TOURISM INFORMATION SYSTEMS INTEGRATION AND UTILIZATION WITHIN THE SEMANTIC WEB

by

Brooke Abrahams

Submitted to Victoria University in Fulfilment of the Degree of:

Doctor of Philosophy

in the School of Information Systems

Faculty of Business and Law

October 2006

ACKNOWLEDGEMENTS

This research was supported by an Australian APA scholarship, ICT top-up scholarship, and CRC top-up scholarship. I gratefully acknowledge this support and thank Lesley Birch for her role in administering these scholarships. In addition I wish to express my thanks to:

My supervisor Professor G Michael McGrath for his guidance and encouragement throughout the course of the PhD and constructive criticism of earlier drafts of the thesis.

Dr Wei Dai for his technical assistance relating to the development of a semantic portal prototype.

Dr Stephen Burgess for his general assistance and feedback on postgraduate presentations.

Professor John Zeleznikow for his role as acting supervisor while my principal supervisor was on leave for one semester.

My parents Don and Josie, and my sister Niki for their encouragement and support during the course of my studies.

DECLARATION

I Brooke Abrahams declare that the PhD thesis entitled 'Tourism Information Systems Integration and Utilization within the Semantic Web' is no more than 100,000 words in length, exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Signature Date

SUMMARY

The objective of this research was to generate grounded theory about the extent to which the Semantic Web and related technologies can assist with the creation, capture, integration, and utilization of accurate, consistent, timely, and up-to-date Web based tourism information.

Tourism is vital to the economies of most countries worldwide (developed and less-developed). Advanced Destination Marketing Systems (DMS) are essential if a country's tourism infrastructure, facilities and attractions are to receive maximum exposure. A necessary prerequisite here is that relevant data must be captured, 'cleansed', organized, integrated and made available to key industry parties (e.g. travel agents and inbound tour operators). While more and more tourists are using the Internet for travel planning, the usability of the Internet as a travel information source remains a problem, with travellers often having trouble finding the information they seek as the amount of online travel related information increases. The problem is largely caused by the current Web's lack of structure, which makes the integration of heterogeneous data a difficult time consuming task.

Traditional approaches to overcoming heterogeneity have to a large extent been unsuccessful. In the past organizations attempted to rectify the problem by investing heavily in top-down strategic information systems planning projects (SISP), with the ultimate aim of establishing a new generation of systems built around a single common set of enterprise databases. An example of this approach to integration is that undertaken by the Bell companies (Nolan, Puryear & Elron 1989), whose massive investment in computer systems turned out to be more of a liability than an asset. The Semantic Web offers a new approach to integration. Broadly speaking, the Semantic Web (Berners-Lee, Hendler & Lassila 2001) refers to a range of standards, languages, development frameworks and tool development initiatives aimed at annotating Web pages with well-defined metadata so that intelligent agents can reason more effectively about services offered at particular sites. The technology is being developed by a number of scientists and industry organizations in a collaborative effort led by the Worldwide Web Consortium (W3C) with the goal of providing machine readable Web intelligence that would come from hyperlinked vocabularies, enabling Web authors to explicitly define

their words and concepts. It is based on new markup languages such as such as Resource Description Framework (RDF) (Manola & Miller 2004), Ontology Web Language (OWL) (McGuinness & Harmelen 2004), and ontologies which provide a shared and formal description of key concepts in a given domain.

The ontology driven approach to integration advocated here might be considered 'bottom-up', since individual enterprises (and parts of the one enterprise) can apply the technology (largely) independently – thereby mirroring the processes by which the Web itself evolved. The idea is that organizations could be provided with a common model (the Semantic Web ontology), and associated (easy-to-use) software could then be employed to guide them in the development of their Websites. As such, because Website production is driven by the common ontology, consistency and convenient integration is almost an automatic by-product (for all companies that take advantage of the technology and approach). In many cases, organizations would not have to change their present data structures or naming conventions, which could potentially overcome many of the change management issues that have led to the failure of previous integration initiatives.

Many researchers (e.g. (El Sawy 2001)) have stressed the necessity to take a holistic view of technology, people, structure and processes in IT projects and, more specifically, Sharma et al. (2000, p. 151) have noted that as significant as DMS technological problems are, they may well pale into insignificance when compared with the managerial issues that need to be resolved. With this in mind, a systems development research approach supported by a survey of tourism operators and secondary interviews was used to generate grounded theory. The systems development and evaluation were designed to uncover technical benefits of using the Semantic Web for the integration and utilization of online tourism information. The survey of tourism operators and secondary data interviews were aimed at providing an understanding of attitudes towards adoption of a radical new online technology among industry stakeholders.

A distinguishing feature of this research was its applied and pragmatic focus: in particular, one aim was to determine just what of practical use can be accomplished *today,* with *current* (albeit, extended) technology, in a real industry setting.

TABLE OF CONTENTS

1 INTRODUCTION

Chapter 1 commences with an introduction to the research topic. Some background information about the topic is provided, which includes a brief history of the Internet and search engines. The research problem is then introduced along with the aims, methodology and research approach. The chapter concludes with an outline of the thesis.

1.1 Research Topic

The World Wide Web (WWW) has evolved to become a major source of information and services. It is decentralized, gigantic with unchecked growth, and constantly changing in its structure. The full potential of the current Web, however, remains untapped because information is rendered to be processed by machines, but understandable to humans only. Hypertext Markup Language (HTML) offers the freedom to present anything about any subject and make it available over the Web. This freedom has created a major problem of heterogeneity making the integration and utilization of information a difficult task. Traditional solutions for information interoperability are essentially top-down, and involve the development of interfaces between pairs of communication systems built around a single common set of enterprise wide databases. These approaches are too expensive and inflexible for many sectors including e-tourism, which is the leading application field in business-to-consumer (b2c) e-commerce (Werthner 2003).

In recent years, the notion of the Semantic Web (Berners-Lee, et al. 2001) has been introduced to define a machine-interpretable Web targeted for automation, integration and reuse of data across different applications. Data instances on the Semantic Web are enriched with metadata, defined as concepts and properties from ontologies, which are formal, explicit specifications of shared conceptualizations of a given domain of discourse. This enables machines to intelligently process and reason more effectively about information on the Web, thus providing an exciting new opportunity for improved information integration. It is the potential benefits and limitations of this opportunity for tourism Information and Communication Technology (ICT) systems that the thesis investigates.

1.2 Research Background

This section provides a brief history of the Internet and search engines.

1.2.1 The Internet

The origins of the Internet, which are summarized by Howe (2005), trace back to a group of people in 1960's who saw great potential value in allowing computers to share information on research and development in scientific and military fields. In 1962 J.C.R. Licklider of MIT first proposed a global network of computers. Later that year he moved to the Defense Advanced Research Projects Agency (DARPA) to lead development of this network. Leonard Kleinrock of MIT and later UCLA developed the theory of packet switching, which was to form the basis of Internet connections. Lawrence Roberts of MIT connected a Massachusetts computer with a California computer in 1965 over dial-up telephone lines. This showed the feasibility of wide area networking, but also showed that existing telephone switching technology was inadequate. Roberts moved to DARPA in 1966 and developed his plan for ARPANET, which as Alesso & Smith (2004e) explain was an initiative intended to promote sharing of super computers among scientists and military researchers in the USA. According to Howe (op. cit), these visionaries (and many more left unnamed here) are the real founders of the Internet. In the 1970's software protocols began to emerge to facilitate file transfer and email. In 1978 Bob Kahn and Vint Cerf along with other project members created TCP/IP, which is a common set of protocols for information exchange on the Internet that are still in use to the present day. Throughout the 1980's corporations increasingly began communicating with each other via the Internet as well as with customers who owned personal computers (PC's).

The transition towards the modern World Wide Web did not occur until 1991 when Tim Berners Lee introduced the concept of HTML, which provided the ability to combine words, pictures, and sounds on Internet pages and access them via a Web browser. Since the advent of the Web browser, the Internet has grown to become a global information superhighway and, in the last few years, there has been a new phase of commercialization.

Originally, commercial efforts consisted mainly of vendors providing basic networking products, and service providers offering the connectivity and basic Internet services. The Internet has now become almost a "commodity" service, and much of the latest attention has been on the use of this global information infrastructure for support of other commercial services. Leiner et al. (2003) state that this has been tremendously accelerated by the widespread and rapid adoption of browsers and the World Wide Web technology, allowing users easy access to information linked throughout the globe. The widespread growth of Internet usage is highlighted by Neilson Net Ratings[1] whose Internet usage statistics show that in the year ending December 31 2005, there were approximately one billion Internet users. This equates to 15.7% of the estimated world population of 6.5 billion. New products increasingly facilitate provisioning Web-based information and many of the latest developments in technology have been aimed at providing increasingly sophisticated information services on top of basic Internet data communications. The Internet continues to change and evolve. It is now beginning to provide new services such as real time transport in order to support, for example, audio and video streams, and services such as dynamic product packaging, as in the case of advanced Destination Marketing Systems (DMS).

1.2.2 Search Engines

Search engines are tools that provide users with a graphical user interface (GUI) to assist locating Websites containing specific categories of information. They exploit both the content of Web documents and the structure implicit in the hyperlinks connecting one document to another (Sheth et al. 2005, p. 11). Alesso & Smith (2004d) classify search engines according to the following two implementation types:

- Individual – Individual search engines compile their own searchable databases on the Web (e.g. Google[2]).
- Meta – Metasearchers do not compile databases. Instead, they search the databases of multiple sets of individual engines simultaneously (e.g. Yahoo![3]).

