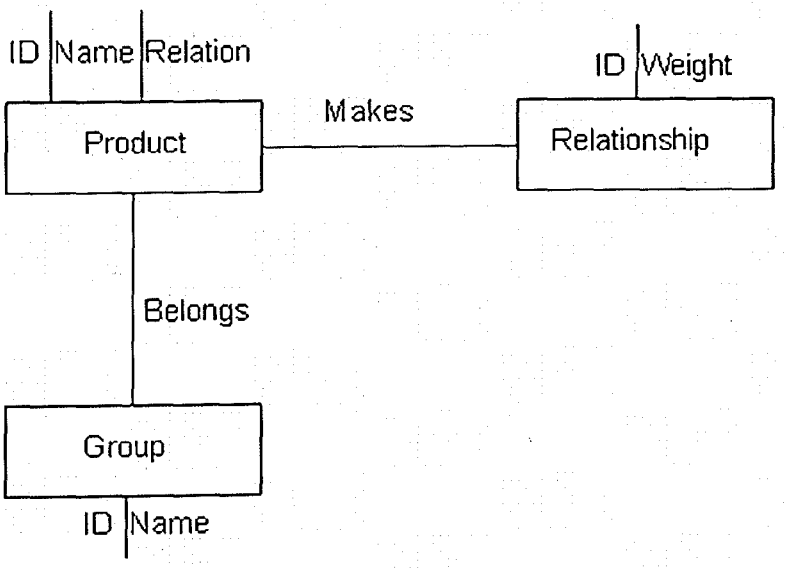


Figure 7.1 Products relationships

To enhance customer satisfaction, semantic search and semantic products must be made available to the customer and they should be easily accessible by the user. As we have explained in the previous chapter an intelligent system which includes the semantic database where the data is indexed and stored, is necessary for efficient retrieval of data. Indeed in the previous chapters we have been explaining about the semantic information, semantic storage of information, the semantic link between data and semantic retrieval of data. The real time semantic information search and discovery are mainly dependant on all these components.

In order to evaluate the approach a case study based on an E-commerce business application is constructed and considered. The choice of the type of business is of secondary importance, since the approach is generic enough to accommodate a variety of E-commerce systems, and all what we need is an application that can be used to demonstrate the benefits of the approach and as a proof of concept. However, in the previous chapters, particularly in chapters four and six the examples chosen were mainly from the retail sector. Therefore, as a case study a retail business that is dealing with clothing is considered. Hence, the list of products will present typical clothing business items such as shirts, dresses, suits etc. as shown in figure 7.2.

Figure 7.2 Simple product list

7.4 Application of the Semantic Approach

In this new approach there are three types of semantic terms that can be considered in a product list. The first type being that the semantic terms are defined and fixed by the business. Secondly, there are optional and negotiable terms i.e. they are defined by the business as well as degree of usage. Finally, the terms are defined according to usage. Thus, the product list can be modified and updated, including inserting or deleting products, either manually by the business and/or automatically according to usage. These automatic updates will take place according to the queries made by the customers. The customer will be able to find the products efficiently because of the semantic connections of products and queries. The customer can therefore submit any terms for locating searched products. The search result

will return all the semantically defined relationships of products. In this new approach the user makes the query and the results will be displayed in the semantic display (see section 7.5.4).

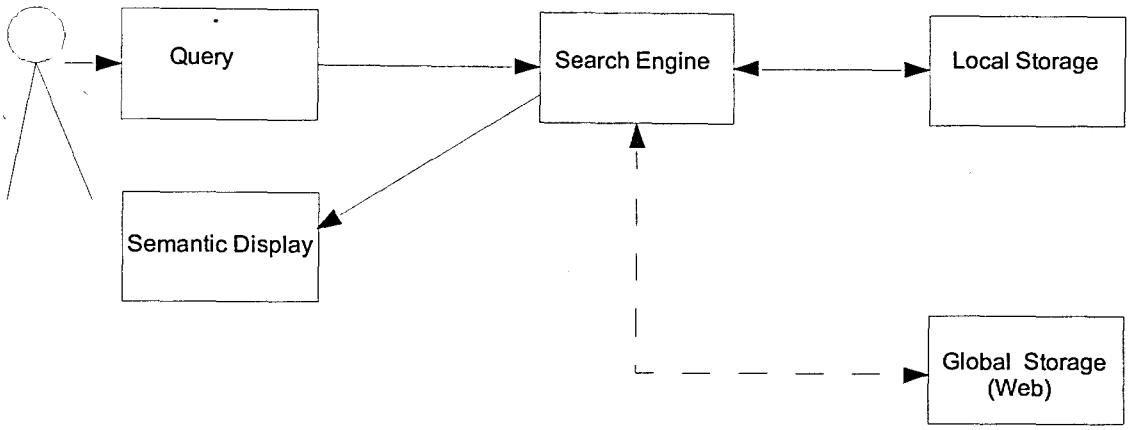


Figure 7.3 The high level view of the system

7.4.1 The Process

The whole process of the new approach starts with the initial query by the user. In order to search for a 'product/list,' the user makes the query in real-time in the query interface of the system. The search begins assisted by the intelligent engine "Search Engine" as shown in the figure 7.3. The search engine looks for some specific information first in the local storage. If the information is not found the searching process looks for information globally i.e. using global storage. Finally, answers to the initial query are displayed semantically that is 'semantic display', according to the ranking of information together with any additional semantic information. Searching locally and globally using semantic search intelligent engine is carried out as discussed in chapters five and six i.e. the indexing and retrieval of semantic information according to this new approach can be applied to two categories. Firstly it can be applied locally, and secondly the search engine techniques are used to retrieve information globally within an enterprise or sector.

The relationships between products and their strength are defined using the approach presented in Chapter 6. However, as the number of products grow storing their relationships may become an issue, therefore relationships weights are updated dynamically according to usage, whether the weights were initially defined or not. This is achieved by linearly increasing or decreasing the weights according to usage. Moreover, if usage drops the weight might become static, which is allowed for a period of time, after which it is decreased in

proportion with the time scale defined by the system. Finally, a threshold weight is set below which it is assumed that no relationships exist. In addition the current implementation takes into account personal occurrence that is based on personal choice i.e. the occurrence weight is not only based on the common occurrence that was generated by many users as was done in chapter six. Therefore, the occurrence weight defined in equation 6.1 is now influenced by two factors, the O_p representing personal choice, i.e. the personal occurrences of entities services etc. and their frequency, and O_c representing common occurrences by many users. In the absence of personal occurrences (choices) only the common occurrences are taken into account. However, as the frequency of personal occurrences (queries) increases the contribution of O_p also increases automatically. This follow the same relationship between the user defined weight and occurrences as shown in equation 6.1. It is expected that personal choices and their frequencies to be available either on the user/customer local environment or in the user profile which can be stored within the E-commerce system.

7.4.2 Semantic Storage.

The storage and retrieval of semantic information is handled in a similar way using local and global enterprise storage. That is based on storing the products, products groups and their relationships using an entity relationships diagram which is, in its simplest form, shown in figure 7.1. Thus, information about products, the relationships between products, association a product with a group, groups of products and group relationships are all recorded and stored. It is expected that products belonging to the same group will have strong relationships.

7.4.2.1 Local Storage

The local storage in the chosen business scenario (retail clothing business) is used to store the local information needed by the business which is primarily the product list. This gives the business the ability to directly assigning the products relationships and their strengths at input stage; clearly only if this brings benefits to the system. Although such an assignment is subjective it should be based the business's experience and history, and its marketing strategy. This will allow the flexibility to deliberately direct users/consumers toward products, which is similar, although more powerful, to what businesses for example supermarkets use to display their products.

However, still the approach discussed earlier will be applied which will take into account the occurrences, but the local business has the ability to manipulate the various weights that contribute to the total strength of the products relationships. Thus products relationships and links are constructed by following pre-established links. The user sends a request and query answers are sent in the form of semantic information. The relationships between the products used in the query are determined by the weighted values which will express the semanticity between them.

7.4.2.2 Global Enterprise Storage

In such a business environment clearly queries are initially directed to local storage that is local business, which is very much driven by the type of business to be carried out. However, if products are not available in the local store it is commonsensical to extend the search across the enterprise. A logical extension is to allow global products, on the web, to be available to local customers. This might take the form of B2B or simply directing customers to other E-commerce sites which is the approach adopted in this work. Obviously, such directions are expected to bring commercial benefits to the local business. The proposed framework can accommodate the global extension as discussed in chapters five and six. In this case the meaningful pre-established links are joined to create the global links, which are established throughout the web.

Global links can be established according to users' queries and these links then respond to a particular query submission which includes keyword, graphical image etc. These links are then automatically indexed and can be stored locally. In this new approach the storage problem can be solved by linking connection numbers between the products/groups of products. The semantic connections established by the new approach makes the system uniquely different from any others in that the existing search engine approaches present all the links connected with say a keyword. However, these connections might or might not be semantic, in fact many, if not the majority, of connections do not present any meaningful relationships to the searched keyword.

7.5 Evaluation

As mentioned in the introduction of this chapter the integration and evaluation of the new approach require the use of a number of technologies, concepts and development tools, including the use of programming and scripting languages as well as databases, many of

which were discussed and reviewed in chapters two and three. In fact, different parts of the system architecture might require different tools as shown in figure 7.4.

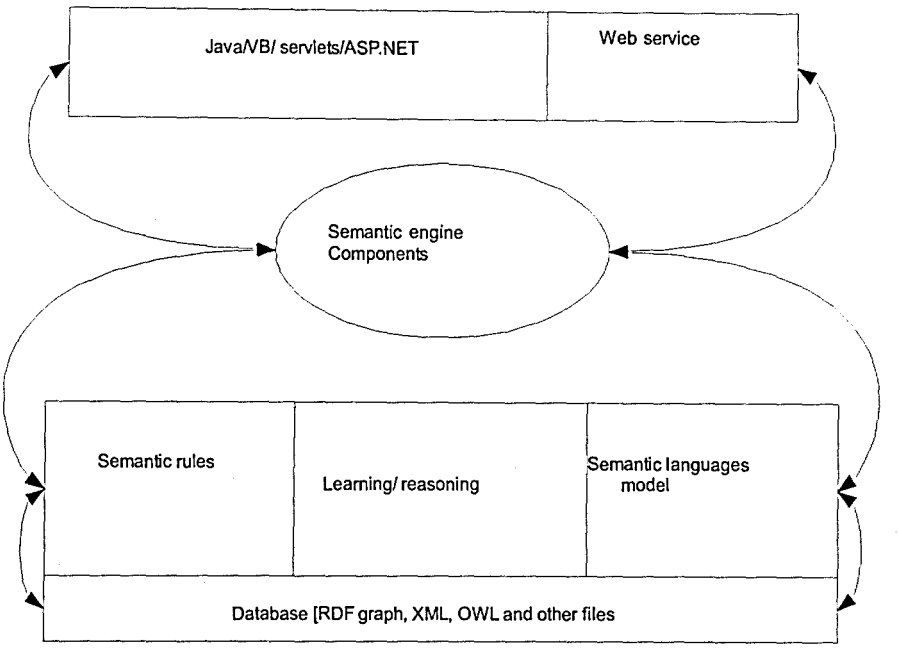


Figure 7.4 System architecture and tools

The system consists of three layers. One of the most important benefits of using such a multi-layer architecture is that the system becomes more adaptable and flexible. Furthermore, the layers can be mapped to the generic three tier architecture, discussed briefly in chapter 2, which is divided into three main components – "user interface tier", "middle tier" and "remote tier". The "user interface" tier provides a physical medium for a software solution to engage with an end-user. The "middle tier" component is responsible for business logic processes and rules incorporating business, technical and data access logics. The "remote tier" component gives a data storage and data management facility.

7.5.1 User Interface

The first layer in 7.4 is "user interface" which is a physical presentation medium, which gives a means to interact with users (i.e. doorway). This presentation medium (web browser) interprets and organises its interfaces (i.e. a presentation layout) according to instructions provided by the middle layer in the form of a standard protocol. It also delivers information entered by end-users to the other layers and returns the output to the end-user. A number of technologies can be used for this layers Java, VB.NET etc. VB.NET is being used in the work

to develop the interface to the retail business (see for example figures 7.5 and 7.6). Moreover, queries can be made either using keywords or graphical presentation of the items.

For example – a customer wants to access his account at a shopping web site. At a client side, a web browser (i.e. the "user interface" component) sends a customer login information to a web server. At the server side, the "middle tier" component accepts this incoming information and an instruction is then sent back to the user interface (e.g. invalid username and password, access has been granted, etc.)

7.5.2 Semantic Engine layer

At the core of the middle layer is the semantic engine components offering an encapsulation of business, technical and physical requirements into one viable software solution. These requirements are integrated into three different types of tiers – Presentation, Business and Data Access tiers. These tiers are able to act as an independent single unit. This layer governs how an entire solution should behave via a predefined business flow process, as well as a specific business consideration such as a provision setting of a content management, a service level agreement (SLA), etc.

Now let us consider an example suggested in the presentation tier (above). A business tier makes a query of a user account via a data access tier, which is based on a data access request received from a presentation tier. Subsequently, the business tier replies an "access denied" instruction because it has discovered that the customer has not renewed his annual subscription.

The semantic engine layer consists of semantic components such as handlers. The semantic rules supported by semantic languages and models and the learning process also takes place at this layer. The users will interact with this layer through the interactive interface. Semantic components will be responsible for receiving the request from the users and forwarding them to the next layer. The semantic engine layer is also responsible for receiving the results from the semantic layer and forwarding them to the user. In this layer VB is also used, however, since this layer is separated from the interactive interface layer, the interactive interface layer can be used with any other technologies if needed.