---

[1] http://www.nielsen-netratings.com/

[2] www.google.com

[3] http://www.yahoo.com/

Alesso & Smith (op. cit) categorize search engines according to the following functionality types:

- Lexical – searches for a word or a set of words, with Boolean operations (AND, OR, EXCEPT).
- Linguistic – allows words to be found in whatever form they take, and enables the search to be extended to synonyms.
- Semantic – the search can be carried out on the basis of the meaning of the query.
- Mathematical – semantic search operates in parallel with a statistical model adapted to it.
- Metasearch – searches the database of multiple sets of individual search engines simultaneously. Metasearchers provide a quick way of finding out which search engines are retrieving the best results for your search.
- Structured Query Languages (SQL) - a search through a sub-set of documents of a database defined by SQL (widely used by Web portals[4]).
- XML structured query – the initial structuring of a document is preserved and the request is formulated in Xpath[5].

Figure 1 illustrates a breakdown of search engine usage on the Web for the year 2005.

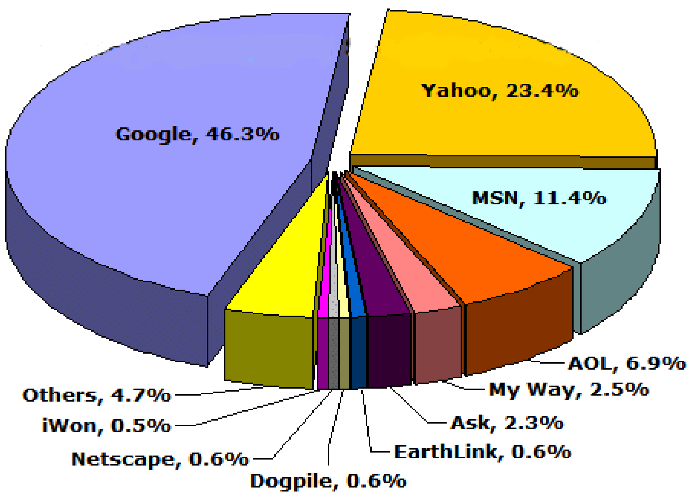


Figure 1: Search engine usage for the year 2005 (Sullivan 2005).

---

4 An SQL search engine type from a conventional portal will be used as the bases for comparison of Semantic Web search and conventional Web search methods in Chapter 4.

5 http://www.w3.org/TR/xpath

Statistics show that the most widely used search engine at present is Google. At the heart of Google's search software is a system for ranking Web pages known as PageRank. PageRage, which was developed at Stanford University by Larry Page and Sergey Brin, uses the Internet's vast link structure as an indicator of the importance of a Web page in relation to the search. The PageRank algorithm combined with sophisticated text-matching techniques measures all aspects of a page's content to determine an importance ranking which Google remembers. Although search engines such as Google are very effective at ranking relevant content, they are still limited by the fact that the ranking analysis is based on keywords rather than the underlying concepts associated with a Web page.

1.3 Research Problem

The research problem is categorized into three distinct parts and should be viewed as follows: 1) there are a number of limitations associated with the current Internet; that 2) create significant challenges for information systems integration; which 3) have negative consequences for tourism ICT applications.

1.3.1 Limitations of the Current Internet

As the World Wide Web's infrastructure, scale and impact have grown, Internet users are increasingly in need of more powerful technologies capable of collecting, interpreting and integrating the vast amount of heterogeneous information available on the Web. This heterogeneity stems from the fundamental disparity of Web domains. In the tourism industry for example, there are numerous Web portals containing vast amounts of information about accommodation, transportation, entertainment, and insurance. Most of the information on the Web is presented as natural-language text with occasional pictures and graphics. Ding et al. (2005) explain that even though this is convenient for human users to read and view, it is difficult for computers to understand. Consequently, current Web technology presents serious limitations for integrating information, and making it accessible to users in an efficient manner. These limitations are summarized in Lausen et al. (2003), who state that the main problem is that searches are imprecise, often yielding matches to many thousands of hits. Users face the task of reading the documents retrieved in order to extract the information desired – thus making information searching, accessing extracting, interpreting and processing a difficult time consuming task.

Today, search engines such as Google and Yahoo! dominate the Web's infrastructure and largely define Web users' experience. Ding et al. (op. cit) contend that conventional search engines have limited indexing capabilities, since they cannot infer meaning. For example, does an occurrence of the word "raven" refer to the bird or to Baltimore's football team? A search relying purely on the keyword "raven" is unable to definitively return answers that relate to the correct context. The ambiguity associated with current search engines is also highlighted by Alesso (2004), who states that because Web search engines use keywords for indexing concepts, they are subject to the two well-known linguistic phenomena that strongly degrade a query's precision and recall; 1) Polysemy (one word might have several meanings); and 2) Synonymy (several words or phrases might designate the same concept). These limitations have resulted in a number of significant problems for accessing reliable up-to-date information that urgently need to be solved. One of the most significant problems, which is described in detail by Stuckenschmidt & Harmelen (2005b), is Information Integration – i.e. even when it is possible to find any particular piece of information, it is very hard to combine it with other information that may already be known.

1.3.2 The Problem of Information Integration

The problem of accessing online information has in the most part been solved by the invention of large-scale computer networks such as the World Wide Web. The problem of combining, interpreting, and processing retrieved information (in other words information integration), however, remains an important research topic. The difficulties of integrating heterogeneous data are well known within the distributed database systems community. Stuckenschmidt & Harmelen (2005b) say that in general, heterogeneity can be divided into three categories[6]:

1. Syntax (e.g. data and format heterogeneity).
2. Structure (e.g. homonyms, synonyms or different attributes in database tables).
3. Semantics (e.g. intended meaning of terms in a special context or application).

[6] For a more detailed description of the types of heterogeneity that may occur please refer to section to sub-section 2.2.8

Stuckenschmidt (2005b) explains that the existence of standardized Web markup languages enables data to be represented and structured on the World Wide Web in a uniform way. According to Stuckenschmidt, this uniformity makes it easier to automatically process not only local data, but also information obtained from other sources. Syntactic homogeneity is an important enabler of information sharing. Experiences from the database area, however, have shown that the existence of syntactic standards is not enough. Even in almost completely homogeneous environments such as relational databases, the exchange of information is a problem, because heterogeneity in the way information is structured and interpreted lead to conflicts when information from different sources needs to be combined. To meet integration requirements, two broad approaches are possible:

- Top-down: the data warehousing approach – for example, where consortia of government bodies, trade organizations and larger tourism industry companies establish a shared data repository, define common metadata standards, coopt key (large) content providers and when "critical mass" is reached use this as a lever to bring smaller enterprises on-board (Sharma, Carson & DeLacy 2000). An example of this approach is the *Australian Tourism Data Warehouse (ATDW)* (Daniele, Misitilis & Ward 2000). Other prominent examples are the destination and product marketing Websites of the Australian state tourism authorities.
- Bottom-up: Websites of customers and suppliers are annotated with metadata describing site contents, consistent with a common ontology (a consensual, shared and formal description of key concepts in a given domain – in this case, tourism). Intelligent agents can then reason about services offered at particular locations through direct access to the relevant Websites. This approach utilizes Semantic Web technologies, tools and frameworks.

Traditional approaches to data integration are all essentially 'top-down', in that they are driven by senior management, or even governments or industry bodies. While these top-down approaches seem to make sense theoretically, the evidence strongly suggests that they do not work in practice (Markus & Tanis 2000). An example of one these failures described by Nolan et al. (1989), is the Bell companies massive 1980's investment in computer systems which turned out to be more of an integration liability than an asset. Reasons for such failures identified by Lederer & Sethi (1992) include technical obstacles, overoptimistic cost and schedule estimates, lack of senior management support, poor communication and change management, inappropriate IT department structures and failure to address people-related issues. The bottom-up Semantic Web approach on the other hand, remains largely untried. A significant exception here is the high-profile, EU-funded ‘Harmo-TEN’ project formally known as ‘Harmonise’ (Dell'Erbra et al. 2005).

1.3.3 Consequences of Internet Limitations for Tourism ICT Applications

The use and application of Internet based technologies in commerce, government, and education, is undergoing extraordinary growth, with the World Wide Web significantly altering the way in which traditional business is conducted (Sandy & Burgess 2003). The travel and tourism industry is no exception, where according to Werthner (2003), the industry’s acceptance of e-commerce has created a new type of tourism customer that now become their own travel agents and build travel packages themselves. Staab (2005) believes that what is most impressive about today’s information systems, is the complexity and the intricate ways that different systems interact with each other in a useful manner. Internationally, perhaps one of the major thrusts of tourism ICT systems research over the past five years has been the development and maturation of intelligent 'Travel Recommender Systems' (TRS). Broadly speaking, TRS aim to: 1) match tourism customers needs to suppliers' offerings; and 2) promote the offerings (destinations) themselves (together with all their delights, features and facilities) through wider and, perhaps, more targeted exposure. These systems make it possible to book services such as air travel or accommodation at any time from virtually anywhere in the world.

TRS are not new: e.g. travel agents, utilizing guide books, brochures, other promotional material, and (perhaps most importantly) their expert knowledge of the industry, key industry contacts and customers, have been developing and utilizing their individual TRS for decades. The difference now is that with advanced computer technology, combined with the ubiquity of the Internet and the Web, much of the functionality of these tools can be automated and their reach greatly extended - leading to much more useful systems (McGrath & Abrahams, 2006b, p. 1). For this new generation of TRS to be effective, customer and supplier data must be 1) available online; and 2) defined consistently, precisely and unambiguously so that its meaning is absolutely clear. In short, distributed, disparate and heterogeneous data sources must be integrated.

The unstructured nature of the Internet as described in section 1.3.1, and lack of global schemas means that much of the available tourism information is meaningful to humans only - and not machines. As a consequence, the success of handling transactions involving heterogeneous data on disparate systems depends on the foresight and analytical capability of the individual software developer to program a system to perform the required integrative tasks. The programmer's capabilities are restricted by the available software and data structures at their disposal, which at present makes the task of integrating tourism information difficult, costly, and time consuming (Staab 2005, p. 181). A better solution for tourism information integration may lie with a bottom-up Semantic Web approach. It is the benefits and limitations of this approach that were investigated by conducting the research.