7.5.3 Data Storage

A "data storage" component is a physical data store medium for an "application structure" component. A database management system (DBMS) is often used for this purpose. Typically, a standard driver (gateway) such ODBC and JDBC is used to establish a connection between "data storage" and "application structure" components for a data retrieval and storage operation. The data access tier (remote tier) provides an information management facility for an "application" component. The user interface and middle tiers use this tier for a data manipulation service (e.g. a facility for the application to make a query of the stored information, which could be in the form of SQL queries). Moreover, since local and global storage can be applied in a similar manner, in this experiment only local storage is used.

The "data storage" component is where the product lists are stored and indexed semantically using XML. The XML format can help in improving the efficiency of information search and discovery. One of the advantages of storing product lists in the XML format is that it is able to hold the raw and structured data and it will enable the web users to manipulate and generate the data in accordance with their representation. The product lists can be exchanged in XML format, and it is easy to the existing database systems to manage and deal with XML. For example it is possible to make use of relational databases for storing XML documents, so that the information can be described in hierarchical, structured tree type relations which would give plenty of room for storing, querying and publishing information. The data stored according to semantic connections and the query will produce only the semantically connected product lists. This is done dynamically by learning the predefined data from the database and storing it within an entity (as explained in chapter five). Thus, the product lists are organised and semantically connected into a hierarchy of categories or classes and sub classes, (parent child relationships) and have many levels. These parent child relationships are represented by XML format. XML is capable of explaining the type of data stored in a product list. Thus the class properties are explained in a more semantic and flexible way.

Customer satisfaction is achieved by way of establishing a semantic relationship between entities (products), and thus the customers are able to identify the products efficiently. The queries are formulated according to the semantic connection and availability of products and therefore the product lists are generated accurately and semantically. Since the new approach makes use of a probabilistic network (chapter six) it can also use historical data to predict the

output of a new query and it can be added in the product list. However, the prediction is based on the strength of the connection and the query is learned only from the customers view point (entries) and of the semantic nodes that correspond to the combination of input nodes indicated by the given weighted value and the product lists are updated accordingly.

7.5.4 Semantic Display and Personalisation

In order to demonstrate visually the proposed approach a case study based on an E-commerce business application, specifically in retail clothing, as pointed out earlier is considered. Thus, a retail clothing site was constructed which emulates the real life business, however, the main focus of the demonstration is to show the benefits of semantic relationships and the potential for personalisation. The first page of the site, shown in figure 7.5, displays the business categories including men's, women's, children's fashion and sports clothing as well as their accessories, so that the customers can browse through the site and make their choices.

Figure 7.5 Home page

The site provides the customers with a login page (figure 7.6), in line with typical E-business applications, where the customers can login. The logging in process help in storing the customers' shopping habits, products and historical usage which gives the ability to personalise the site according to the customers' needs as discussed further. In summary when logging into the system, the customer leaves his/her behaviour/ history details to the system

to establish relationships for him/her. The account holder receives the personalised 'system established relationships.'

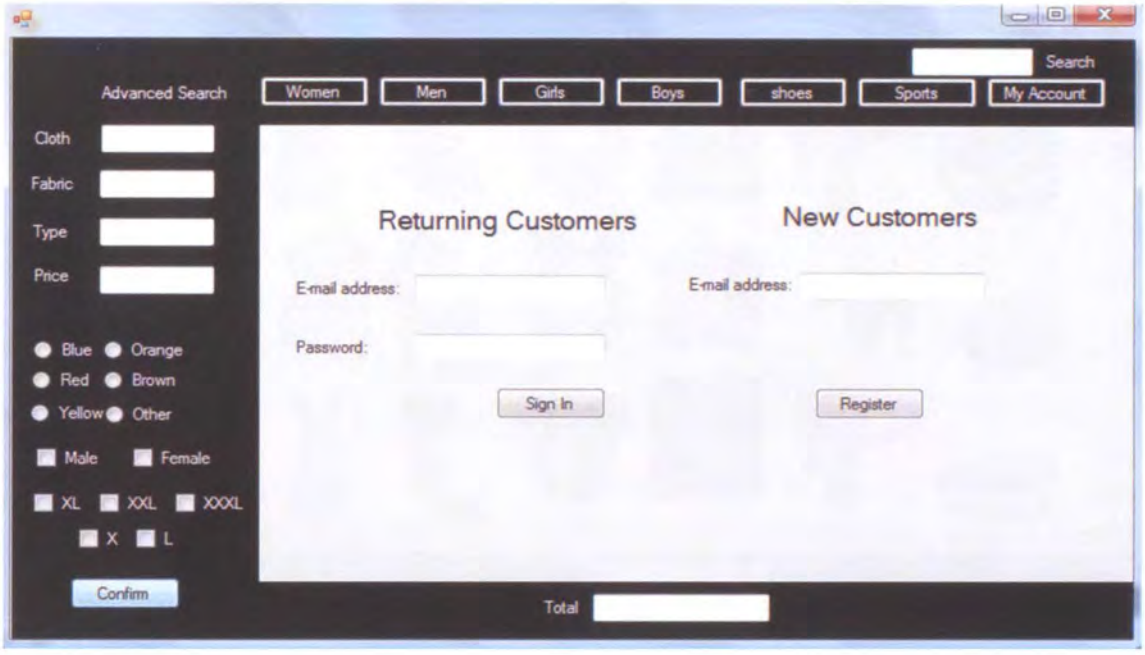


Figure 7.6 Login page

An example page showing a particular set of items such as women clothing is shown in figure 7.7. This is based on the business view of the products popularity.

Figure 7.7 Women Clothing.

Figure 7.8 Women dress query

For the purpose of demonstrating the products availability and the potential semantic relationships a 'keyword' based query is used as shown in figure 7.8. The customer searches for women dress which returns dresses with additional products that are associated with them. The dresses types as well as the relationships with other products are set by the business manager according to his/her experience in the sector. Therefore, these preferences are predefined by the manager and are displayed accordingly. Although the system is producing the highest semantic results for 'women dress' based on the weighting factors, these at this point are defined by the business. The products that are displayed nearest to the dresses represent stronger relationships.

Moreover, additional information for the query of a specific product such as 'full blouse' by selecting an item (clicking an image), can be made available as shown in figure 7.9. The approach semantically connects with the selected item and it brings forth additional information which includes the availability of items in the shop.

Figure 7.9 Additional information

In order to demonstrate the impact of usage on semantic relationships the same query 'women dress' is made and the products as well as the relationships are displayed according to both usage (occurrence) and predefined weight entered by the manager. This is demonstrated in figures (7.10 – 7.13) which show that the semanticity of products vary according to its demand. The further experiments show that the system establishes relationships through the semantic strength in relation to the usage. In these examples the system re-arranges the shop according to the popularity of products (usage) and their relationships. Indeed the query 'Dress + female' finds out other products with the highest semantic strength with 'Dress+female'.

As mentioned earlier the products that are displayed nearest to the dresses represent stronger relationships, therefore this is represented by 'bag' in figure 7.10 which according to usage has the strongest relationship to the 'dress' which is followed by 'shoe' etc. The page shows that bags are in demand, when buying dresses, more than other items and therefore bag is semantically building up close relationship with 'Dress.'

Figure 7.10 Dress semantic relationships (I)

As can be seen from figure 7.10 the order of displayed items reflect their semantic strength to the item 'dress', in this case 'bag' has the highest strength (weight) followed by 'shoe', 'necklace'. 'glass' and finally 'belt'. It is important to mention that this is order is based on occurrences and usage i.e. it is based on system decision, which is also applied to figures 7.11 and 7.12 where the order of displayed items is changing. However, considering figure 7.10 again, the human logic might imply that 'belt' should have a stronger relationship to 'dress' than say 'necklace' and 'glass' but the usage implies otherwise. This is the reason that the proposed approach gives the option to user/manager to interfere and apply their business logic by modifying the weight of the relationships according to the business needs.

As the system is now driven by demand and as the demand changes so does the order of products as shown in figure 7.11. The page shows bag, shoes maintaining the same strength of relationship with 'Dress' because they are highest in demand still. Whereas, sunglasses have taken the place of necklaces, which means necklaces are not in demand by the customer in this combination.

Figure 7.11 Dress semantic relationships (II)

Figure 7.12 Dress semantic relationships (III)

Next figure (7.12) shows bag, shoes, sunglass maintaining the same strength of relationship with 'Dress', however belts are highest in demand by the customers and therefore belts take the place of necklaces as they are not in demand by the customer.

Figure 7.13 Dress semantic relationships (IV)

Finally, figure 7.13 show that bag, shoes, sunglasses and belts maintaining the same strength of relationship with 'Dress' but necklaces are no longer in demand by the customer and therefore they have gone down the history whereas a new item 'gloves' replaces necklaces.

Apart from indentifying the relationships between products which can be very helpful for the customers, the approach can be considered for the personalisation of the customer choice which primarily based on their historical preferences but also can be influenced by the global customers' usage and the decision of business manager. There are two given pages (figures 7. 14 and 7.15) which represent some personal choices and needs such as designer and oversize clothing.

Figure 7.14 Designer clothing

Figure 7.14 shows that using historical preference the system establishes strong semantic strength between various designer products, therefore the page "shop" is re-organised to display only the designer's clothes. i.e. the system generates results semantically and personalises the pages. The same applies in figure 7.15 when the customers preferences/needs is for oversize clothing which also is organised following the same principles.

Figure 7.15 Oversize clothing

7.6 Observation

Having considered the visual output, in the form of semantic display, obtained from the application of the proposed approach, in this section some observations regarding the semantic relationships including the personalisation factors are considered. Thus, the experimentation will follow the strategy applied in chapter 6, but with the aim is to produce an evaluation of on E-commerce scenarios as discussed earlier in the chapter. Moreover, the observations in terms of relationships weight represent the changes shown in the semantic display as shown in the previous section.

The starting point of the experiment is with the business manager defining the products to be displayed and offered to the costumers as well as their relationships with other products. This process is subjective and is which is based on any previous retail experience the business manager might have. At the early stage there is very little historical customer experience, therefore the weighting factor is purely fixed and defined by the business. Thus, an initial query such as "women dress" (figure7.8) produces a relationship weight between products as set by the business manager as shown in figure 7.15.

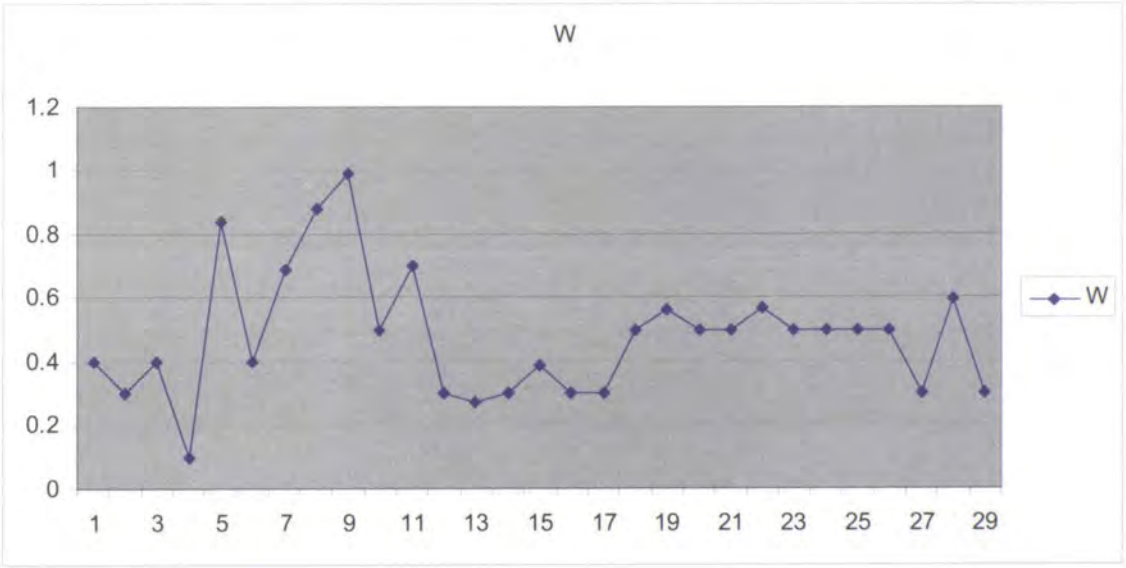


Figure 7.16 Fixed and defined relationships

However, this changes as demand for particular products grow and when usage is taken into account with products being ranked according to their combined weight that is based on usage and business decision as shown in (figures 7.16 – 7.18).

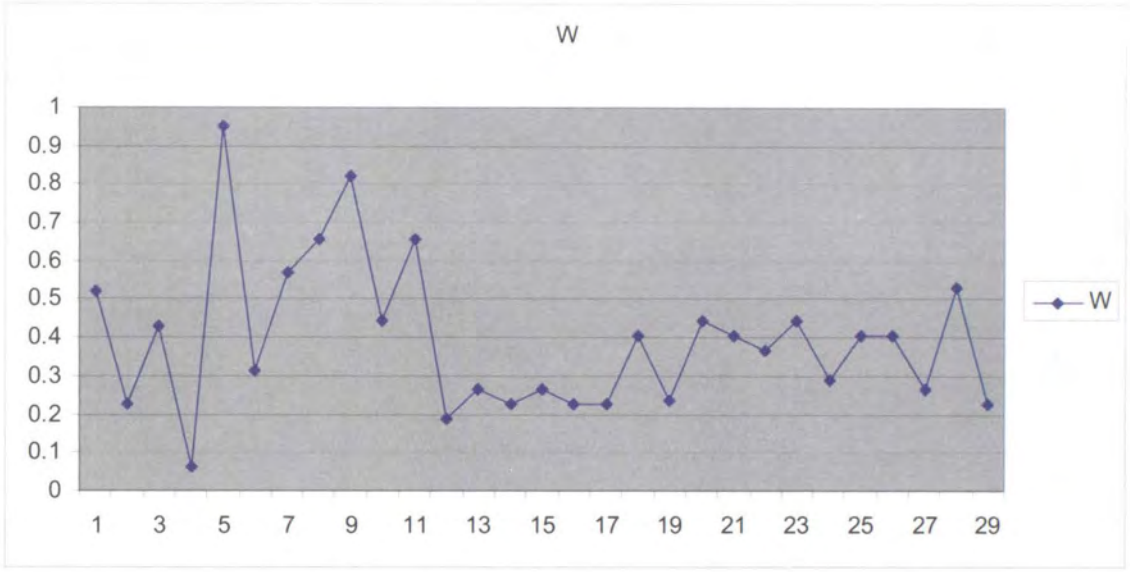


Figure 7.17 Varied relationships based on usage and business defined weight (I)

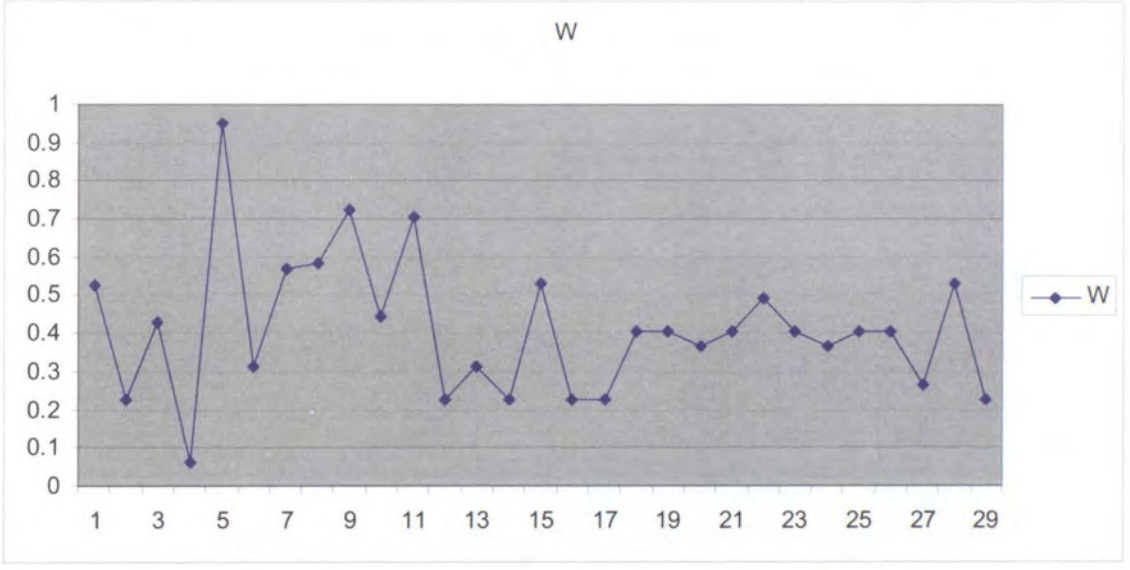


Figure 7.18 Varied relationships based on usage and business defined weight (II)

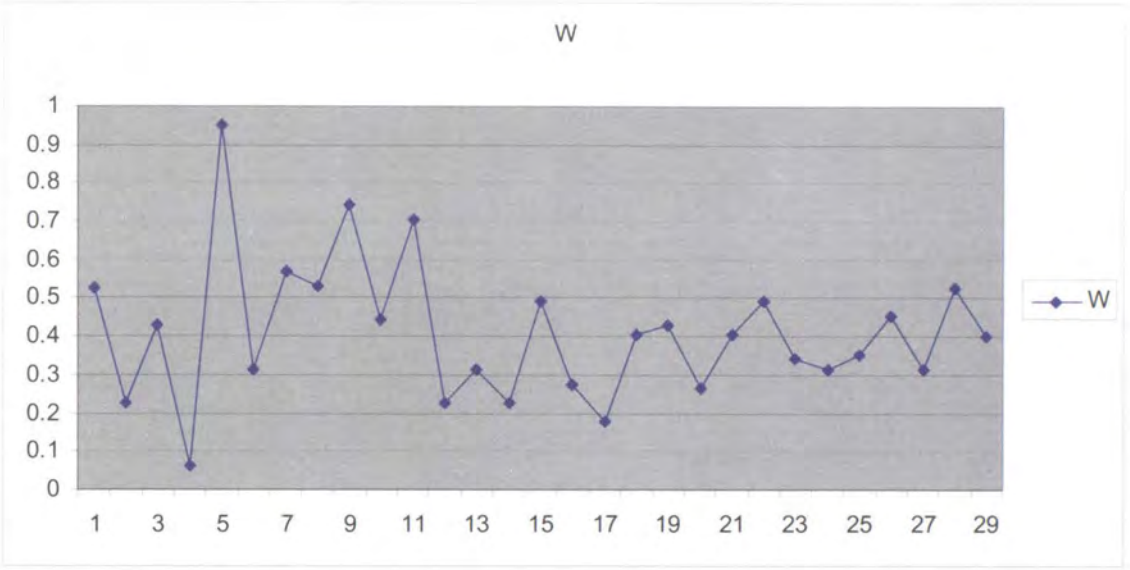


Figure 7.19 Varied relationships based on usage and business defined weight (III)

Moreover, according to usage eventually certain items weight will drop, so they attract a low ranking and they are not even shown on the screen which is represented in figure 7.19.

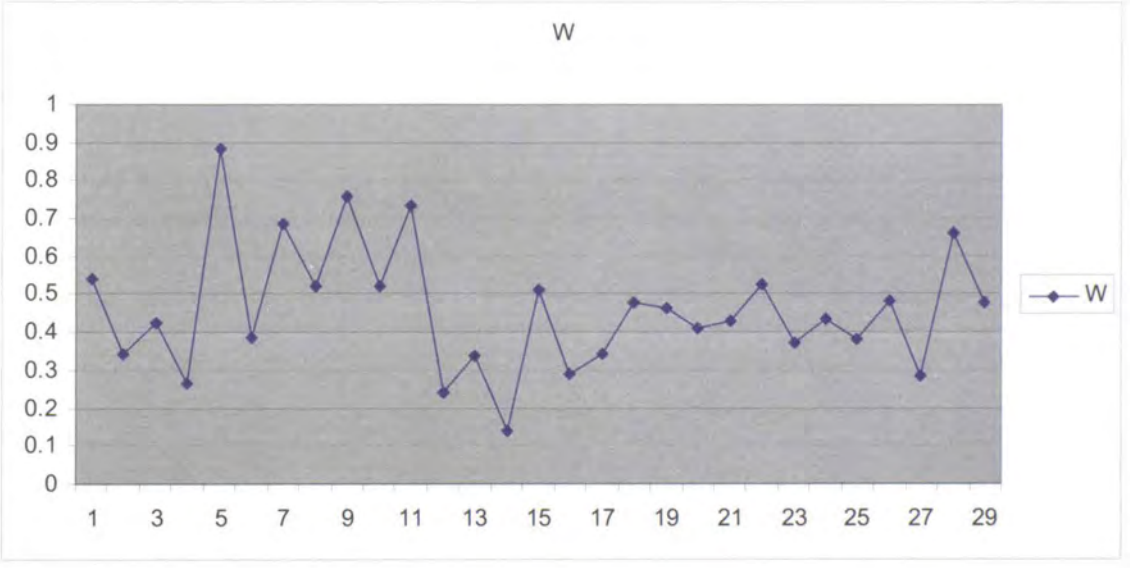


Figure 7.20 Varied relationships based on usage and business defined weight (IV)

The next set of experiments is based on usage only i.e. the influence of the business choice is removed completely as shown in figure 7.20.

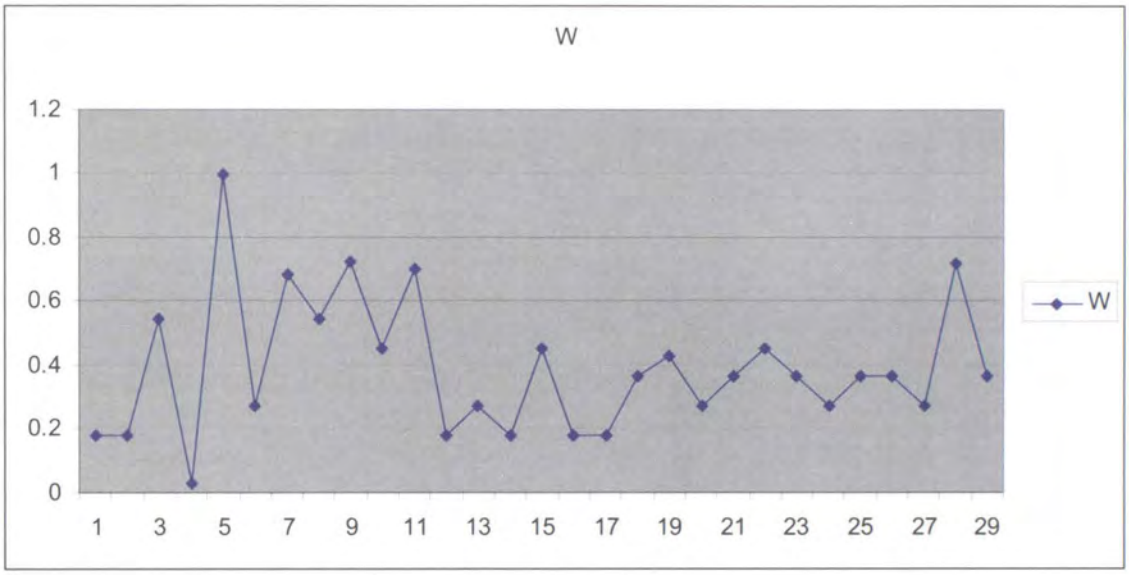


Figure 7.21 Varied relationships based on usage

Moreover, some results showing some of the personalisation aspects including designer as well as oversize products are shown in figure 7.21 and figure 7.22.

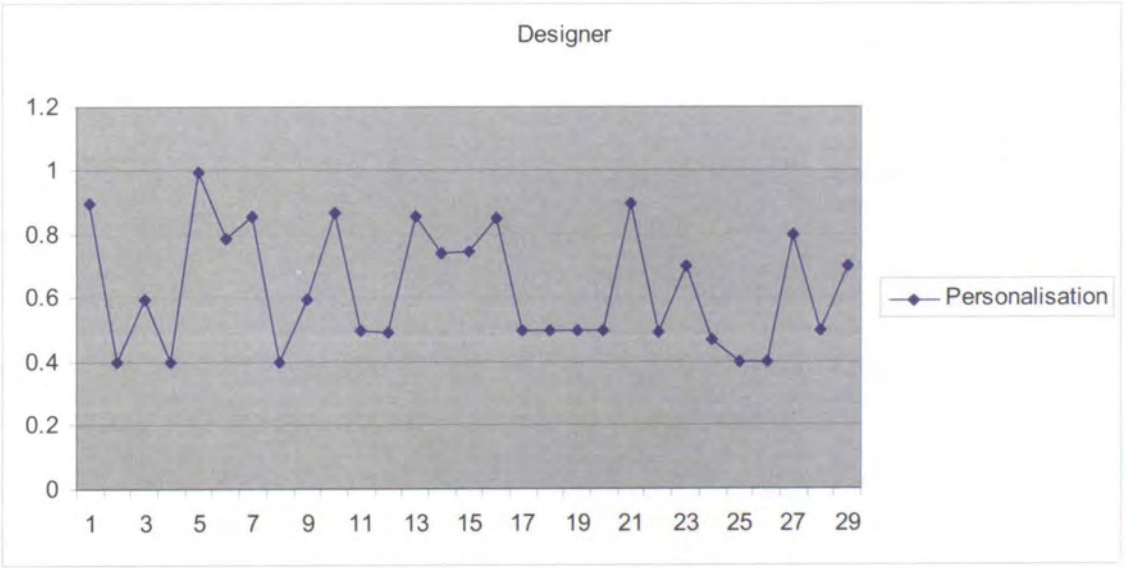


Figure 7.22 Personalisation according to design

Figure 7.23 Personalisation according to size

Finally, some comparison between the initial manager choice and final usage driven relationships are shown in figure 7.23 and figure 7.24 correspondently.