1.4 Aims of the Study

At a theoretical level, the research attempted to provide a comprehensive understanding, from a tourism ICT systems perspective, of the benefits and limitations associated with a novel approach to tackling one of the more critical problems currently confronting information systems researchers (systems and data integration). At a physical level, the research investigated state of-the-art tools, development techniques, applications, standards, limitations, and likely future trends associated with the Semantic Web and its application to tourism. On the social side, the study attempted to build on previous research into online technology acceptance among small-to-medium tourist enterprises (SMTEs) (e.g. Morrison & King 2002), and provide an understanding of the managerial issues faced, and possible solutions for gaining wider industry acceptance as a practical means for tourism information integration and utilization. Specific aims of the research were to:

- Provide an understanding of issues and problems involved in defining, establishing, capturing, integrating and using the heterogeneous, scattered and diverse supplier source data necessary for the development of Semantic Web based tourism applications.

- Specify a theoretical and conceptual solution to these data-related problems that addresses technical limitations with existing integration approaches and takes into account the critical social dimension.
- Develop a proof of concept DMS prototype (based on the conceptual model discussed above), restricted to matching tourism customers accommodation needs to suppliers' offerings. This prototype (titled *AcontoWeb)* will be 'ontology-driven'.
- Demonstrate the effectiveness of the DMS with regard to usability and value-adding potential for tourism industry customers and service providers – via a survey and experiment.
- Gain an insight into the attitudes towards the adoption of semantic technology by SMTEs and their requirements and preferences in relation to implementation and usability of such systems.
- Generate a grounded hypotheses that can be tested in further research.

It is important to note here that the focus of this study was on information integration *via the Semantic Web*. Thus, while acknowledging the importance of integration theory in areas such as integration methodologies, data mapping algorithms and approaches, data integration in the absence of commonly-accepted international standards, and the implications of information loss during data mappings, a systematic evaluation of all types of possible model differences using for example, the metadata categorization scheme presented by Hsu (1996), was not undertaken. A rigorous investigation of this is beyond the scope of the study, but has been identified as a promising area for further research, that indeed could build upon the framework established here.

1.5 Research Question

The Australian tourism industry is an ideal domain for testing a new approach to online information integration because there are large numbers of SMTEs offering dispersed and unstructured information about services and attractions, which need to be matched to customers individual travel preferences. This provides the perfect opportunity to investigate how successfully tourism information can be integrated using Semantic Web technologies from a technical perspective and, from a managerial perspective, how likely it is that such an initiative will gain wider industry acceptance. The main outputs of this, essentially exploratory, study are tentative hypothesis to be validated in later research. The major research question is therefore defined as:

To what extent can the Semantic Web and related technologies assist with the creation, capture, integration, and utilization of accurate, consistent, timely, and up-to-date Web based tourism information?

The following minor research questions will also be investigated:

- What is the ease of ontology development, availability, and Website annotation?
- What level of ontology and Website annotation richness can be obtained?
- What is the maturity and ease of use of Semantic Web development tools?
- How robust are Semantic Web operational environments at present?
- How can the Semantic Web best be queried?
- What are the potential query results and accuracy?
- How do query results compare to that of conventional database systems?
- How useful is the Semantic Web and what are its limitations?
- How successfully can tourism information be integrated on the Semantic Web?
- What are the managerial issues faced in gaining user acceptance of Semantic Web technology in the tourism industry?

1.6 Research Approach

This section outlines the research approach, which was to formulate grounded theory through a systems development research method, supported by a survey and secondary data analysis.

1.6.1 Grounded Theory

The research aimed to generate grounded theory (Glaser 1967). Grounded theory is concerned with the generation of theory from research, as opposed to research that tests existing theory. With this approach, theories and models should be *grounded* in real empirical observations, rather than being governed by traditional methodologies and theories (Ticehurst & Veal 2000b). As Jones (1987, p. 25) notes, research should be used to generate grounded theory which "fits" and "works" because it is derived from the concepts and categories used by social actors themselves to interpret and organize their worlds. In the generation of theory, the researcher approaches the data with no pre-formed notions in mind, instead seeking to uncover patterns and contradictions through intuition and feelings. To achieve this, the researcher needs to be very familiar with the data, the subjects and the cultural context of the research. The process is a complex and personal one, as described in Strauss (1987) and Strauss and Corbin (1994).

Although a detailed review of grounded theory is outside the scope of this thesis, the theory is briefly described above to provide an understanding of the underlying philosophy of the research that was undertaken. A grounded theory approach was applied because it was best suited to the exploratory nature of the study, where notably the overarching aim was to observe and evaluate the implications or any other effects of introducing a new technology for the integration and utilization of Web based tourism information. In this case, the grounded hypothesis expresses a viewpoint as to the extent that the Semantic Web and related technologies are likely to assist with the creation, capture, integration, and utilization of accurate, consistent, timely, and up-to-date Web based tourism information.

1.6.2 Systems Development and Survey Type Research

A systems development approach, as described by Burstein (2002), supplemented with survey type research (i.e. Tanner 2002), was used to generate grounded theory. According to Cerez-Kecmanovic (1994), systems development has also been referred to as *engineering* type research also known as *social engineering*. Nunamaker et al. (1990-1991) assert that it is a developmental and engineering type of research, which falls under the category of applied science. It is grounded on the philosophical belief that development is always associated with exploration, advanced application and operationalization of theory (Hitch & McKean 1960 cited in Burstein 2002 p.151). The research approach may be classified as 'research and development' where scientific knowledge is used to produce '...useful materials, devices, systems, or methods, including design and development of prototypes and processes' (Blake 1978 cited in Nunamaker and Chen 1990, p. 631 and Burstein 2002, p.151).

Burstein (2002) explains that systems development denotes a way to perform research through the exploration and integration of available technologies to produce an artefact, system or system prototype. The design of such a system needs to be justified by some preliminary research undertaken to identify a problem and predict the likely success or failure of such a design for addressing the problem. Once the theory is proposed it needs to be tested to show its validity and to recognize its limitations, as well as to make appropriate refinements according to new facts and observations made during its application (Burstein 2002, p. 151).

In consideration of the available resources and the large scale of the tourism industry itself, it was decided that it would be more informative from a research perspective to focus on a specific sector of the tourism industry. Accommodation services represent the largest single economic sector of the Australian tourism industry[7]. It is for this reason, as well as geographical convenience, that data was collected and analysed from within the *Accommodation Services domain of the Australian Tourism Industry*. To provide the required holistic view of technology, people, structure and processes within this domain, systems development and an experiment were combined with a survey of tourism operators and analysis of secondary data interviews designed to provide insight into attitudes towards the adoption of a radical new Internet technology. These components of the research were conducted in the following two largely concurrent phases:

Phase 1: Development of a proof-of-concept DMS prototype called AcontoWeb, with the aim of evaluating and demonstrating the efficiency, benefits and limitations of the bottom-up (Semantic Web) approach to DMS development and implementation. The DMS prototype was in the form of a semantic portal, based on the layout, functionality, and data structure of the RACV (AAA tourism) accommodation portal. The system also contained an annotation module to allow tourism operators to add RDF metadata automatically to their Websites. The scope of the system was limited to the bare minimum, consistent with research objectives. An evaluation was made of perceived advantages for information integration of a portal based on Semantic Web standards, where a collection of resources is indexed using a rich domain ontology and a semantic search tool is applied, as opposed to a

7 Source: ABS (cat. no. 8635.0) available at: http://www.abs.gov.au/

conventional portal where information is indexed using a flat keyword list backed by a relational database and an SQL search engine. The SQL type search engine that the RACV portal uses (briefly described in section 1.2.2) is one of the main categories of search engines identified by Alesso & Smith (2004d).

Phase 2: An investigation into the issues affecting attitudes towards adoption of new on-line technologies among tourism operators, as well as the needs and preferences that operators have for implementation of such technologies. This investigation was designed to produce results that indicate potential interest in Semantic Web-based DMS, and may also be used to enhance SME take-up of the technology. To maintain a level of consistency with phase 1, a survey was conducted of accommodation providers listed on the RACV[8] (AAA tourism) portal. The survey included the following categories of accommodation:

- Hotel/Motels
- Apartment/Holiday Units
- Caravan Park/Camping Areas
- Chalet/Cottages
- Backpacker/Hostels
- Bed and Breakfast/Guesthouses
- Houseboat/Cruisers

The survey was supported by secondary data obtained from a research project documented in McGrath et al. (2005c), where semi-structured interviews were conducted about attitudes to adoption of online technology in the Australian tourism industry.

1.7 Thesis Outline

The thesis commences in Chapter one by providing background to the research topic. A brief history of the Internet and search engines is followed by a description of the integration problem, which is largely caused by the limitations of the current Internet. The research aims, questions, and approach are also discussed in Chapter 1.

In Chapter 2 the Semantic Web, tourism, and tourism ICT literature is reviewed. This provides an overview of previous work undertaken in the areas of the Semantic Web and

[8] http://www.accommodationguide.com.au/searchgateway.asp?sit=2&aid=1

tourism ICT, including the current state of tools, standards, applications, projects, managerial, and other issues.

Chapter 3 presents the research methodology. The chapter describes the research philosophy and phases, followed by designs for the query experiment and survey of tourism operators. Research limitations and threats to external validity are also discussed.

Chapter 4 presents a detailed software requirement specification (SRS) of the AcontoWeb semantic portal, and documents the results of the query experiment.

Chapter 5 discusses the findings of the tourism operator survey, with secondary interview data used to support the findings.

Finally, Chapter 6 presents the conclusion and discusses the overall research outcomes. This includes answers to major and minor research questions, and the proposition of a grounded hypothesis based on research findings. Potential directions for future research in the topic area are also discussed.