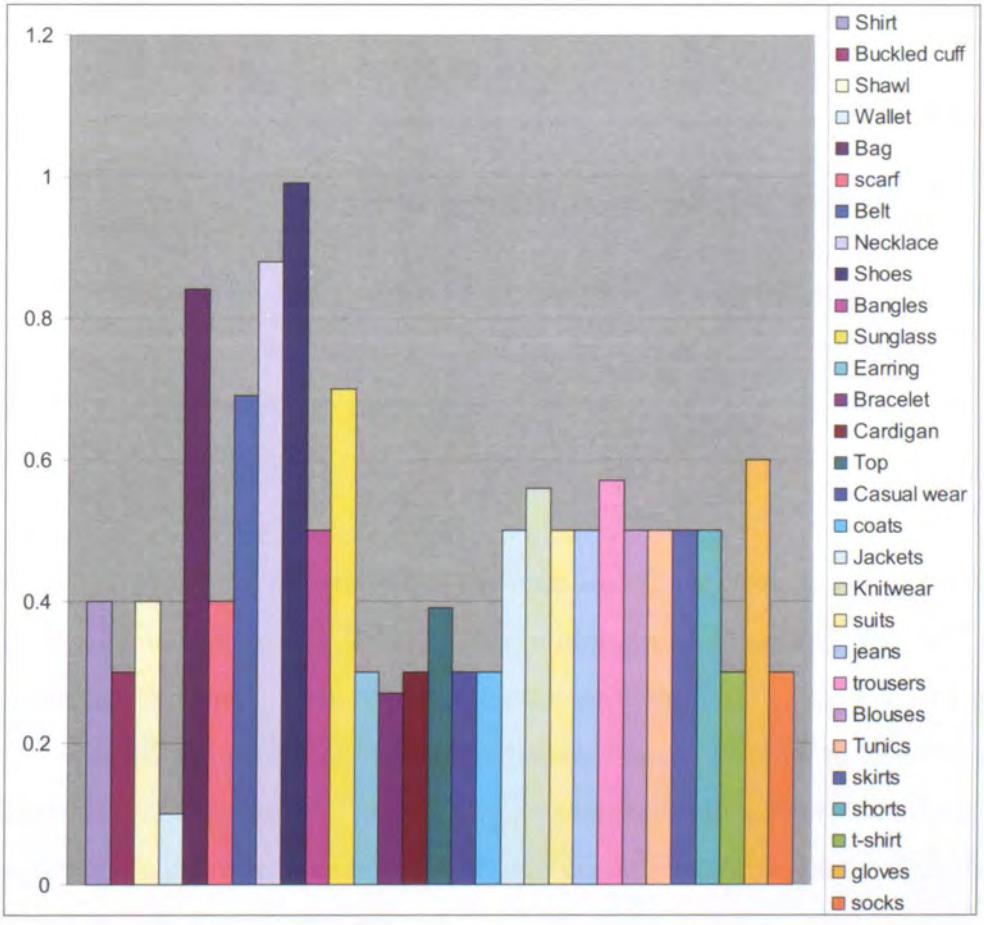


Figure 7.24 The pictorial presentation according to business choice

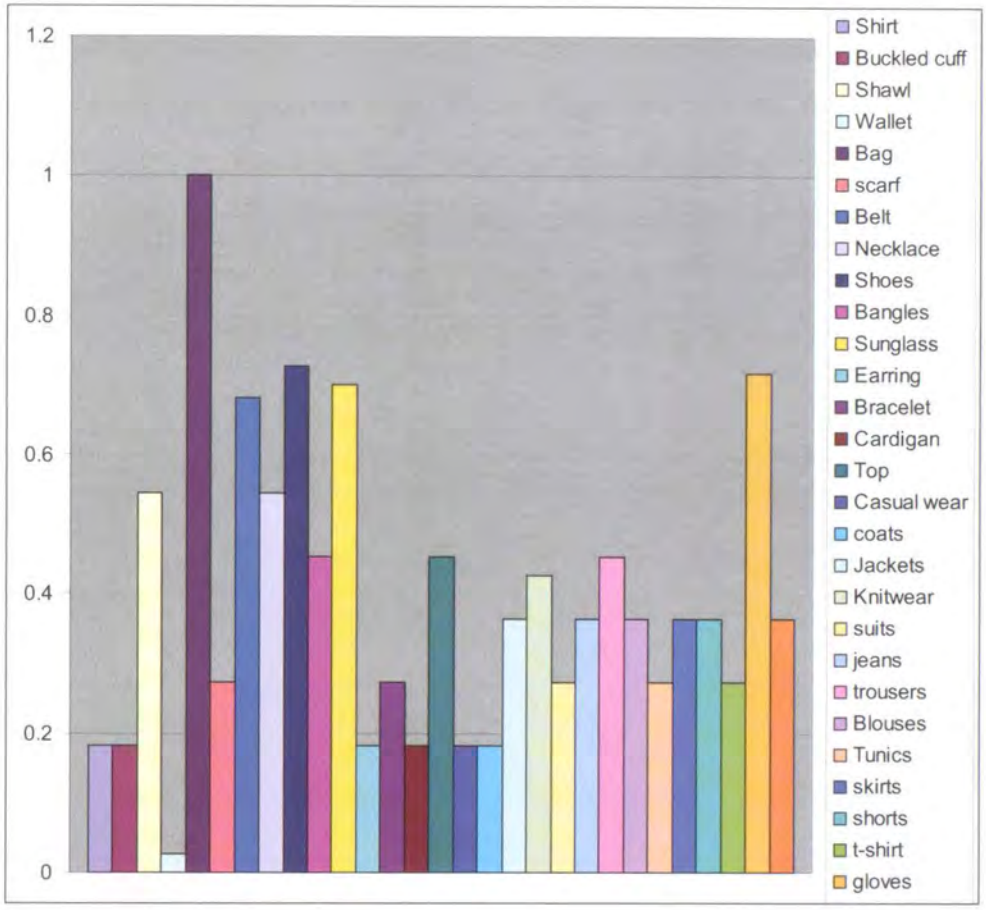


Figure 7.25 The pictorial presentation according to usage only

In summary the approach has performed in line with the initial evaluation in chapter 6, however, on a more specific business scenarios. This also shows that semantic data can be retrieved, discovered and ranked and displayed semantically.

7.7 Conclusion

This chapter presented the integration of the proposed framework and architecture that were proposed in chapters five and six, together with semantic storage, semantic search and semantic display concepts. This required the use of a number of technologies and concepts that were discussed and reviewed in chapters two and three. In order to evaluate the potential of the approach a discussion on E-commerce as a background was presented and then a case study, chosen from the retail business sector, was constructed to support and evaluate the new approach and to demonstrate its potential.

Adapting the new approach for the current application for information searches is relatively straightforward. The results show that the approach can establish entity relationships semantically and it can prepare a semantic search within a business or an organisation over the web. By using this approach, semantic data can be retrieved, discovered and displayed accurately and semantically. Therefore there are clear benefits from using the approach not only for businesses but also for the customers/end users. Moreover, although the approach was evaluated for a particular E-commerce application it is generic enough to be used in other applications and online businesses and it can serve as a frame work for designing future semantic websites.

Chapter 8 – Conclusions and Future Work

8.1 Conclusion

The primary aims and contributions of this thesis are the design, construction, and evaluation of a new framework for semantic information search and discovery in E-commerce oriented applications. The work which was carried out to achieve these aims, as well as the results of the work, was reported in the previous chapters of this thesis. In this section the achievements and conclusions which have been previously drawn will be summarized.

The research began by a background overview of internet and website technologies and their applications particularly in E-commerce. This includes reviewing the internet technologies network structure and protocols, mark-up languages, the three tier systems, databases, the different types of websites etc. A particular attention was paid to E-commerce systems and their design aspects which require careful planning and development if they are to be achievable in terms of semantic, real-time information requirements. Moreover, some of the issues in web search, information overload etc. were identified, the existent methodologies that are being used in information searches were briefly reviewed and the concepts of semantic information search were considered.

Next, the thesis has investigated the meaning of the term "semantic" in an attempt to present some definitions and characterisations of semantic web. Therefore, it has presented the historical evolution of the word "semantic" and semantic web. After considering the architectural issues in semantic web, several knowledge representation languages were presented and their use in semantic web and their associated processes were examined. Moreover, the review continues with architectural issues particularly in term of web services, their interoperability problems and the potential semantic links of web services. Overall, the literature review provided some background, in terms of basic concepts, definitions and technologies to the work presented in this thesis. The review also identified some of the issues with existing approaches, particularly in information search and linkage, which can affect the quality of service provided by online systems.

The quality aspects and issues of common websites, particularly E-commerce sites, were considered, the limitations of some of the existing approaches were discussed and a quality model, with a number of quality factors, was proposed. The model defines some fundamental factors that influence and maintain the quality of websites, therefore impacting their users/customers satisfaction. Indeed, the model evaluation, using a number of businesses with active E-commerce systems, has produced a variety of responses which showed satisfactions in some areas but not in others; however some overall satisfaction/un-satisfaction trends were identified. After considering the quality factors at the highest level, the specific characteristics of the some of the factors together with the potential sites structures were also considered. An evaluation using the identified characteristics were conducted which has shown the impact of the structure and layers of websites on some of the quality characteristics. Thus, improving the quality factors/characteristics can contribute towards creating and modeling an effective semantic E-commerce system.

Furthermore, some specific aspects of semantic real-time information search were considered, and an indexing semantic architecture that discovers and integrates information using self organising hyperlinks was proposed, which can help in locating semantic information more accurately. The semantic real-time information search approach is based on a powerful semantic frame work and architecture that can automatically self organise and establish relationships according to the semantic search, among semantic content, and thus creates a semantic network of hyperlinks. There are a number of other features associated with this new approach. A key feature is that a customer/user can easily search for information, just by clicking the object (product) itself. Each product has additional information associated with it which also can be ranked. Moreover, the approach will enable complex queries, thus allowing the search for complex answers on each product. This information could be used to rearrange the site according to user requirements which will enhance the user experience.

However, the semantic architecture requires the application of some artificial intelligence techniques for the establishment of semantic relationships of different entities. Thus, an intelligent engine in the semantic layer was proposed together with the choice of a network topology for its design. The investigation into probabilistic networks for semantic real-time information search and discovery was focused on the weighting factors that define the strength of the semantic relationships between different entities. The presented approach

takes into account both system determined and user defined weights, as well as the ratio between the two weights in its determination of the total weight that defines the strength of the semantic relationships. The user defined weight can be applied when a local E-commerce business, with its products and services and its own data storage, needs to define the relationships between its products/services according to its business logic and the potential profits, which can be done by setting the relationships at input stage (or even get them modified at a later stage), but which also can be modified automatically according to usage. The probabilistic networks of the engine together with the ways how the connection weights are calculated were illustrated and a number of experiments to evaluate the approach were conducted. The conducted experiments show the impact on the total weight of the individual weights, it also show that results from system only determined weight and user defined weight can vary considerably, therefore a combination of both was recommended, however, the final decision should be driven by the business case and the potential benefits.

Finally, the integration and evaluation of the framework and architecture together with the indexing, semantic storage, semantic search and semantic display concepts, was presented. This started with a discussion of E-commerce systems and the requirement for product E-catalogues/lists, then, it carried on with a description of the components of the system. An experimental study with the aim to evaluate the system using a case study based on a retail E-commerce business application was conducted, which highlighted the significant benefits of the new approach.

8.2 Future Work

In this section we give suggestions about how the work presented in this thesis can be carried out further. The future work which can presently be seen in this field may be classified into a number of categories.

Firstly, it should be noted that this research study has focused on the design and construction of a framework for semantic information search and discovery, and although some implementation and evaluation were carried out as proof of concept, there is certainly a need for its integration and evaluation in a large scale and complex real time dynamic web environment. However, the obtained results suggest that such a test could provide a further proof of the proposed framework. Moreover, when fully implemented, the proposed

framework should be able to be operated dynamically and automatically to semantically link information while providing aspects of personalisation in the form of semantic display.

Secondly, although in this thesis the framework and its application was focused on E-commerce systems, the approach is designed to be generic therefore it can be applied in a number of areas which can be specific applications such as those in health informatics, GIS etc. or more generic applications such as knowledge management, search engines etc.

Thirdly, there are a number of further advancements which can be made to the approach, particularly in adopting the distributing computing paradigm. During the course of this work it was clear that using the probabilistic network (or any other intelligent engine) to work out the many to many semantic relationships, and their strength, between entities/keywords even within a specific domain requires a huge amount of data to be processed in order to produce meaningful information. Therefore, it will be essential for data processing to be enhanced so that it takes as little time as possible. One of the best ways of performance enhancement is distributing the processes associated with each request across a number of servers, instead of doing huge processing on a single one. Therefore, designing distributed semantic web applications so that the large scale processes can be distributed among a number of servers will be very important. This, however, requires the development of efficient strategies/algorithms for web process allocation, synchronization and data distribution among a number of servers according to the nature of semantic web applications.

Finally, having proposed a programmed network topology to identify the relationships between entities/keywords and quantify the weight of each relationship, a further extension could include a similar approach using a neural network with self adapting topology and weight calculations automatically with automatic connection on a global scale. The almost deterministic approach proposed can be used to verify (as measurement) the work of neural network. The neural network can be used to indemnify ontologies automatically as well as potential relationships. Moreover, other data mining and AI techniques can be investigated for their suitability and applicability in the intelligent engine of the proposed architecture which might lead to the design of other and perhaps more efficient strategies to intelligently navigate the massive data available on the web in a useful, fast, efficient and meaningful manner.