2 LITERATURE REVIEW

2.1 Chapter 2 Overview

The purpose of the literature review is to provide an in-depth analysis of previous research and industry work undertaken in both the fields of Semantic Web and tourism ICT. The chapter commences with an overview of the Semantic Web. This includes a discussion about existing and potential future application areas, markup languages associated with the Semantic Web, and a comprehensive introduction and overview of ontologies. Semantic search is introduced to provide background to the application development and experimental part of the thesis (Chapter 4). The benefits that can be achieved for integration and utilization of information through the use of semantics and inference are also demonstrated.

The chapter then focuses on state-of-the-art art applications and techniques available to assist with Semantic Web application development. This is to show the options that were available for developing the prototype semantic portal, as well as to help evaluate and report on Semantic Web applications and technologies in the thesis conclusion (Chapter 6). In order to provide a broad view of the Semantic Web, other important areas are covered including ontology merging and alignment techniques, Semantic Web services, and future trends and challenges associated with the technology. The other major topic area covered by the literature review is tourism ICT. The second part of the chapter focuses on this as well as the use of Semantic Web technologies in tourism.

2.2 The Semantic Web

This section reviews key aspects of the Semantic Web - including applications, markup languages, ontologies, semantic search, application development, Semantic Web services, ontology integration, challenges and future trends.

2.2.1 The Semantic Web Initiative

In 1992 Tim Berners-Lee created the World Wide Web Consortium (W3C) with the goal to develop, extend, and standardize the Web. W3C research eventually led to the conceptual development of the so called Semantic Web, that is described by Berners-Lee et al. (2001) as an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. Van Harmelen et al. (2000) describe the Semantic Web as a range of standards, modelling languages and tool development initiatives aimed at annotating Web pages with well-defined metadata, so that intelligent agents can reason more effectively about services offered at particular sites. Alesso & Smith (2004a) state that the goal of the initiative is to provide a machine-readable intelligence that would come from hyperlinked vocabularies that Web authors could use to explicitly define their words and concepts. and that the idea allows software agents to analyse the Web on our behalf, making smart inferences that go beyond the simple linguistic analysis performed by today's search engines. The foundations of the Semantic Web are based on powerful new markup languages such as Resource Description Framework (RDF) (Manola & Miller 2004), Ontology Web Language (OWL) (McGuinness & Harmelen 2004), and ontologies. Berners-Lee et al. (op. cit) identified the following three components as essential for the Semantic Web to function:

1. **Knowledge Representation** - Structured collections of information and sets of inference rules that can be used to conduct automated reasoning. Knowledge representation must be linked into a single system.
2. **Ontologies** - A document that formally describes classes of objects and defines the relationships among them.
3. **Agents** - Programs that have the ability to act autonomously by collecting content from diverse sources and exchange the results with other programs.

The following special research groups, which are listed on the W3C Website[9] as charted and part of the Semantic Web Activity, have formed to lead work on the creation of standards, as well as technology development:

- **Rules Interchange Working Group**[10] - is chartered to produce a core rule language with extensions that together allow rules to be translated between rule languages and rule systems. The group has to balance diverse community needs including business

[9] http://www.w3.org/2001/sw/

[10] http://www.w3.org/2005/rules/wg

rules, and a semantic users Web specifying extensions that can be used to articulate a consensus design sufficiently motivated by use cases.

- **RDF Data Access Working Group**[11] – has the task of evaluating the requirements for a query language and network protocol for RDF. The group also defines formal specifications and test cases for supporting such requirements.
- **The Semantic Web Coordination Group**[12] - is tasked to provide a forum to manage interrelationships and interdependencies among groups. The focus here is on standards and technologies relating to the goals of Semantic Web Activity. The group aims to avoid duplication of effort and fragmentation of the Semantic Web by way of incompatible standards and technologies through coordination, facilitation, and (where possible) helping to shape the efforts of other related groups.
- **Semantic Web Best Practices and Deployment (SWBPD) Working Group**[13] - this group provides hands-on support for developers of Semantic Web applications.
- **Semantic Web Interest Group**[14] - is a forum for W3C Members and non-Members for discussing innovative new Semantic Web applications. The group also initiates discussion about potential future work items for enabling technologies to support the Semantic Web, as well as the relationship of that work to other activities of the broader social and legal context in which the Web and the W3C are situated.
- **Semantic Web Services Interest Group**[15] - provides an open forum for W3C members and non-members to discuss Web services topics oriented towards the integration of Semantic Web technology into the ongoing Web services work at the W3C.
- **Semantic Web Health Care and Life Sciences Interest Group**[16] - aims to improve research and development, collaboration, and innovation adoption in the life science

[11] http://www.w3.org/2001/sw/DataAccess/

[12] http://www.w3.org/2001/sw/CG/

[13] http://www.w3.org/2001/sw/BestPractices/

[14] http://www.w3.org/2001/sw/interest/

[15] http://www.w3.org/2002/ws/swsig/

[16] http://www.w3.org/2001/sw/hcls/

and health care industries. The group aids decision making in clinical research, so that Semantic Web technologies will one day be capable of bridging many forms of biological and medical information across institutions.

2.2.2 Semantic Web Application Domains

This sub-section provides an overview of the various Semantic Web application domains that exist today, as well as potential future application areas.

2.2.2.1 Semantic E-Business

The following areas of e-business are widely reported in the Artificial Intelligence (AI) literature as most likely to benefit by future adoption of Semantic Web technologies:

- **Supply Chain Management (SCM)** - Described by Poirer & Bauer (2001), as a common strategy employed by businesses to improve organizational processes to optimize the transfer of goods, information and services between buyers and suppliers in the value chain. Singh, Lakshmi et al. (2005) believe that a standard ontology for trading partners is necessary for seamless transformation of information, and that knowledge is essential for supply chain collaboration.
- **E-Marketplaces** – in these environments intermediaries perform a critical role in bringing together buyers and suppliers in an e-marketplace and facilitating transactions between them. Singh & Iyer (2003) contend that the integration of intelligence and knowledge within and across e-marketplaces can enhance the coordination of activities among collaborating firms.
- **Healthcare** - Pollard (2004) states that knowledge management activities in healthcare centre on the acquisition and storage of information, and presently lack the ability to share and transfer knowledge across systems and organizations to support individual user productivity. Semantic Web technologies can enable health information integration, thus providing the transparency for healthcare-related processes involving all entities within and between hospitals, as well as stakeholders such as pharmacies, insurance providers, healthcare providers, and clinical laboratories. According to Eysenback (2003), such innovations can lead to enhanced caregiver effectiveness, work satisfaction, patient satisfaction, and overall quality in healthcare.

- **E-government** - refers to the use of Internet technologies for the delivery of government services to citizens and businesses. The aim of e-government is to streamline processes and improve interactions with business and industry, empower citizens with the right information, and improve the efficiency of e-government management (Teswanich, Anutariya & V 2002, p. 30). Teswanich et al. (op. cit) state that there is a critical need to manage the knowledge and information resources stored in these disparate systems, and that emerging Semantic Web technologies can enable transparent information and knowledge exchange to enhance e-government processes. After comprehensively examining the use of Semantic Web based e-commerce applications for e-government services, Klischewski & Jeenicke (2004) concluded that although such applications and functions are integral, at present it is very difficult to recommend technical solutions and identify best practices in this area, and that further research is therefore required.
- **E-Learning** - Semantic Web technologies are widely used in e-learning because they meet the most important e-learning requirements: quickness, just-in-time learning, and pertinence (Castellanos & Fernández 2004, p. 61). Learning materials can be efficiently semantically annotated so these materials can be reused in different courses. Moreover, access to content can be customized according to student needs and preferences. The adjustment of the Semantic Web to e-learning needs is illustrated by Stojanovic (2001). There, the following issues concerning the Semantic Web were considered: 1) knowledge items are distributed on the Web and they are linked to consensus ontologies; 2) the user makes semantic searches for desired materials; 3) the Semantic Web has the potential to become an integration platform for business processes; 4) there is active information delivery to create a dynamic learning environment; 5) authority is as decentralized as possible; 6) users search for material suited to their needs; 7) the Semantic Web allows for using the knowledge provided in different forms through the semantic annotation of materials; and 8) each user has a personalized agent that communicates with other agents to obtain materials.

A major area of semantic e-business not covered in this sub-section is e-tourism, which is reviewed in section 2.3.

2.2.2.2 Semantic Portals

Web portals are entry points for information presentation and exchange over the Internet used by a community of interest. Hence, they require efficient support for communication and information sharing. Lara et al. (2004) state that current Web technologies present serious limitations regarding information search, access, extraction, interpretation and processing, and that these limitations are naturally inherited by existing Web portals, thus hampering the communication and information sharing process between community members. The application of Semantic Web technologies has the potential to overcome these limitations and, therefore, used to evolve current Web portals into semantically enhanced Web portals.

The notion of semantic portals is that a collection of resources is indexed using a rich domain ontology, as opposed to say, a flat keyword list. Search and navigation of the underlying resources then occur by exploiting the structure of this ontology. Reynolds (2001) explains that this allows search to be tied to specific facets of the descriptive metadata and to exploit controlled vocabulary terms – leading to much more precise searches. There are several advantages inherent in using Semantic Web standards for portal design compared to traditional portals. Lara et al. (op. cit) believe that a main benefit is the ability to model a portal's structure using ontologies as the starting point. Ontologies are best suited to represent consensus knowledge and its structure. According to Lara et al. (op. cit), this is exactly what is needed to exchange information within a community of interest and to enable automated processing of information items.