References

(Aar02) Aaron Swartz, *"IEEE Intelligent systems, ed James handler"* (university of Maryland) in journal of The semantic web, 2002

(Aar01) Swartz, A. and Hendler, J., "*The Semantic Web: A Network of Content for the Digital City"* Proceedings Second Annual Digital Cities Workshop, Kyoto, Japan, October, 2001

(Aar00) Aaron Marcus et.al., "*Cultural Dimensions and Global Web User-Interface Design: What? So What? Now What?"* 6th Conference on Human Factors and the Web in Austin, Texas, 19 June 2000

(Abu85) Abu-Mostafa, Y., St. Jacques, J.,*"Information capacity of the Hopfield model" IEEE Trans. Inf. Theory*, vol.7, pp.1-11, 1985

(Ado05) Adobe.com, "*Lesson 3.1 Site Design"* [Online] Available at

<http://www.adobe.com/education/webtech/CS/unit_planning2/sd_hierarchical_id.htm>[Accessed 10th Nov 2005]

(Ala01) Alain Durand, *"Deploying IPv6, IEEE Internet computing"* Volume 5, number 1 January/February 2001

(Alb09) Alberto Fernandez et.al, "*Closing the Service Discovery Gap by Collaborative Tagging and Clustering Techniques*" National University of Ireland, Galway, 2009

(Alt08) [Online] Available at <http://www.altova.com/semantic_web.html> [Accessed on 25th Nov 2009]

(And08) Andy Mulholland, "*The Cloud and SOA Creating an Architecture for Today and for the Future"* Capgemini [Online] Available at<http://www.hp.com/hpinfo/analystrelations/wp_cloudcomputing_soa_capgemini_hp.pdf>[Accessed on 19th April 2009]

(Arm05) Armin Haller et.al, "*WSMX –A Semantic Service-Oriented Architecture*" *Proc. of the 3rd Int. Conf. on Web Services*, pages 321 – 328. IEEE Computer Society, 2005.

(Ant08) Antti Nurminen, "*Mobile 3D city maps*" IEEE Computer Graphics and Applications, volume 28, number 4, pages 2031, 2008

(Ari05) library.arizona.edu [Online] Available at <http://www.library.arizona.edu/help/>[Accessed 17th Oct 2005]

(Ant04) Antoniou G. & Harmelen et.al, " *A semantic web primer"* The MIT Press, 2004

(Ant05) Antonia Stefani et.al, " *In-depth analysis of selected topics related to the quality assessment of e-commerce systems"* Proceedings of the 2nd IEEE International Conference on E-Business and Telecommunication Networks, ICETE UK, 2005

(Apo05) Apostolos Dollas, "*An Open TCP/IP Core for Reconfigurable Logic*" Proceedings of the 13th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM'05), 2005

(Ari09) Arie Ben-David and Eibe Frank, "*Accuracy of machine learning models versus "hand crafted" expert systems - A credit scoring case study"* Expert Systems with Applications: An International Journal, Volume 36 , Issue 3, Pages: 5264-5271, Pergamon Press, Inc. Tarrytown, NY, USA, 2009

(Bal05) Baldoni et.al, "*Personalisation For The Semantic Web, Lecture Notes in Computer Science: Reasoning Web"* First International Summer School 2005 Vol 3564 2005

(Bai08) Bailin Adam, "*Website evaluation, usage and analysis"* COI, [Online] Available at< http://www.coi.gov.uk/documents/guidance/website-evaluation-usage-and-analysis.pdf > [Accessed 25th Jan 2009]

(Bar97) Barry M. Leiner et.al, "*The Past and Future History of the INTERNET*" February 1997/Vol. 40, No. 2 COMMUNICATIONS OF THE ACM, 1997

(Bar01) Barbara S. Chaparro, "*Top Ten Mistakes of Shopping Cart Design. Internet working 2001"* [Online] Available at<http://www.internettg.org/newsletter/dec01/article_chaparro.html>[Accese d 5th Oct 2005]

(Bel03) Belatreche, A., Maguire, L. P., McGinnity, M., & Wu, Q. X, " *A method for supervised training of spiking neural networks*" Proceedings of IEEE Cybernetics Intelligence--Challenges and Advances (CICA) 2003 (pp. 39-44). Piscataway, NJ: IEEE, 2003

(Ben05) J Ben Schafer, Joseph A. Konstan, John Riedl, "*E-Commerce Recommendation Applications*" [Online] Available at<http://www.grouplens.org/papers/pdf/ECRA.pdf>[Accessed 25th Jan 2009]

(Ben03) Ben Shepherd, "*What to Expect in PHP 5.0"* [Online] Available at <http://www.devarticles.com/c/a/PHP/What-to-Expect-in-PHP-5.0/1/> [Accessed on 5th Jan 2005]

(Ber06) Tim Berners-Lee, "*Artificial Intelligence and the Semantic Web: AAAI2006 Keynote"* [Online] Available at <http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html>[Accessed 17th Dec 2007]

(Ber05) Tim Berners-Lee, "*WWW2005 Keynote" [Online] Available at*

<http://www.w3.org/2005/Talks/0511-keynote-tbl/>[Accessed on 19th Jan 2007]

(Ber05) Tim Berners Lee et. Al, *"IETF RFC3986 - Uniform Resource Identifiers Generic Syntax"* [Online] Available at< http://www.rfc-archive.org/getrfc.php?rfc=3986> [Accessed 29th May 2008]

(Ber01) Tim Berners Lee et. al *"The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities"* [Online] Available at <http://www.sciam.com/article.cfm >[Accessed 29th May 2004]

(Ber01) Tim Berners Lee et. al, *"The Semantic Web"* Scientific American: Feature Article: The Semantic Web: May 2001 *.[Online]* Available at *<http://wwwsop.inria.fr/acacia/cours/essi2006/Scientific%20American/FeatureArticleThe/Semanti/Web/May/2001.pdf* >[Accessed 17th Nov 2005]

(Ber01) Tim Berners Lee et.al, *"The Semantic Web,* Scientific American, 284 (2001), 34-43"

(Ber 98) Tim Berners Lee, *"Frequently asked questions"* W3C website [Online] Available at <http://www.W3.org/people/Berners-Lee/FAQ.html >[Accessed 29th May 2004]

(Ber94) Tim Berners-Lee and Masinter L, *"IETF RFC1738 -Uniform Resource Locators (URL): Generic Syntax"* [Online] Available at <http://www.rfc-archive.org/getrfc.php?rfc=1738> [Accessed on 29th May 2004]

(Ber94) Tim Berners-Lee, et.al, *"The World-Wide Web"* Communications of the ACM, 37(8):76-82, 1994

(Ber99) Tim Berners-Lee and Mark Fishetti, *"Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web"* Harper San Francisco, 1999

(Ber96) Tim Berners-Lee, *"The Web Maestro: An interview with Tim Berners-Lee"* Technology Review (July), 1996

(Ber96) Tim Berners-Lee, *"WWW: Past, Present, and Future"* IEEE Computer 29, 10 (Oct), 69-77, 1996

(Ber92) Tim Berners-Lee et.al, *"World-Wide Web: The Information Universe"* Electronic Networking: Research, Applications, and Policy, 1(2):52-55, 1992

(Bob03) Bob Baxely *"Universal model of a user interface"* In: Proceedings of DUX03: Designing for User Experiences, ACM, 2003. pp. 1-14, 2003

(Bou04) P. Bouquet et.al, *"Querying the Semantic Web: a new approach"* Department of Information and CommunicationTechnology – University of Trento, Italy [Online] Available at <http://semanticweb.deit.univpm.it/swap2004/cameraready/scoz.pdf>[Accessed 29th May 2004]

(Bra00) Brajnik, G. *"Automatic Web Usability Evaluation: What Needs to Be Done?"* In P. Kortum and E. Kunzinger (Eds.), Proceedings of the 6th Conference on Human Factors and the Web (HF&W 2000), 2000

(Bro02) Broekestra et.al, "*Sesame: A Generic Architecture For Storing And Querying RDF And RDF Schema:*" In Proceedings of the Semantic Web, ISWC 2002, First International Semantic Web Conference, Italy, Vol. 2342/2002.

(Bri10) Brian K. Walker et.al, "*Trends 2010: eCommerce Platform And Technology eCommerce Technology And Business Tools Continue To Advance And eBusiness Leaders Must Prepare To Use Them Effectively"* Forrester Research, Inc., 400 Technology Square, Cambridge, MA, USA, 2010

(Bri09) Britannica Online Encyclopedia [Online] Available at <http://www.britannica.com> [Accessed on 9th April 2007]

(Bri04) [Online] Available at <http://www.brighton.co.uk> [Accessed on 5th May 2005]

(Bry 00) Boston Consulting Group, "*The State of Online Shopping*" Boston. Brynjolfsson, E., & Kahin, B. (2000) *Introduction. In E.* Brynjolfsson & B. Kahin (Eds.), *Understanding the digital economy: Data, tools, and research* (pp. 1-10). Cambridge, MA: MIT Press. 2000

(Bur99) Bur Goode, "*Voice Over Internet Protocol (VoIP)"Proccedings of the IEEE, Vol. 90, No.9, September, 1999*

(Can06) Canny George, "*5-point vs. 6-point Likert Scales"* [Online] Available at http://searchwarp.com/swa69773.html> [Accessed on18th April 2011]

(Car99) Carson, et.al, "*Understanding Institutional Designs Within Marketing Value Systems*" Journal of Marketing 63 (Special Issue) 115-130, 1999

(CAR04) Catarina Sismeiro and Randolph E. Bucklin, "*Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach"* [Online] Availableat<http://www.anderson.ucla.edu/faculty/randy.bucklin/papers/SismeiroandBucklin2004.pf>[Accessed 19th Oct 2005]

(Cha99) Chandrashekaran et.al, "*What are Ontologies, and Why Do We Need Them?"* IEEE-Information Systems, 14, 1999

(Cha06) H. K. Chang and W. A. Chien, "*Neural network with multi-trend simulating transfer function for forecasting typhoon wave*" Advances in Engineering Software Volume 37 , Issue 3, Pages: 184-194, Elsevier Science Ltd. Oxford, UK, 2006

(chu07) Chun-Hsiu Yeh et.al, "*Building B2B2C Web-Page Generator base on OOPM model and Intelligent Agent"* Proceedings of the 13th Asia Pacific Management Conference, Melbourne, Australia, 2007, 888-895, 2007

(Chr07) Christoforos Ntantogian et.al, "A security Protocol For *Mutual Authentication and Mobile VPN Deployment in B3G Networks"* The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07), 2007

(Chr05) Chris Ramsey, "*Managing Web Sites as Dynamic Business Applications*" NCompass Labs [Online] Available at <http://www.intranetjournal.com/articles/200006/wm_b.html>[Accessed 5th Nov 2005]

(Cla07) Clay Mayers, "*A Service-Oriented Architecture Approach To Enterprise Capture*" EMC Captiva [Online] Available at <http://www.ecmconnection.com/article.mvc/A-Service-Oriented-Architecture-Approach> [Accessed on 19th April 2009]

(Cla97) D. Clark, M. Lambert, and L. Zhang, "*NETBLT: A high throughput transport protocol, in Frontiers in Computer Communications Technology*" Proc. of the ACM-SIGCOMM '1987, Stowe, VT, 1997

(Cra09) Crane, N, Downes et.al, "*Leadership Imperatives of the Information Age*" Electronic Journal of Knowledge Management Volume 7 Issue 2, (pp219 - 226), [Online] available at <www.ejkm com> Accessed on 10th June 2010

(Cro70) S. Crocker et.al, "*Network Working Group Request for Comments*: 33", 12 February 1970, UCLA, 1970

(Cue04) Bernardo Cuenca Grau, "*A possible simplification of the semantic web Architecture*" In Proc. of the Thirteenth InternationalWorldWideWeb Conference(WWW 2004), 2004

(Cue04) Bernardo Cuenca Grau, "A *possible simplification of the semantic web architecture*" International World Wide Web Conference, Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, Semantic web foundations, Pages: 704 - 713,ACM New York, NY, USA 2004

(Dan05) Daniel Oberle et.al, "*Supporting application development in the semantic web*" ACM Transactions on Internet Technology (TOIT), pages 328 - 358 NY, USA , 2005

(Dan04) Dan Brickley R.V. Guha, and Brian McBride (editors), "*RDF vocabularydescription language 1.0: RDF Schema*, W3C recommendation" World Wide Web Consortium (W3C), February, 2004

(Dan02) Dan Davis and Manish Parashar, "*Latency Performance of SOAP Implementations*" 2nd IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2002), 22-24 May 2002, Berlin, Germany, 2002

(Dan99) Dan Libby, "*The History of RSS Feeds*" [Online] Available at <http://www.rssfeeds.org> [Accessed on 21st July 2009]

(Deb06) Debby Young, "*Interactivity features can boost e-commerce revenue*" [Online] Available at <http://www.microsoft.com/midsizebusiness/businessvalue/interactivity.msp> [Accessed 3rd April 2007]

(Dav08) David Corsar & Derek Sleeman, "*Developing Knowledge-Based Systems using the Semantic Web*" Presented at SSS-08 on Symbiotic Relationships between the Semantic Web & Knowledge Engineering [Online] Available at <http://www.csd.abdn.ac.uk/~sleeman/published-papers/publications-2008/p187-BCS-DavidC.pdf> [Accessed on May 5th 2009]

(Dav07) David L Barkley et.al, "*E-commerce As A Business Strategy: Lessons Learned From Case Studies of Rural And Small Town Businesses*" UCED working Paper 2007, University Centre for economic Development, Clemson University, California, 2007

(Dev 05) Devin Comskey, "*Virtual Iris Set To Revolutionize e-commerce*" [Online]Available at <http://www.eco ecommerceguide.com/news/trends/article.php/>[Accessed 5th Oct 2005]

(Dei09) [Online]Available at <http://www.dei.polimi.it/> [Accessed 18th Nov 2009]

(Dia10) Dianne Cyr , Milena Head , Hector Larios, "*Colour appeal in website design within and across cultures: A multi-method evaluation*", International Journal of Human-Computer Studies, v.68 n.1-2, p.1-21, Academic Press, Inc. Duluth, MN, USA, 2010

(Don00) Don Box et.al, "*Simple Object Access Protocol (SOAP)* 1.1" W3C [Online] Available at <http://www.w3.org/TR/2000/NOTE-SOAP> Accessed on 1st Jan, 2009

(Don00) Don Box et.al. "*Simple Object Access Protocol (SOAP) 1.1*" World Wide Web Consortium, Note NOTE-SOAP-20000508, May 2000

(Dra99) Dran, et.al "*Quality Web sites: an application of the Kano model to Web-site design*" Paper presented at the 5th Americas Conference on Information Systems, Milwaukee USA, 1999

(Dum06) Dumitru Roman and James Scicluna "*Ontologybased choreography of wsmo services*" Wsmo final draft v0.3, DERI, 2006[Online] Available at <http://www.wsmo. Org/TR/d14/v0.3> [Accessed on 9th April 2008]

(Edo05) Ed Ort, "*Service-Oriented Architecture and Web Services: Concepts,Technologies, and Tools*" Sun Microsystems, Inc [Online] Available at <http://java.sun.com/developer/technicalArticles/WebServices/soa2/soa2.pdf> [Accessed 7th Nov.2007]

(Elf02) Elfwing, et.al "*Performance of SOAP in Web Service Environment Compared to CORBA*" In Proceedings of the Ninth Asia-Pacific Software Engineering Conference, IEEE, 2002

(Ema04) eMarketer, "*B2C e-commerce in the US*" [Online] *Available at <http://www.asc.upenn.edu/courses/comm530/secure/b2c_us_jul04.pdf*>[Accessed on 9th April 2008]

(Eme04) Emery, Charles R.; Tolbert, Samuel H, "*Using the Kano Model of customer satisfaction to define and communicate supervisor expectations*" Journal of Organizational Culture, Communications and Conflict, 2004

(Euz02) Jérôme Euzanat, (editor), "*Research challenges and perspectives of the Semantic Web*" Report of the EU-NSF strategic workshop, Sophia-Antipolis, France, January, 2002

(Fen02) D. Fensel and C. Bussler, "*TheWeb ServicesModeling Framework (WSMF)*" Electronic Commerce Research and Applications, 1(2):113–137, 2002

(Fen02) D Fensel, "*Language Standardization for the Semantic Web: The LongWay from OIL to OWL*" Distributed Communities on the Web: 4th International Workshop, DCW 2002, Sydney, Australia, vol. 2468 / 2002, pp. 215_227.

(Fil10) Filip Ponulak and Andrzej Kasiński, "*Supervised learning in spiking neural networks with resume: Sequence learning, classification, and spike shifting*" Neural Computation, Volume 22 , Issue 2 (February 2010) Pages: 467-510, MIT Press Cambridge, MA, USA, 2010

(Fot04) Fotini Patsioura, et.al, "*Evaluation of an agricultural web site*" University of Macedonia, Greece [Online] Available at <http://biofood.sbbe.gr/dnfiles/agricultural_website_evaluation.PDF>[Accessed 5th Nov 2005]

(Fra04) Frank Manola, Eric Miller, Brian McBride,(editors) "*RDF primer. W3C recommendation*" World Wide Web Consortium (W3C), 2004

(Gir00) M. Girardot and N. Sundaresan. Millau, "*An Encoding Format for E_cient Representation and Exchange of XML Documents over the WWW*" In Proceedings of the 9th international World Wide Web conference on Computer networks, pages 747-765, 2000.

(Geo08) George A. Komatsoulis et.al, "*caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability*" Journal of Biomedical Informatics, 41(1):106–123, 2008

(Gol90) Goldfarb, C., "*In The SGML Handbook*" Rubinsky, Y., Ed.; Clarendon Press: Oxford, UK, 1990

(Gom04) Gomez-Perez, A., Lopez, M. F., and Corcho, O., "*Ontological engineering with examples from areas of knowledge management, e-commerce and semantic web*" Springer-Verlag, Berlin, 2004

(Goo87) Goodwin, N. C., "*Functionality and usability*" Communications of the ACM, 30(3), 229-233

(Gru93) Gruber, Thomas R, "*Toward principles for the design of ontologies used for knowl-edge sharing*" *Padua workshop on Formal Ontology,* 1993

(Gua97) Guarino N, "*Understanding, Building and Using Ontologies*" *IJHCS,* 46, 1997

(Guh 05) R. Guha et.al, "*Unweaving a Web of Documents*" KDD'05 ACM, 2005

(Guh 03) Guha et. al, *"Semantic search"* In: WWW '03: Proceedings of the 12th international conference on World Wide Web, ACM Press (2003) 700–709, 2003

(Ham90) Hampshire, J.B. and Waibel, A.H. "*A Novel Objective Function for Improved Phoneme Recognition Using Time-Delay Neural Networks*" IEEE Transactions on Neural Networks, Vol. 1, No. 2, pp. 216-228, 1990

(Hea02), Hearst, M., Elliott et.al, "*Finding The Flow In Web Site Search," Communications of the ACM, 45 (9), 42-50, 2002*

(Hef04) Heflin J, "*OWL Web Ontology Language, Use Cases and Requirements"* [Online] Available at <http://www.w3.org/TR/ 2004/REC-webont-req> [Accessed on 28th Jan 2005]

(Hel07) Helmut Petritsch "*Service-oriented architecture (SOA) vs. Component based architecture"* [Online] Available at <http://petritsch.co.at/download/SOA_vs_component_based.pdf[Accessed on 9th Jan 2009]

(Hhs06) USDHS, *" Health Information"* U.S Department of Health and Human Services [Online] Available at <http://www.hhs.gov> [Accessed on 4th Feb 2008]

(Hir08) Y Hiroyuki Torikai et.al, "*Artificial Spiking Neurons and Analog-to-Digital-to-Analog Conversion*" IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,Volume E91-A , Issue 6 , Pages: 1455-1462, Publisher Oxford University Press Oxford, UK, 2008

(Hor05) Horrocks I et.al, "*Semantic Web Architecture: Stack or Two Towers?"* Lecture Notes in Computer Science: Principles and Practice of Semantic Web Reasoning: Third International Workshop, vol. 3703 / 2005. ISSN0302-9743, 2005

(Hor09) G. Horgan, S. Khaddaj, *"Use of an Adaptable Quality Model in a Production Support Environment"* in 'Journal of Systems and Software', (8) Elsevier, pp. 730-738, *2009*

(Hun98) J Hunter and R Iannella, "*The application of Metadata Standards to Video Indexing*" Second European Conference on Research and Advanced Technology for Digital Libraries, Crete, Greece, September, 1998.

(Ido08) Idoia Berges et.al, *"Semantic Web Technology for Agent Communication Protocols"* 5th European Semantic Web Conference (ESWC2008) Spain, 2008

(Ing05) Ingo Wassink et.al, "*Applying a user-centered approach tointeractive visualisation design*" Human Media Interaction Group, University of Twente, 2005

(Jae03) Jae Kyu Lee and Mye M. Sohn, "*Enhanced Knowledge Management with eXtensible Rule Markup Language*" Proceedings of the 36th Hawaii International Conference on System Sciences, Korea, 2003

(Jaf07) A. Jaffri, H. Glaser, and I. Millard, "*Uri identity management for semantic web data integration and linkage*" In 3rd International Workshop On Scalable Semantic Web Knowledge Base Systems. Springer, 2007

(Jam05) James Aflaki, Lawrence Schmitt, "*Learning Telecommunications: TCP/IP Configurator*" Allied Academies International Conference Proceedings of the Academy of Information and Management Sciences, Volume 9, Number 1 Memphis, 2005

(Jam09) James A. Hendler, "*Tonight's Dessert: Semantic Web Layer Cakes, Lecture Notes In Computer Science*" Vol. 5554, Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications,Heraklion, Crete, Greece Pages: 1 - 1, Springer-Verlag Berlin, Heidelberg, 2009

(Jan03) Jane Barton et.al, "*Building Quality Assurance into Metadata Creation:an Analysis based on the Learning Objects and e-Prints Communities of Practice*"[Online]
Available at <http://www.siderean.com/dc2003/201_paper60.pdf> Accessed on 10th Feb., 2006

(Jea03) Jean Tourrilhes, "*On-Demand BlueTooth: Experience integrating BlueTooth in Connection Diversity*" Jean Tourrilhes Mobile and Media Systems Laboratory HP Laboratories Palo Alto, USA HPL-2003-178 [online]
Available at <http://www.hpl.hp.com/techreports/2003/HPL-2003-178.pdf>
Accessed on 13th Jan., 2009

(Jef08) Jeffrey Nichols, Tessa Lau "*Mobilization by Demonstration: Using Traces to Re-author Existing Web Sites*" *IUI'08*, Maspalomas, Gran Canaria, Spain, ACM, 2008

(Jem10) Jem Rayfield "*BBC World Cup 2010 dynamic semantic publishing*" [Online]
Available at
<http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html > [Accessed on 17th August 2010]

(Jin01) Jin, X., D. Srinivasan, and R.L. Cheu, "*Classification of Freeway Traffic Patterns for Incident Detection Using Constructive Probabilistic Neural Networks*" IEEE TRANSACTIONS ON NEURAL NETWORKS, 2001. 12(5): p. 1173-1187, 2001

(Joh10) John O' Donovan, "*The World Cup and a call to action around Linked Data*" [Online] Available at <
http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html > [Accessed on 17th August 2010]

(Joh02) John Mack , "*An interview with John Mack, president of the Internet Healthcare Coalition. Interview by Richard D Lang*" [Online] Available at <http://www.hon.ch/HONcode> [Accessed on 5th July 2009

(Joh97) John MacKrell, Miller, Ed, and Alan Mendel, "*PDM buyer's guide.* 6th edition.2 vols. CIMdata Corporation., 1997

(Joh06) John L. Arnott, "*Information Visualisation in the User Interface for Augmentative and Alternative Communication*" Technology & Disability Vol.18, No.3, 2006, pp.147-161, 2006

(Jon02) H. Jonathan Chao "*Next Generation Routers*" Proceedings of the IEEE, VOL. 90, NO. 9, SEPTEMBER 2002

(Jon07) Jonathan Lazar et.al, "*Understanding web credibility: a synthesis of the research literature,Foundations and Trends in Human-Computer Interaction*" Volume 1 , Issue 2, pages 139 - 202 (January 2007),Now Publishers Inc. Hanover, MA, USA, 2007

(Jor07) *Jorge Ardenghi and Javier Echaiz, "Peer-to-Peer Systems: The Present and the Future"* The Journal of Computer Science and Technology (JCS&T) JCS&T Vol. 7 No. 3, 2007

(Jos05) Josenildo C. da Silva et.al, "*Distributed data mining and agents*" Engineering Applications of Artificial Intelligence, Pages: 791-807 Pergamon Press, Inc. Tarrytown, NY, USA, 2005

(Jos04) Joseph F. McCarthy and Eric S. Meidel, "*AgentSheets®: an Interactive Simulation Environment with End-User Programmable Agents*" [Online] Available at http://stud.hsh.no/lu/inf/im2/Interaction2000.pdf >[Accessed 17th July 2006]

(Jus03) Justin Picard and Jacques Savoy, "*Enhancing Retrieval with Hyperlinks: A General Model Based on Propositional Argumentation Systems*" Journal of the American Society For Information Science And Technology, 54(4):347–355, 2003

(Kay07) Kaylene C. Williams et.al, "*Fine-Tuning Useful E-Commerce Practices*" Journal of Technology Research [Online] Available at <http://www.aabri.com/manuscripts/08094.pdf> [Accessed on 5th January 2008]

(Kel04) Keller et al, Keller, U et.al editors, "*WSMO Discovery*" WSMO Working Draft D5.1v0.1, 2004 [Online]. Available at <http://www.wsmo.org/2004/d5/d5.1/v0.1/> [Accessed 7th July 2007]

(Kan84) N. Kano et.al, "*Attractive and normal quality, Quality*" 14 (2) (1984

(kan84) Kano, N., Seraku, N., Takahashi, F., & Tsuji, S, "*Attractive quality. Journal of Japanese Society for Quality Control, 14*, 39-48, 1984

(Kat10) Katharina Siorpaes and Elena Simperl, "*Human Intelligence in the Process of Semantic Content Creation*" Journal, World Wide Web, Volume 13 , Issue 1-2 (March 2010), Pages: 33 - 59, Kluwer Academic Publishers Hingham, MA, USA, 2010

(Ken 03) Kent R, "*The IFF Foundation for Ontological Knowledge Organisation. Knowledge Organisation and Classification in International Information Retrieval*" Cataloging and Classification Quarterly, 37, 2003

(Kha05) S. Khaddaj, G. Horgan, *"A Proposed Adaptable Quality Model for Software Quality Assurance"* in 'Journal of Computer Science', 1(4) Science Publications, pp. 481-486, *2005.*

(Kif05) Kifer M et.al, "*A Realistic Architecture for the Semantic Web*" Lecture Notes in Computer Science, vol. 3791: pp. 17 -29, 2005

(Kim05) Kim M. and Stoel, L., "*Salespersons roles: Are online retailers meeting customer expectations?*" Internationa Journal of Retail & Distribution Management, Vol.33 No.4. pp. 284-297, 2005

(Kle61) Kleinrock, L, "*Information Flow in Large Communication Nets*" RLE Quarterly Progress Report, July 1961.

(Kle62) Kleinrock, L , "*Message Delay in Communication Nets with Storage. Message Delay in Communication Nets with Storage*" P h.D. diss., MIT, December, PP 1-26, 1962

(Kle64) Kleinrock, L "*Waiting Times for Certain Queue Disciplines*. Extract from Chapter 5 of *Communication Nets: Stochastic Message Flow and Delay*" McGraw-Hill, NewYork, 1964

(Kle69) Kleinrock, L "*UCLA to be First Station in Nationwide Computer Network*" Press release dated July 3, 1969.