Reynolds (op. cit) sees the decentralized nature of Semantic Web technologies as another major advantage, because this makes it possible for portal information to be an aggregation of a large number of small information sources instead of a single central location where people submit information. The portals can be reorganized to suit different user needs while the domain indexes remain stable and reusable. Communities of interest can share access to the same underlying information using a different navigation structure, search facility and presentation format. Reynolds (op. cit) adds that in this situation, central organization is still needed in the initial stages to provide the start-up impetus and ensure that appropriate ontologies and controlled vocabularies are adopted. Once the system reaches a critical mass though, information providers can then take responsibility for publishing their own information, provided it is annotated in accordance with the correct domain ontology.

An example of this decentralized approach is the ARKive portal[17], which publishes multimedia objects depicting endangered species. ARKive just provides the backbone structure of resources by making its ontology available for use. Individual communities of interest then supply the additional classifications, annotations, and navigational interfaces to suit their needs. The application of Semantic Web technologies also makes it easier to integrate data across portals by applying mapping and merging techniques to shared or compatible ontologies. Techniques for ontology integration are discussed in sub-section 2.2.8. Table 1 shows a comparison of traditional and semantic portals.

Traditional design approach	Semantic Portals
Search by free text and stable classification hierarchy.	Multidimensional search by means of rich domain ontology.
Information organized by structured records, encourages top-down design and centralized maintenance.	Information semi-structured and extensible allows for bottom-up evolution and decentralized updates.
Community can add information and annotations within the defined portal structure.	Communities can add new classification and organizational schemas and extend the information structure.
Portal content is stored and managed centrally.	Portal content is stored and managed by a decentralized Web of supplying organizations and individuals. Multiple aggregations and views of the same data are possible.
Providers supply data to each portal separately through portal-specific forms. Each copy has to be maintained separately.	Providers publish data in reusable form that can be incorporated in multiple portals but updates remain under their control.
Portal aimed purely at human access. Separate mechanisms are needed when content is to be shared with a partner organization.	Information structure is directly machine accessible to facilitate cross-portal integration.

Table 1: Comparison of traditional and semantic portals (Reynolds 2001).

2.2.3 Semantic Web Projects

The following are examples of real world Semantic Web project initiatives:

- **The DARPA Virtual Soldier Project**[18] – aims to enhance diagnosis and prognosis of battlefield injuries by using an OWL ontology. The goal is to investigate methods that will revolutionize medical care for the soldier. The project is integrated into the

[17] http://www.arkive.org/

[18] http://www.virtualsoldier.net/

protégé general framework and uses ontology reasoning to produce complex mathematical models that create physiological representations of individual soldiers. These holographic medical representations (known as Holomers) can be used to improve medical diagnosis on and off the battlefield. The Holomers coupled with predictive OWL reasoning, facilitate a new level of integration in medical procedures. The Virtual Soldier provides multiple capabilities, including automatic diagnosis of battlefield injuries, prediction of soldier performance, evaluation of non-lethal weapons, and virtual clinical trials.

- **CS AKTive Space (CAS)**[19] **-** winner of the 2003 Semantic Web challenge[20], this is an integrated Semantic Web application that provides a way to explore the UK Computer Science research domain across multiple dimensions for multiple stakeholders, from funding agencies to individual researchers. One of the challenges for the Semantic Web is to represent large ontological spaces in meaningful ways to people who wish to explore them. The goal of the interaction design for CS AKTive Space has been to explore this Semantic Web challenge by providing a user interface to millions of triples from multiple heterogeneous sources that represent the UK Computer Science domain. The project uses an ontology to provide seamless integration and on-demand semi-automatic content harvesting from multiple semantically heterogeneous data sources to provide information access.

- **Swoogle**[21] - is a search engine / Web crawler-based indexing and retrieval system for Semantic Web documents in RDF or OWL. It is being developed by the Computer Science and Electrical Engineering Department of the University of Maryland Baltimore County. It extracts metadata and computes relations between documents. Discovered documents are also indexed by an information retrieval system to compute the similarity among a set of documents and to compute rank as a measure of the importance of a Semantic Web document. Swoogle facilitates the development of the Semantic Web by finding appropriate ontologies, and helping users specify

[19] http://cs.aktivespace.org/

[20] http://www.informatik.uni-bremen.de/agki/www/swc/swc2003submissions.html

[21] http://swoogle.umbc.edu/

terms and qualify types (class or property). In addition, the ranking mechanism sorts ontologies by their importance.

In order to help users integrate Semantic Web data (SWD) distributed on the Web, Swoogle enables querying SWDs with constraints on the classes and properties. By collecting meta-data about the Semantic Web, Swoogle reveals interesting structural properties such as how the Semantic Web is connected, how ontologies are referenced, and how an ontology is modified externally. Swoogle is designed as a system that will scale up to handle millions of documents. Moreover, Swoogle also enables rich query constraints on semantic relations. The Swoogle architecture consists of a database that stores metadata about SWDs. Two distinct Web crawlers discover SWDs and components to compute semantic relationships among the SWDs. Also, an indexing and retrieval engine and simple user interface for query and agent Web service APIs provide useful services.

- **MUSEUMFINLAND**[22] - is a semantic portal that was built by the Museum of Finland using the Ontoviews tool (see sub-section 2.2.7.3). In MUSEUMFINLAND, the content consists of collections of cultural artefacts and historical sites in RDF format consolidated from several heterogeneous Finnish museum databases (Hyv¨onen, Salminen & Kettula 2004). M¨akel¨a et al. (2004) explain that the RDF content is annotated using a set of seven ontologies. From the seven ontologies, nine view-facets are created. The ontologies underlying the application consist of some 10,000 RDF(S) classes and individuals, half of which are in use in the current version on the Web. There are some 7,600 categories in the views. Search for museum content can be done via keywords which return a list of semantically related categories displayed as hyperlinks. Searches may also be performed by navigating hyperlinks alone without entering keywords.
- **INWISS knowledge portal prototype**[23] - is a semantic portal that demonstrates an approach for communicating user context (revealing the user's information need) among portlets (components of a Web portal) utilizing Semantic Web technologies. For example, the query context of an OLAP portlet, which provides access to

[22] http://www.museosuomi.fi/

[23] http://www-ifs.uni-regensburg.de/inwiss/index.jsp

structured data stored in a data warehouse, can be used by a search portlet to automatically provide the user with related intranet articles or documents in an organization's document management system.

- **The ARTENIS project**[24] - is a peer-to-peer network based on Semantic Web services and suitable domain ontologies. The project, which is funded by the European Union, aims at improving interoperability among different health organizations. Organizations can join the ARTEMIS network and advertise electronic services, such as access to a patient's health care record (with appropriate authorization), as well as access to different subsystems such as patient admission or laboratory information systems. The network also allows other services to be invoked dynamically. One such example is the possibility to dynamically map different representations of health care information. ARNENIS' participating partners originate from Turkey, Greece, Germany, and the United Kingdom.

- **The DartGrid**[25] - is a Semantic Web based toolkit that mainly aims to resolve the problem of heterogeneous database integration in a specific VO (virtual organization). The system combines Grid and Semantic Web technologies together to provide a uniform semantic query interface to sets of distributed and heterogeneous relational data sources. What the DartGrid project contributes is at the semantic service level. The services at this level are designed for semantic-based relational schema mediation and semantic query processing by using the semantics of an RDFS ontology.

- **The Bioinformatics project for Glycan Expression**[26] **-** has the overarching objective of building a semantic framework for integration, sharing, storage, and retrieval of vast amounts of data generated by high-throughput glycomics experiments. Glycomics is the study of glycans (modifications of sugar molecules) expressed by an organism, which plays a critical role in the life functions of organisms. The project, which is explained in more detail by Sheth (2005), forms the bioinformatics component of the Biomedical Glycomics Research Resource Centre at the Complex Carbohydrate Research Centre (CCRC). As part of the semantic

[24] http://www.srdc.metu.edu.tr/Webpage/projects/artemis/

[25] http://ccnt.zju.edu.cn/projects/dartgrid/

[26] http://lsdis.cs.uga.edu/projects/glycomics/

framework, two large ontologies were developed, namely GlycO (a glycoproteomics domain ontology with extremely fine granularity) and ProPreO (a process ontology with comprehensive modelling of glycoproteomics experimental lifecycle). An XML-based standard for representation of glycan structures called GLYDE (Glycan Data Exchange) has been proposed and implemented. GLYDE is now being seriously considered for adoption by the international glycomics community for data exchange.

2.2.4 Semantic Web Markup Languages

This sub-section describes the evolution and capabilities of mainstream Semantic Web markup languages.

2.2.4.1 XML and XML Schema

XML (Bray et al. 2004) was developed by the W3C XML working group in the late 1990's to provide rules for creating vocabularies that can structure both documents and data on the Web. The aim of XML was to overcome some of the drawbacks of HTML, which was designed for information presentation rather than machine processing. XML provides clear rules for syntax, while XML schemas extend these capabilities by serving as a method for composing XML vocabularies against which documents can be validated. XML is a powerful, flexible surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents. It is not possible for example, to deduce new knowledge from an XML statement. More powerful Web markup languages are required to perform sophisticated information processing tasks.

2.2.4.2 Resource Description Framework (RDF)[27]

RDF (Manola & Miller 2004), which stands for Resource Description Framework, is a data model and syntax specification for representing information about Web resources. An RDF Model is a set of statements, each consisting of a triple (i.e. subject, predicate, object). RDF statements can either be represented as a graph or in an XML format known as RDF/XML serialization. Figure 2 is an example of a simple RDF graph which is demonstrated in an RDF tutorial by Decker et al. (2000). The resource

[27] The RDF syntax specification is available at: http://www.w3.org/TR/rdf-syntax-grammar/

“http://www.daml.org/projects/#11” is a subject with a property “http://www.SemanticWeb.org/schema-daml-01/#hasHomepage.” The value of the property is the object “http://wwwdb.stanford.edu/OntoAgents.” The property “http://purl.org/DC/#Creator” with value “Stefan Decker” (a literal) is joined to form the graph.

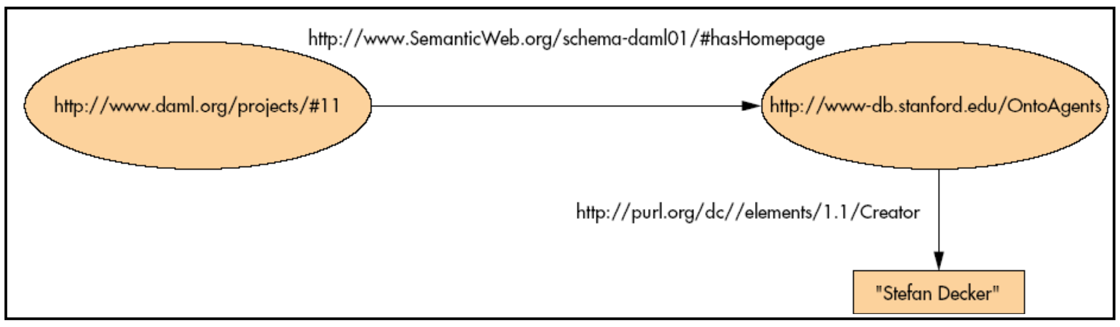


Figure 2: RDF graph (Decker, Mitra & Melnik 2000).

An RDF graph is not the most efficient means of storing and retrieving data. RDF/XML serialization is a process that converts the graph into an XML format that is machine processable. The Figure 2 graph is represented in Figure 3 using RDF/XML serialization.

```
<Project rdf :about="http://www.daml.org/projects/#11">
    <hasHomepage>
        <rdfs:Resource rdf:ID="http://www-db.stanford.edu/OntoAgents">
            <dc:Creator>Stefan Decker</dc:Creator>
        <rdfs:Resource>
    </hasHomepage>
</Project>
```

Figure 3: RDF/XML serialization (Decker, Mitra &

2.2.4.3 RDF Namespaces

To ensure that RDF/XML data from various documents can be successfully merged, namespaces are added to RDF specifications. Namespace declarations act as a prefix for identifying the vocabulary in a document, as well as pointing to the URI of any external RDF vocabulary that is used. Figure 4 is an example RDF namespace declaration presented by Manola & Miller (op. cit).

```
<?xml version="1.0"?>
 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
             xmlns:exterms="http://www.example.org/terms/">
   <rdf:Description rdf:about="http://www.example.org/index.html">
       <exterms:creation-date>August 16, 1999</exterms:creation-
date>
   </rdf:Description>
 </rdf:RDF>
```

Figure 4: RDF namespace (Manola & Miller 2004).

2.2.4.4 RDF Schema (RDFS)

RDF is limited to descriptions about individual resources, and does not provide any modelling primitives for defining the relationships between properties and resources. To demonstrate this, Gomes-Perez et al. (2004b) provide the example that a relation arrivalPlace expressed in RDF can only hold between instances of the classes *Travel* and *Location*. This limitation is solved by RDFS (Brickley & Guha 2003), which extends RDF by providing a vocabulary by which classes and their subclass relationships can be expressed, and properties defined and associated with classes. RDFS achieves this by adding 16 new modelling primitives for organizing Web objects into hierarchies. This allows objects to be grouped together into classes, making it possible to link together instances of these classes. For example, a class called *Accommodation* could be linked to a class called *Location* via a predicate relation (property) called *hasLocation*. This means that any instance of the *Accommodation* class could be defined as having a particular location which would be an instance of the *Location* class.

It is also demonstrated by Antoniou et al. (2005), that the application of predicates can be restricted with RDFS through the use of domain and range restrictions. For example, it becomes possible to restrict the property *hasLocation* to apply only to instances of the class *Accommodation*, and have only instances of the class *Location* as values. This way, nonsensical statements (e.g. Accommodation has the Location of Five Star Rating) due to user errors can be automatically detected. Even though RDF and RDFS are building blocks for defining a Semantic Web, they still lack sufficient expressive power for building sophisticated ontologies. They are not capable for example, of defining properties of properties, equivalence and disjointness of classes. Even more expressive markup languages are therefore required.

2.2.4.5 DAML + OIL

Ontology Interchange Language (OIL) (Horrocks 2000), was developed by Dr Ian Horrocks at the University of Manchester as an extension to RDFS. OIL added more frame based KR primitives and used description logics to give clear semantics to these primitives. At the same time that OIL was being developed, the Defense Advanced Research Agency (DARPA) began working on DARPA Agent Markup Language (DAML) (Horrocks & Harmelen 2001). Similar to OIL, DAML was designed with a greater capacity than RDF and RDFS for describing objects and the relationships between them. DARPA developed DAML as a technology with semantics built into the language to assist the capabilities of Web agents, which are programs that can dynamically identify and comprehend sources of Information (see 2.2.5.10). Soon efforts were underway around the world to unify the various ontology languages. These efforts led to a new language called DAML+OIL, which consolidated the capabilities of DAML and OIL to further overcome many of the expressive capability inadequacies of RDFS. DAML+OIL was soon to be superseded, however, by another language known as OWL.

2.2.4.6 Ontology Web Language (OWL)[28]

The OWL language, which was created by the W3C Web Ontology (WebOnt)[29] Working group derive from DAML+OIL. Like DAML+OIL, OWL builds on RDF and RDF Schema and adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes (McGuinness & Van Harmelen 2004). OWL ontologies have three species or sub-languages: OWL-Lite, OWL-DL and OWL-Full. A defining feature of each sub-language is its expressiveness. OWL-Lite is the least expressive sub-language. It is intended to be used in situations where only a simple class hierarchy and simple constraints are needed. OWL-Full is the most expressive sub-language. It is intended to be used in situations where very high expressiveness is more important than being able to guarantee the decidability or computational completeness of the language. OWL-DL was designed to be processed by description logic reasoners. The expressiveness of OWL-DL falls

[28] The OWL language and abstract syntax specification can be found at: http://www.w3.org/TR/owl-absyn/

[29] http://www.w3.org/2001/sw/WebOnt/

between that of OWL-Lite and OWL-Full. OWL-DL is an extension of OWL-Lite, and OWL-Full an extension of OWL-DL. Horridge (2004) contends that there are simple rules of thumb that should be considered when deciding which language to use when building an ontology. For example, the choice between OWL-Lite and OWL-DL may be based upon whether the simple constructs of OWL-Lite are sufficient or not. Also, the choice between OWL-DL and OWL-Full may be based upon whether it is important to carry out automated reasoning on the ontology, or whether it is important to be able to use highly expressive and powerful modelling facilities such as meta-classes (classes of classes).

2.2.4.7 Markup Language Pyramid

Figure 5 represents the stack of mainstream markup languages for the Semantic Web starting from XML and XML Schema, followed by RDF and RDFS. On top of the pyramid sits OWL. The languages offer an increasing degree of expressiveness with lower lever languages syntactically compatible with those at the upper levels. Their development is based on a history of different languages which have all to some degree, contributed to the final W3C standards that now form a stable basis for Semantic Web development (Stuckenschmidt & Harmelen 2005a, p. 61).

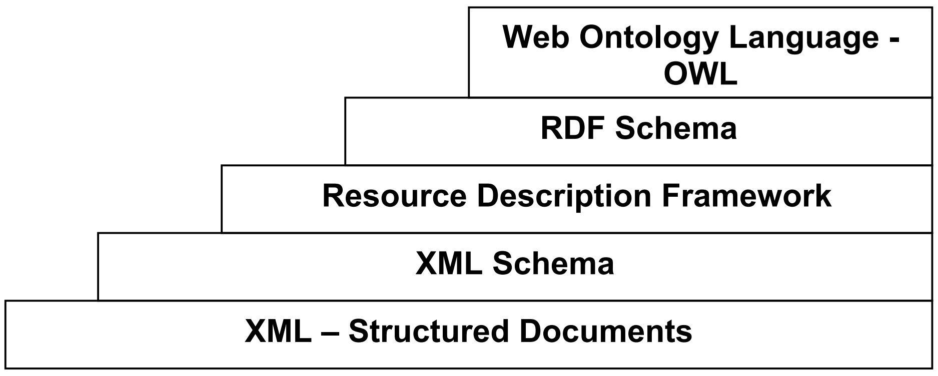


Figure 5: Markup language pyramid (Alesso, PH & Smith, CF 2004).

The language pyramid is sometimes presented in a more generalized manner based on logical classifications at the upper levels. The Semantic Web tower by Antoniou et al. (2005) (see Figure 6) presents four logical levels on top of RDFS.

1. **The ontology language layer** - expands RDFS and allows the representation of more complex relationships between Web objects. Languages such as DAML + OIL and OWL fit into this category.

2. **The logic layer** - is used to enhance the ontology language further, and to allow the writing of application specific declarative knowledge.
3. **The proof layer** - involves the actual deductive process, as well as the representation of proofs in Web languages and proof validation.
4. **Trust layer** - will emerge through the use of digital signatures, and other kind of knowledge based on recommendations by agents and consumer bodies.

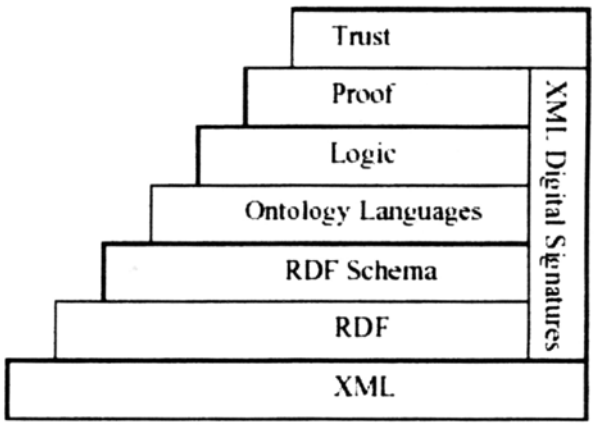


Figure 6: The Semantic Web tower (Antoniou et al. 2005).