(Kor90) Korsten, M.J., Houkes, Z "*The estimation of geometry and motion of a surface from image sequences by means of linearisation of a parametric model*" Computer Vision, Graphics and Image Processing, vol.50, pp.1-28, 1990

(Las01) Lassila, O. and McGuinness. D "*The Role of Frame-Based Representation on the Semantic Web" Linköping Electronic Articles in Computer and Information Science,* 6, 2001

(Lau04) Laurentiu Vasiliu et.al, "*WSMO in DIP"* Technical report, June 2004. [Online] Available at <http://www.wsmo.org/2004/d19/d19.1/v0.1/20040621/> [Accessed on 19th Dec 2007]

(Law67) Lawrence G. Roberts, "*Multiple Computer Networks and Intercomputer Communication*" *Proceedings* of the First ACM Symposium on Operating System Principles,1967

(Law98) Lawrence, S. and Giles C, "*Context and Page Analysis for Improved Web Search*," IEEE Internet Computing, vol. 2, no. 4, 1998

(Lic68) J. C. R. Licklider and Robert W. Taylor, "*The Computer as a Communication Device*" Science and Technology, April, 1968

(LiD06) Li Ding and Tim Finin, "*Characterizing the Semantic Web on the Web"* The proceedings of the 5th International Semantic Web Conference, 2006, Athens GA USA, 2006

(Lil05) Lili Wang "*Evaluating Web-based e-government services with a citizen-centric approach"* Proceedings of the 38th Hawaii International Conference on System Sciences - 2005

(Luc00) Luca Chittaro and Angelo Montanari "*Temporal representation and reasoning in artificial intelligence: Issues and approaches,Annals of Mathematics and Artificial Intelligence*" Journal, Volume 28 , Issue 1-4 (2000) Pages: 47 - 106,Kluwer Academic Publishers Hingham, MA, USA,2000

(Luy05) X. Lu and W. Zhao (Eds.) "*Multiple Schema Based XML Indexing*" ICCNMC 2005, LNCS 3619, pp. 891 – 900, Springer-Verlag Berlin Heidelberg, 2005

(Lyn04) Lynch and Horton, "*Web style guide"* [Online] Available at <http://www.webstyleguide.com/site/basic_structures.html> [Accessed 3rd April 2007]

(Lyn02) Lynch, P.J., & Horton, S. "*Web Style Guide: Basic Design Principles for Creating Web Sites"* 2nd ed. 2002

(Mac05) Maciej Zaremba and Eyal Oren, "*WSMX Execution Semantics,* Technical report" D13.2v0.2, Feb 2005, [Online] Available at < http://www.wsmo.org/TR/d13/d13.2/v0.2/> [Accessed on Feb 2007]

(Mar06) Martin Hepp et.al,*"Towards The Semantic Web In E-Tourism: Can Anntation Do the Trick?"* Proceedings of the 14th European conference on Information System, Sweden June 2006, (ECIS2006)

(Mar05) Martin Dodge et.al "*Virtual Cities on the World-Wide Web Towards a Virtual City Information System"* [Online] Available at <http://www.casa.ucl.ac.uk/martin/virtual_cities.html>[Accessed 5th Nov 2005]

(Mar04) marumushi.com, *"News Map"* [Online] Available at <http://www.marumushi.com/apps/newsmap 2004>[Accessed 17th January 2006]

(Mar98) Martin Dodge et.al, "*Towards the Virtual City: VR & Internet GIS for Urban Planning*" Virtual Reality and Geographical Information Systems, Birkbeck College, 22nd May 1998

(Mat10) Matías Recabarren , Miguel Nussbaum, "*Exploring the feasibility of web form adaptation to users' cultural dimension scores, User Modeling and User-*

Adapted Interaction" v.20 n.1, p.87-108, Kluwer Academic Publishers Hingham, MA, USA, 2010

(Mat09) Matthias Mayer, "*Web History Tools and Revisitation Support: A Survey of Existing Approaches and Directions*", Foundations and Trends in Human-Computer Interaction" archive,Volume 2 , Issue 3,Pages: 173-278,Inc. Hanover, MA, USA, 2009

(Mau04) Mauro Andreolini et.al, " *Fine grain performance evaluation of e-commerce sites*" Volume 32 , Issue 3 ,_ACM SIGMETRICS Performance Evaluation Review, ACM, NY, 2004

(Med01) MedNet [Online] Available at <http://www.mednet.ucla.edu> [Accessed on 7th May 2005]

(Meh10) Mehdi Achour et.al ed (Philip Olson) "*PHP Manuel*" [Online] Available at <http://php.net/manual/en/index.php> [Accessed on 18th August 2010]

(Mel02) Melody Y. Ivory and Marti A. Hearst, "*Improving Web Site Design*" IEEE internet computing, 2002 IEEE, 2002

(McG04) McGuinness D.L. and van Harmelen F, "*OWL Web Ontology Language Overview*" [Online] Available at < http:// www.w3.org/TR/2004/REC-owl-features> [Accessed on 28th Jan 2005]

(McI01) McIlraith et.al, "*Semantic Web Services*" IEEE Intelligent Systems, 16 (2001), 46-53.

(Mic05) Michal Zaremba et.al "*WSMX Deliverable D13.4 v0.2 WSMX Architecture*" WSMX Final Draft 2005, [Online] Available at http://www.wsmo.org/TR/d13/d13.4/v0.2/20050613/20050613_d13_4.pdf> [Accessed on 10th April 2007]

(Mic07) Michael Rappa,"*Business Models On The Web*" [Online] Availabale at <http://digitalenterprise.org/models/models.html>[Accessed 3rd April 2007]

(Mic04) Michal Zaremba et.al, " *WSMX – Infrastructure For Execution of Semantic Web Services*" [Online] Available at <http://iswc2004.semanticweb.org/demos/38/paper.pdf> [Accessed on 1st Nov 2005]

(Mic05) Mick Kerrigan, "*The Web Services Modelling Toolkit (WSMT)*" Technical report [Online] Available at <http://www.wsmo.org/TR/d9/d9.1/v0.2/> D9.1v0.2 [Accessed on 21st Nov 2007]

(Mic00) Micklethwait, J., & Wooldridge, A. "*A future perfect: The challenge and hidden promise of globalization*" New York: Crown, 2000

(Mir98) Miroslav Kubat, Robert C. Holte et.al "*Machine Learning for the Detection of Oil Spills in Satellite Radar Images*" Journal, Machine Learning,Volume 30, Numbers 2-3 / February, 1998,Pages 195-215, Springer Netherlands, 1998

(Mit97) Mitch Waldrop, "*DARPA and the Internet Revolution*, 50 Years of Bridging the Gap" DARPA, 1997

(Miy93) Miyoshi T. and Azuma M., "*An Empirical Study of Evaluating Software Development Environment Quality*" IEEE Transactions on Software Engineering. Vol 19, No 5, pp 425-435., 1993

(Mor90) Morgan, N. and Bourlard, H, "*Generalization and parameter estimation in feedforward nets: some experiments*" In Neural Information Processing Systems II, Ed. D.S. Touretsky, Morgan Kaufmann, NIPS II, pp. 630-637, 1990

(Moh01) Mohammed E. Haque, "Interactive Animation And Visualisation In A Virtual Soil Mechanics Laboratory" 31th ASEE/IEEE Frontiers in Education Conference, T1C-5,Reno, NV, 2001

(Nan05) Nancy M Levenburg, Simha R Magal "*A-Framework-for-Deciding-Website-Shape-and-Content"* [online] Available at <http://www.ezinearticles.com/>[Accessed 12 Oct 2005]

(Nat07) Natalya et al. "*Ontology Development 101: A Guide to Creating Your First Ontology"* [online] Available at <http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html>[Accessed 7th July 2007]

(Nie00) Nielsen J, "*Designing Web Usability*" New Riders Publishing, 418 pages, 2000

(Nie03) Nielsen J, "*Top Ten Web Design Mistakes of 2003"* [Online].Available at <http://www.useit.com/alertbox/20031222.html>[Accessed 29th May 2004

(Nik08) Nikolaos Simou et.al, "*Semantic Adaptation of Neural Network Classifiers in Image Segmentation*" Department of Electrical and Computer Engineering, National Technical University of Athens, V. K urkov´a et al. (Eds.): ICANN 2008, Part I, LNCS 5163, pp. 907–916, 2008. Springer-Verlag Berlin Heidelberg, 2008.

(Nik04) Nik Looker and Jie Xu, *"Assessing the Dependability of SOAP RPC-Based Web Services by Fault Injection"* Proceedings of the Ninth IEEE International Workshop on Object-Oriented Real-Time Dependable Systems, 2004

(Oec08) OECD "*Internet 2018: An Essential Platform for the Global Economy and Society"* Business Vision Paper to the OECD Seoul Ministerial Meeting On the Future of the Internet Economy, 2008 [Online] Available at <http://www.biac.org/statements/iccp/Business_Seoul_Ministerial_Vision_Paper.pdf> Accessed on 15th June 2010

(Ore05) Oreste Signore " Towards A Quality Model For Websites, CMG Poland Annual Conference – Warsaw May, 2005

(Oxf02) Oxford Brookes University, "*XML Prime*" [Online]Available at<http://www.w3c.rl.ac.uk/primers/xml/xmlprimer.pdf> [Accesed on 11th Feb 2007]

(Pat04) Patel-Schneider et.al "*OWL Web Ontology Language Semantics and Abstract Syntax*" *W3C* [Online] Available at <http://www.w3.org/TR/2004/ REC-owl-semantics> [Accessed 17th May 2007]

(Pau05) Paul Rand, "*Making the Web Work*" Webstyle guide 3rd edition Bob Baxley New Riders Publishing, 2005

(Pao04) Paolucci et.al "*Expressing WSMO Mediators in OWL-S*" Carnegie Mellon UniversityPittsburgh, Pennsylvania, USA, 2004

(Paw01) Pawan R. Vora "*XOR, Designing Usable Web Sites for e-commerce*" [Online] Available at <http://www.upassoc.org/conferences_and_events/upa_conference/2001/reg/program/tutorials/t3.html>[Accessed 5th March 2007]

(Pau06) Paul Shabajee "*Informed Consent on the Semantic Web - Issues for Interaction and Interface Designers*" [online]Available at <http://swui.semanticweb.org/swui06/papers/Shabajee/Shabajee.pdf >[Accessed 5th Oct 2006]

(Pau05) Paul Chin "*Human-Centered Intranet Design*" [Online] Available at<http://www.intranetjournal.com/articles/200511/ij_11_28_05a.html>[Accessed 5th Oct 2005]

(Pet 02) Peter et.al "*Layering the Semantic Web: Problems and Directions*" Proceedings of the First International Semantic Web Conference (ISWC2002) Italy, 2002

(Pet 03) Peter Lyman and Hal R. Varian, "*How much information*" Technical report, *UC Berkeley, 2003.*[Online] Available at <http://www.sims.berkeley.edu/research/ projects/how-much-info-2003/. >[Accessed 5th Oct 2006]

(Pet98) Peter T. Kirstein "*Early Experiences with the ARPANET and INTERNET in the UK*" History of the UK ARPANET/Internet links,1998

(Pet89) Peter J. Denning, "*The ARPANET after Twenty Years*" RIACS Technical Report TR-89.38, 1989

(Pet02) Petra Wilson [Online] Available at <http://www.bmj.com/cgi/content/full>[Accessed 11th March 2006]

(Phy09) Online] Avialable at <http://www.physics.csbsju.edu/stats/descriptive2.html> [Accessed on18th April 2011]

(Pin 09) Pingdom [online] Available at <http://royal.pingdom.com/2009/01/22/internet-2008-in-numbers> [Accessed on 11th Nov.2009]

(Que05) Queens- land.govt [online] Available at <http://queens-land.gov>[Accessed on 11th Nov.2009]

(Qui 08) B. Quilitz, U. Leser, "*Querying distributed RDF data sources with SPARQL*" Proceedings of the Fifth European Semantic Web Conference, Lecture Notes in Computer Science, Vol 5021, pp. 524-538, 2008

(Qvo02) L.Qvortrup ed., "*Virtual Space: spatiality in virtual inhabited 3D worlds*" Springer, Great Britain: 2002

(Rac04) Rachatrin Tongrungrojana and David Lowe "*WIED: A Web Modelling Language for Modelling Architectural-Level Information Flows*" Journal of Digital Information, Vol 5, No 2, 2004 [Online] Available at < http://journals.tdl.org/jodi/article/view/139/137> [Accessed on 24th Nov2008]

(Rac07) Rachatrin Tongrungrojana and David Lowe, "*WebML+ : a Web modeling language for forming a bridge between business modeling and information modelling*" [Online] Available at < http://services.eng.uts.edu.au/userpages/dbl/public_html/archive/2003-Ton03c.pdf> [Accessed on 15th April 2008]

(Ric05) Riccardo Albertoniet.al, "*Information Search: the Challenge of Integrating Information Visualization and Semantic Web*" Proceedings of the 16th International Workshop on Database and Expert Systems Applications (DEXA'05)1529-4188/05 $20.00 © 2005 IEEE

(Rim09) Rimantas Gatautis, "*Towards eBusiness Support Policy Deelopment: Lithuania Perspectives*" Kaunas University of Technology, Lithuania, ECONOMICS & MANAGEMENT: 14 , 2009

(RME07) Ramesh Loganathan, "*Contrary Opinion: "SOA Has Its Limitations"* [Online]Available at <http://webservices.sys-con.com/read/44978.htm> 2007 [ACCESSED 17TH APRIL 2007]

(Ram06) Ramon F. Brena, Eduardo H. Ramirez, "*A Soft Semantic Web*" , IEEE, 2006

(Reh00) Rehman, A "*Holiday 2000 E-Commerce: Avoiding $14 Billion in "Silent Losses"* (Research Report). New York, NY: Creative Good, 2000

(Riz00) Rizzuti, K. and Dickinson, J, "*Satisfying the Experienced On-Line Shopper: Global E-Shopping Survey*" (Research Report). London, 2000.