2.2.5 Ontologies

This sub-section provides a detailed description of ontologies, including their various definitions, purposes, application areas, and issues concerning their development.

2.2.5.1 Defining Ontologies

The word ontology stems from philosophy where it means a systematic explanation of being (Gomes-Perez, Fernandez-Lopez & Corcho 2004c, p. 6). Since the late 1990's the word has become relevant in the information systems and artificial intelligence community. It is within this context that Neches et al. (1991) defines ontology as: *The basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary*. This definition identifies that an ontology consists of basic terms, relations between terms and rules that combine terms. Neches et al. (op. cit) also contend that an ontology includes both explicitly defined terms and the knowledge that can be inferred from it.

The most widely quoted definition of ontology in the AI literature is by Gruber (1993a), who defines an ontology as: *A formal specification of a shared conceptualization*. Many definitions have since appeared that are based on Gruber (op. cit). For example, Borst (1997, p. 12) has slightly modified the definition to: *A formal specification of a shared conceptualization*. Struder et al (1998, p. 185) merged Gruber's (op. cit) and Borst's (op. cit) definitions and described an ontology as: *A formal, explicit specification of a shared conceptualization*. Conceptualization here refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that an ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, i.e. it is not private to some individual, but accepted by a group.

There are many definitions of *ontology* in the Artificial and Web Intelligence literature - many more than presented here. It can be said for the most part, though, that these definitions only provide different perspectives of the same reality which is that: *ontologies aim to capture consensual knowledge in a generic way*, and as Gomes-Perez et al. (op. cit) explain: *they may be reused and shared across software applications and by groups of people.*

2.2.5.2 Types of Ontologies

This sub-section describes the various types (or categories) of ontologies that exist. There are many ways in which ontologies can be categorized. For Example, Mizoguchi et al. (1995) define the following four ontology classifications:

1. **Content** - for reusing knowledge. These ontologies include other subcategories: task ontologies, domain ontologies and general or common ontologies.
2. **Communication (tell & ask)** - for sharing knowledge.
3. **Indexing** - for case retrieval.
4. **Meta-ontologies** - Mizoguchi (op. cit) say that these are equivalent to what other authors refer to as a knowledge representation ontology.

Van Heijst et al. (1997) classify ontologies by two orthogonal dimensions:

1. **The amount and type of structure of the conceptualization** – these are divided into three categories: 1) terminological - ontologies such as lexicons; 2) information ontologies such as database schema; and 3) knowledge modelling ontologies that specify conceptualizations of the knowledge.
2. **The subject of the conceptualization in the second dimension** – these are divided into four categories: representation, generic, domain and application ontologies.

Guarino (1998) defines the following three classifications of ontologies according to their level of dependence on a particular task or point of view:

1. **Top Level** - Guarino (op. cit) says that these contain very general concepts like space, time, matter, object, event, action. etc., which are independent of a particular problem or domain. It seems therefore reasonable, at least in theory, to have unified top level ontologies for large communities of users.
2. **Domain Level** – are task ontologies and describe respectively, the vocabulary related to a generic domain (like medicine, or automobiles) or a generic task or activity (like diagnosing or selling). This is done by specializing the terms introduced in the top-level ontology.
3. **Application Level** – describe concepts depending on both a particular domain and task, which are often specializations of both the related ontologies. These concepts often correspond to roles played by domain entities while performing a certain activity like replacing a unit or spare component.

Gomes-Perez et al. (2004c) classify ontologies similarly to Guarino (op. cit). In this case they are viewed as either:

- **Upper Level** - describing general concepts and providing general notions under which all root terms in existing ontologies should be links.
- **Domain Level** - provide vocabularies about concepts within a domain and their relationships about the activities taking place in that domain, and about the theories and elementary principles governing the domain.

Finally, Lassila & McGuinness (2001) classified ontologies according to the information that the ontology needs to express and the richness of its internal structure. They identify the following nine categories:

1. **Controlled vocabularies (i.e. a finite list of terms)** - a typical example of this category is a catalogue.
2. **Glossaries** - a list of terms with their meanings specified as natural language statements.
3. **Thesauri** - provides some additional semantics between terms. They give information such as synonym relationships, but do not supply an explicit hierarchy. For instance, traveller and passenger could be considered as synonyms in a travel domain.
4. **Informal is-a hierarchies** - taken from specifications of term hierarchies like Yahoo's. Such a hierarchy is not a strict subclass or "is-a" hierarchy. For instance, the terms car rental and hotel are not kinds of travel, but they could be modelled in informal is-a hierarchies below the concept travel, because they are key components of the travel and allow the user to select either a car rental for the trip or an accommodation option.
5. **Formal is-a hierarchies** - in these systems, if *B* is a subclass of *A* and an object is an instance of *B*, then the object is an instance of *A*. Strict subclass hierarchies are necessary to exploit inheritance. For example, subclasses of a concept *Travel* could be: *Flight*, *Train-Travel*, etc.
6. **Formal is-a hierarchies that include instances of the domain** - this case would include instances of flights: the *flight AA7462* arrives in *Seattle*, departs on *February 8*, and costs *$300*.
7. **Frames** - the ontology includes classes and their properties, which can be inherited by classes of the lower levels of a formal is-a taxonomy. For example, travel has a unique *Departure-date* and an *Arrival-date*, a company *Name*, and at most one *Price* for a single fare with the company. All these attributes could be inherited by the subclasses of a concept *Travel*.
8. **Ontologies that express value restriction** - these are ontologies that may place restrictions on the values that can fill a property. For instance, the type of the property *Arrival-date* is a date.

9. **Ontologies that express general logical constraints** - these are the most expressive. Ontologists can specify first-order logic constraints between terms using expressive ontology languages. A logical constraint in Lassila & McGuinness' (op. cit) travelling domain for example, is that it is not possible to travel from the USA to Europe by train.

2.2.5.3 Ontology Application Domains

Ontologies are widely used in Knowledge Engineering, Artificial Intelligence, and Computer Science in applications related to knowledge management, natural language processing, e-commerce, intelligent information integration, information retrieval, database design, bio-informatics, education, and in new emerging fields like the Semantic Web (Gomes-Perez, Fernandez-Lopez & Corcho 2004c, p. 1). In recent times, considerable progress has been made in developing the conceptual bases to build technology that allows reusing and sharing knowledge-components. According to Gomes-Perez et al. (op. cit), ontologies and Problem Solving Methods (PSMs) have been created to share and reuse knowledge and reasoning behaviour across domains and tasks.

Ontologies are concerned with static domain knowledge, while PSMs deal with modelling reasoning processes. Benjamins & Gomez-Perez (1999) state that a PSM defines a way of achieving the goal of a task. It has inputs and outputs and may decompose a task into subtasks and tasks into methods. Benjamins & Gomez-Perez (op. cit) add that a PSM specifies the data flow between its subtasks, and that an important PSM component is its method ontology because it describes the concepts used by the method on the reasoning process, as well as the relationships between such concepts.

Bylander et al. (2003) first raised the idea that the integration of ontologies and PSMs is a possible solution to the interaction problem. They state said that representing knowledge for the purpose of solving some problem was strongly affected by the nature of the problem and the inference strategy applied to the problem. Through ontologies and PSMs, this interaction can be made explicit in the notion of mappings between the ontology of the domain and the method ontology. Previously there have also been interesting studies done on the integration of ontologies and PSMs, such as that by Park (1998). The emergence of the Semantic Web marked another stage in the evolution of ontologies (and PSMs). The first ontologies represented static domain knowledge, but with the advent of more expressive Web markup languages such as OWL and RDF, PSMs are now used inside Semantic Web services that model reasoning processes and deal with that domain knowledge.

2.2.5.4 Ontology Development Process

Denny (2002) proposes that a number steps are required for developing an ontology. According to Denny (op. cit), the steps are straightforward and typically involve the following processes:

1. **Acquire domain knowledge** - assemble appropriate information resources and expertise that will define, with consensus and consistency, the terms used formally to describe things in the domain of interest. These definitions must be collected so that they can be expressed in a common language selected for the ontology.
2. **Organize the ontology** - design the overall conceptual structure of the domain. This will likely involve identifying the domain's principal concrete concepts and their properties, identifying the relationships among the concepts, creating abstract concepts as organizing features, referencing or including supporting ontologies, distinguishing which concepts have instances, and applying other guidelines of the chosen methodology.
3. **Flesh out the ontology** - add concepts, relations, and individuals to the level of detail necessary to satisfy the purposes of the ontology.
4. **Ontology check** - reconcile syntactic, logical, and semantic inconsistencies among the ontology elements. Consistency checking may also involve automatic classification that defines new concepts based on individual properties and class relationships.
5. **Commit the ontology** - incumbent on any ontology development effort is a final verification of the ontology by domain experts and the subsequent commitment of the ontology by publishing it within its intended deployment environment.

2.2.5.5 Ontology Development Methodologies

In recent years, a series of different methodologies designed to assist with carrying out development tasks have been reported in the Artificial Intelligence literature. Classical methods include Cyc (Lenat & Guha 1990), Uschold and King (Uschold & King 1995), Gruninger and Fox (Gruininger & Fox 1995), Kactus (Kactus 1996), and Methontology (Fernandez-Lopez, Gomes-Perez & Juritso 1997). The methodologies provide common and structured guidelines, which if followed, can speed up the development process and improve the quality of the end result. A survey conducted by Mendes (2003) identified thirty three proposed methodologies for ontology construction. Mendes (op. cit) classified these methodological approaches into five categories: 1) constructing from the beginning; 2) integration or fusion with other ontologies; 3) re-engineering; 4) collaborative constructing; and 5) evaluation of built ontologies. Arguably the most popular ontology design methodology (supported by ontology engineering environment WebODE) is "Methontology". Cristani et al. (2005) explain that Methontology defines a flow of ontology development processes for three different types of activities: 1) management; 2) technical; and 3) supporting. The complete Methontology framework is presented as Appendix A.