(Rob06) Roberto et.al, "*Web Services Description Language (WSDL) Version 2.0Part1:CoreLanguage*" World WideWeb Consortium, W3C Recommendation CR-wsdl20-20060327, March 2006

(Rob07) Rob Hicks, "*Health information on the internet*" [Online] 2007, Available at <http://www.bbc.co.uk/ > [Accessed on 9th Nov.2008]

(Rom08) Roman Dunaytsev, "*Networking Concepts: Layers, Models, and Protocols*" Department of Communications Engineering

Tampere University of Technology TLT-2106, Lecture 2 [Online] 2008, Available at<http://www.cs.tut.fi/~dunaytse/slides/lectureOSI.pdf> [Accessed 17th Oct 2010]

(Rom05) Roman et.al, "*Web Service Modeling Ontology, Applied Ontologies*" pp77 – 106, 2005.

(Rog09) Roger E. Bohn and James E. Short, "*How Much Information? 2009, Report on American Consumers*" Global Information Industry Center University of California, San Diego, December, 2009 [Online] Available at <http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf> Accessed on 10th June 2010

(Rum86) Rumelhart, D.E., McClelland, J.L, "*Parallel distributed processing: explorations of the microstructure of cognition*" MIT Press, 1986

(Rum86) Rumelhart et.al, "*Learning internal representations by error propagation, in*: *Parallel distributed processing: explorations of the microstructure of cognition*" vol.1: Foundations (D.E. Rumelhart and J.L.McClelland eds.), MIT Press, 1986

(Sam10) SMax L. Wilson, "*From Keyword Search to Exploration: Designing Future Search Interfaces for the Web*, Journal, Foundations and Trends in Web Science" Volume 2 , Issue 1 (January 2010) Pages: 1-97, Now Publishers Inc. Hanover, MA, USA, 2010

(Sea09) Sean B. Palmer, "*The Semantic Web: An Introduction*"[online] 2009, Available at <http://infomesh.net/2001/swintro>[Accessed 5th Oct 2006]

(Sea04) Sean Bechhofer et.al "*OWL Web Ontology Language Reference*" World Wide Web Consortium, Recommendation REC-owl-ref-20040210, February, 2004

(Sch01) Schaffer, E, "*Quickness and Usability Keys to Successful Web sites, Bank Systems & Technology*" 38 (2), 50, 2001

(Ser99) R. M. Lawrence et.al, "*The pagerank citation ranking: Bringing order to the web" Stanford Digital Library working paper SIDL-WP—0120, 1999*

(Sha02) Sharon Machlis, "*Virtual Interview: Rasmus Lerdorf*" [Online] 2002, available at<http://www.devarticles.com/c/a/Interviews/Virtual-Interview-Rasmus-Lerdorf> [Accesed on 11th Feb 2007]

(Shi10) Shieh, J.I., H.H. Wu and K.K. Huang, "*A dual importance diagram approach to evaluate service quality" Information technology journal 9, 2010*

(Spre08) Spremić, M., Strugar, I, "*Research study on managing web site quality in Croatia, Advances on Applied Computer Science*" Proceedings of the 8th WSEAS International Conference on Applied Computer Science (ACS '08), WSEAS Press, Series of Reference Books and Textbooks, pp. 165-171, November, 2008.

(Sri04) Sriram Ramabhadran et.al, *"Prefix Hash Tree An Indexing Data Structure over Distributed Hash Tables"* Computer Science Department, University of California, San Diego, November 2004.

(Sta04) Staab, S. and Studer, R.(eds),*"Handbook on ontologies"* Springer-Verlag, Berlin, 2004

(Ste10) Steve Harris, Garlik, Andy Seaborne (eds), *"SPARQL 1.1 Query Language"* W3C Working Draft, June, 2010

(Ste06) Steve Harris and Jim Davies. *"CancerGrid: Model-driven and metadata-driven clinical trials informatics"* In Microsoft eScience Workshop, John Hopkins University, 2006

(Ste09) Steve Wallace , Hsiao-Cheng Yu, *"The Effect of Culture on Usability: Comparing the Perceptions and Performance of Taiwanese and North American MP3 Player Users"* JUS journal of usability studies Vol. 4, Issue 3, pp. 136-146, May 2009

(Ste08) Stephen Bryant, *"The Story of the Internet"* Teacher's notes, Pearson Education Limited, 2008

(Ste00) S. Mazzocchi, *"Toward the semantic web"* [Online] 2000, Available at <http://www.betaversion.org/~stefano/papers/semanticweb.pdf> [Accessed on 23rd Oct 2005]

(Ste07) Stefano Ceri et.al, *"Web Modeling Language (WebML): a modeling language for designing Web sites"* Dipartimento di Elettronica e Informazione, Politecnico di Milano Piazza L. da Vinci, 32 - 20133 Milano, Italy [Online]. Available at <http://www.webml.org/webml/upload/ent5/1/www9.PDF> [Accessed 7th July 2007]

(Str89) Straub, D. *"Validating instruments in MIS research"* MIS Quarterly, 13(2), 147-169, 1989

(Swi01) SW *"The Semantic Web Identification Problem"* [Online] 2001-03, Available at <http://www.w3c.org/2001/03/ identification-problem. >[Accesed on 9th May 2007]

(Syb89) Sybex, *"CCNA .Cisco.Certified.Network.Associate.Study.Guide.Exam"* .640.802.6th.Edition.Aug., 2007

(Tar05) Tarafdar, Monideepa, *"Analysis of Critical Website Characteristics: A Cross-Category Study of Successful Websites"* [Online] 2005, Available at <http://www.allbusiness.com/technology/computer-software-customer-relation/1056608-1.html> [Accessed on 7th Feb 2006]

(The03) Theophilus et al, *"B2B e-commerce in the United States, Europe, and Japan: a comparative study, Journal of the Academy of Business and Economics"*

[Online] 2003, Available at <http://www.findarticles.com> [Accessed 1 Jan 2005]

(Tho06) Thomas, E, "*Service-Oriented Architecture (SOA): Concepts, Technology, and Design*" The Prentice Hall, 2006

(Tho06) Thomas Haselwanter et.al, "*WSMX: A Semantic Service Oriented Middleware for B2B Integration*" Lecture Notes in Computer Science, Volume 4294/2006 Springer Berlin / Heidelberg, 2006

(Tho08) Thomas G. Dietterich, Pedro Domingos et.al, "*Structured machine learning: the next ten years*" journal, Machine Learning, Volume 73 , Issue 1,Kluwer Academic Publishers Hingham, MA, USA, 2008

(Tim03) Timothy Falconer, "*The meaning of semantic* [Online] 2003, Available at <http://bigfractaltangle.com/archive/2003/12/20.jsp>[Accessed 28th Sept 2006]

(Tiz07) Tiziana et.al,"*The SWS Mediator With WEBML/WEBRATIO And JABC/JETI: A Comaprison*" Institute for Informatics, University of Potsdam,Germany, 2007

(Tod05) Todd et.al, "*The Semantics of Ontology Alignment*" [Online] 2005, Available at < http://www.atl.external.lmco.com/projects/ontology/papers/SOA.pdf> [Accessed on 7th Nov 2006]

(Tom07) Tomas Vitvar et.al, "*Semantic SOA to Promote Integration of HeterogeneousB2B Services*" [Online] 2007, Available at <http://www.vitvar.com/!publications/EEE2007-VitvarMZHK.pdf>[Accessed 7th July 2007]

(Tor06) Torre Zuk et.al, "*Heuristics for Information Visualization Evaluation*" BELIV Italy, ACM, 2006

(Tor09) Torunn Vølstad et.al, "*Layered RDF: An object-oriented approach to web information representation*" Volume 7 , Issue 3 (August 2009) Pages: 281-301, IOS Press Amsterdam, The Netherlands, The Netherlands, 2009

(Unc07) Uncle Bob,"*A Rational Plan to Start using SOA*" [Online] 2007, Available at <http://blog.objectmentor.com/articles/2007/10/21/a-rational-plan-to-start-using-soa>[Accessed on 8th April 2009]

(Usc98) USCHOLD, M, "*Knowledge level modelling: concepts and terminology*" Knowledge Engieering Review, 13, 1998

(Won10) Won Kim et.al, "*On social Web sites*, Information Systems archive" Volume 35 , Issue 2, Pages: 215-236, Elsevier Science Ltd. Oxford, UK, 2010

(Van45) Vannevar Bush, "*As We May Think*" The Atlantic Monthly,176, July, 1945

(Vin78) Vinton G. Cerf and Jonathan B. Postel, *"Specification of Internetwork Transmission Control Program TCP" version3*, University of southern California, 1978

(Vin74) Vinton G. Cerf and Robert E. Kahn, *A Protocol for Packet Network Intercommunication*, IEEE Transactions on Communications, COM-22, no. 5 (May 1974): 637–648, 1974

(Vit07) Vitaly Friedman, "*Data Visualization: Modern Approaches*" Smashingmagazine [Online] 2007, Available at <http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches> [Accessedon on 1st July 2008]

(Web05) [Online] 2005, Available at <http://www.webml.org/webml/page9.do;jsessionid> [Accessed on 10th Feb 2007]

(Wil08) Wilson, M.L, "*A Transfer Report on the Development of a Framework to Evaluate Search Interfaces for their Support of Different User Types and Search Tactics*" School of Electronics and Computer Science, University of Southampton, 2008

(Wol03) Wolfinbarger M. and M.C. Gilly, "*eTailQ: Dimensionalizing, Measuring and Predicting etail Quality*" Journal of Retailing, 79 (3), 182-198, 2003

(Woo08) Wooju Kim et.al, "*Intelligent Product Information Search Framework Based on the Semantic Web*" Department of Information Industrial Engineering, Yonsei University, Korea, 2008

(Wsm07) wsmx.org, [Online] 2007, Available at <http://www.wsmx.org/papers/publications/kerrigan-wiw05.pdf>[Accessed 10th May 2007]

(W3c05) W3C, *"RDF Data Access Use Cases and Requirements"* [Online] 2005, Available at < http://www.w3.org/TR/2005/>WD-rdf-dawg uc20050325/> [Accesed on 11th Feb 2007]

(W3c06) W3C, *"SPARQL Query Language for RDF"* [Online] 2006, Available at <http://www.w3.org/TR/rdf-sparql-query/> [Accessed on 10th Feb 2007]

(W3c04) W3C, *"RDF Primer"* [Online] 2004, Available at <http://www.w3. org/TR/rdf-primer/>[Accessed on 10th June 2007]

(W3c04) W3C, [Online] 2004, Available at <http://www.w3.org/TR/owl-features/>

(W3c04) W3C, [Online] 2004, Available at <http://www.w3.org/TR/owl-ref/> [Accessed on 28th Jan 2005]

(W3c01) W3C, [Online] 2001, Available at <http://www.w3.org/2001/sw/interest/> [Accessed on 17th Feb 2007]

(Xia07) Xiaoguang Qi and Brian D. Davison, "*Web Page Classification: Features and Algorithms*" Technical Report LU-CSE-07-010, Department of Computer Science and Engineering, Lehigh University, Bethlehem,PA, 18015, 2007

(Xue04) Xueming Luo, Mojtaba Seyedian, "*Contextual Marketing and Customer-Orientation Strategy for E-Commerce: An Empirical Analysis*" International Journal of Electronic Commerce, Volume 8 , Issue 2 2004, M. E. Sharpe, Inc. Armonk, NY, USA, 2004

(Yij04) Yijian Wang and David Kaeli, "*Execution-driven Simulation of Network Storage Systems*" Proceedings of The IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, MASCOTS-04, 2004

(Yon02) Yonina Cooper and Hal Berghel, "*World Wide Web 10 Years Later*" (Part1) [Online] 2002, Available at <http://berghel.net/publications/web10/web10-1.php>[Accessed on 11th May 2007]

(Yua04) Yuanbo Guo et.al, "*An Evaluation of Knowledge Base Systems for LargeOWL Dataset*" [Online] 2004, Available at <http://swat.cse.lehigh.edu/pubs/guo04c.pdf> [Accessed on 10th Feb 2007]

(Yve03) Yves Lafon, "*HTTP - Hypertext Transfer Protocol*" W3C 1996-2003 [Online] 2003, Available at <http://www.w3.org/Protocols/Overview.html>[Accesed on 9th May 2007]

(Zha01) Zhang, P. and G.M. von Dran, "*User Expectations and Rankings of Quality Factors in Different Web Site Domains*" International Journal of Electronic Commerce 6 (2), 9-34, 2001

(Zon99) Zona Research, "*Shop Until You Drop? A Glimpse into Internet Shopping Success*" (Zona Assessment Paper). Redwood City, CA, 1999

(Ziq07) Ziqiang Xu et.al, "*Reputation-Enhanced QoS-based Web Services Discovery*" IEEE International Conference on Web Services ICWS, 2007

(Zol05) Zoltan Gyongyi and Hector Garcia-Molina, "*Web Spam Taxonomy*" Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web AIRWeb 2005 14th International World Wide Web Conference, Chiba, Japan, May, 2005

(Zsz09) Z. Szalavári, "*Overview of City Tales Second City Database and Client Description*" 05/2009, IP City, Vienna