2.2.5.6 Ontology Devlopment Tools

A number of development and editing tools are available to ease the complex and time consuming task of building ontologies. Tools such as Kaon[30], OileEd[31], *and* Protégé[32] provide interfaces that help users carry out some of the main activities required for developing an ontology. Selecting the most appropriate editor, however, is a challenge because each ontology construction initiative requires its own budget, time, and resources. To help overcome this challenge, Singh & Murshed (2005) proposed criteria to evaluate ontology construction tools. The criteria include functionality, reusability, data storage, complexity, association, scalability, resilience, reliability, robustness, learn-ability, availability, efficiency, and visibility. Protégé and OntoEdit[33] Free (the predecessor to Ontostudio), were evaluated by Singh & Murshed (op. cit) using this criterion. The evaluation concluded that the editors provide similar functionality.

[30] Kaon version 1.2.7 available at: http://kaon.semanticWeb.org/

[31] OildEd version 3.5 available at: http://oiled.man.ac.uk/

[32] Protégé 3.2 available at: http://protege.stanford.edu/

[33] http://www.ontoknowledge.org/tools/ontoedit.shtml

A survey of ontology editors conducted by Denny (2002) classified available commercial products as either standalone editors designed exclusively for building ontologies in any domain, or editors that are part of commercial software suites designed to deliver broad enterprise integration solutions. Denny (op. cit) concluded that non-commercial editing software were generally the outcome of academic and government funded projects investigating the technical application of ontologies, with some intended for building ontologies in a specific domain. The later type of editors were still capable of general-purpose ontology building regardless of content focus.

Probably, the most comprehensive survey of ontology editors conducted to date is that of Damjanoviæ et al. (2004). Their survey was based on the following six criteria: 1) general description of the tools (such as information about developers, releases and availability); 2) software architecture and tool evolution; 3) interoperability with other ontology development tools and languages; 4) knowledge representation paradigm (knowledge model used); 5) inference services attached to the tool; and 6) tool usability. Damjanoviæ et al. (op. cit) concluded that:

- Ontology languages from the pre-XML era have matured. Unfortunately, Damjanoviæ et al. (op. cit) found that unlike these, tools and ontology development languages from the XML-era still aren’t mature. Hence, the tools are continuously evolving. New research areas emerge from deploying intelligent Web services (a combination of the emerging Semantic Web and Web services technologies), but require new research efforts, new development tools, and new tools for dynamic management of the Web.
- From the criteria for ontology development tool extensibility, Damjanoviæ et al. (op. cit) reported a trend of further adaptation of existing ontology development tools to the new Web standards (W3C recommendations), such as RDF (Resource Definition Framework), OWL (Web Ontology Language). They also stressed the importance of the newly proposed ISO standard, known as CL (Common Logic) that will be compatible with all the accepted W3C standards. Damjanoviæ et al. (op. cit) state, however, that this trend is not equally represented in all of these tools. This is because certain problems relate to ontology development tool interoperability. Usually, different research groups develop different tools and as a consequence, ontology development environments and tools are not interoperable. These tools have different

knowledge models, use different technology, and it is often difficult to integrate them. More recent ontology development tools allow for exporting and importing ontologies in XML and other markup languages as a means of exchanging ontologies between the tools. This can improve interoperability.

- Like the ontology development tools extensibility criteria, the portability criteria pertain to the ability of a tool to adapt easily to a new environment. Damjanoviæ et al. (op. cit) say that a good example of this is Protégé, which has a component framework for easily integrating other components via plug-ins. Thus, it was concluded that Protégé brings with it a great potential to expand, and also to adapt itself to the new (ontology-based) development environment. But this is not the case with all tools.
- The 'ease-of-use' criteria was claimed to be very important since it implies a necessity to use intuitive screen designs for anyone who will work in the area of ontology development, maintenance, deployment, merging, and update. However, current ontology development tools require their users to be trained in knowledge representation and abstraction.
- Finally, the study found that the use of ontology development tools in the sense of discovery and search criteria is important in the Web environment to find some potentially interesting new knowledge. Moreover, this criterion is related to the ability of validating, evolving, and maintaining this knowledge.

A number of popular ontology editing tools were experimented with while conducting this research in order to gain first hand knowledge of their functionality. Table 2 is a list of the ontology editing tools that were tried by the researcher:

Developer	Product	Availability	Language Support
FZI – AIFB http://kaon.semanticWeb.org/frontpage	KAON 1.2.7	Open source	KAON RDF(S)
IMG (University of Manchester) http://oiled.man.ac.uk/index.shtml	OilEd 3.5	Open source	RDF(S) OIL DAML+OIL OWL SHIQ
Ontoprise http://www.ontoprise.de/content/e3/e43/index_eng.html	Ontostudio 1.4	Freeware Licenses	RDF(S) OWL F-Logic OXML
SMI (Stanford University) http://protege.stanford.edu/	Protégé 3.2	Open source	XML RDF(S) XML Schema OWL
KMI (Open University) http://kmi.open.ac.uk/projects/Webonto/	WebOnto 2.3	Free access	OCML RDF(S)
Mindswap http://www.mindswap.org/2004/SWOOP/	Swoop 2.3	Open Source	RDF(S) OWL

Table 2: Ontology development tools.

Protégé 3.2, which now supports OWL, is one of the oldest and most widely used ontology editors available today. It allows the user to define and edit ontology classes, properties, relationships and instances using a tree structure. Ontologies can be exported into a variety of formats including RDF(S), and XML Schema. Protégé 3.2 was the most user friendly and functionally superior tool tried. This assessment was based on the fact that the Protégé platform supports two main ways of modelling ontologies via the Protégé-Frames and the Protégé-OWL editor, and the fact that Protégé has many useful plugins. A visualisation tab, for instance, called OWL Viz allows ontologies to be viewed in graph form and exported to a JPEG file. The application also includes a SPARQL query tab.

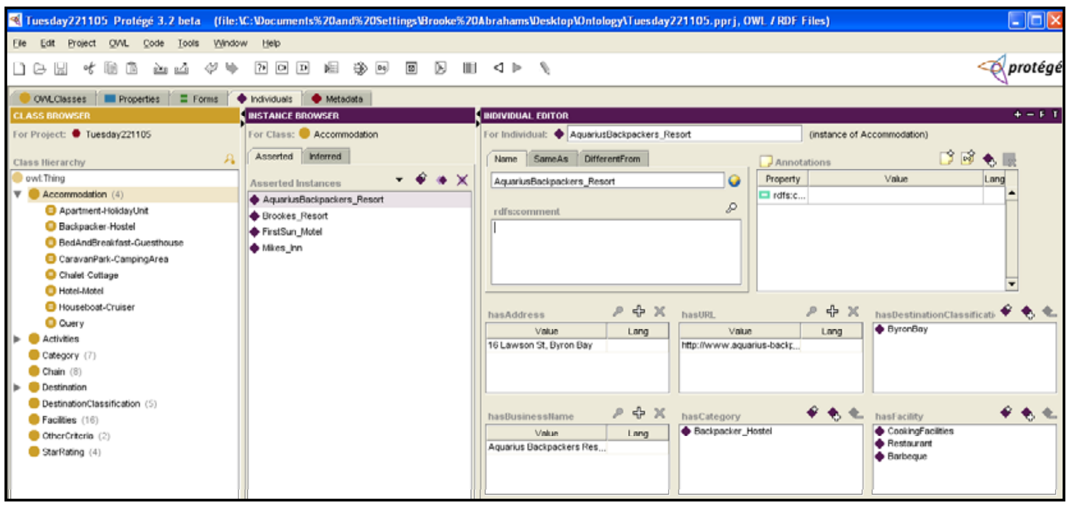


Figure 7: Protégé ontology editor.

2.2.6 Semantic Search

Semantic search is one of the key topics of the literature review. This sub-section discusses RDF query languages and the capabilities of semantically enabled search engines.

2.2.6.1 RDF Query Languages

Work on RDF query languages has been progressing for a number of years. Several different approaches have been tried, ranging from familiar looking SQL-style syntaxes, such as RDQL (Seaborne 2004) and Squish (Miller 2001), through to path-based languages like Versa (Ogbuji 2005) and RQL[34]. The SPARQL query language (Prud'hommeaux & Seaborne 2005) is (as of 6th April 2006) a W3C candidate recommendation and protocol for querying RDF. Furche (2004), who conducted a comprehensive survey of existing Semantic Web query languages, states that the challenge of serializing RDF graphs and the dissatisfaction of the Semantic Web community with RDF/XML has brought forward numerous proposals for alternate serialization formats. Furche (op. cit) found that after early attempts to simplify RDF/XML failed to gain support, the idea of directly mapping RDF nodes and edges to XML elements appears to have been abandoned in favour of a more triple-centric view of RDF graphs. Figure 8 is an example SPARQL query presented by McCarthy (2005). The query searches an RDF graph for the 'URL' of a person called 'Jon Foobar'.

```
PREFIX foaf:
<http://xmlns.com/foaf/0.1/>

SELECT ?url

FROM         <bloggers.rdf>

WHERE  {

    ?contributor foaf:name "Jon Foobar"
.

    ?contributor foaf:Weblog ?url .

}
```

Figure 8: SPARQL query example.

The first line of the query simply defines a PREFIX for the FOAF namespace, so that it doesn't have to be typed in full each time it is referenced. The SELECT clause specifies

[34] http://139.91.183.30:9090/RDF/publications/www2002/www2002.html

what the query should return -- in this case, a variable named URL. SPARQL variables are prefixed with either ? or $ -- the two are interchangeable, but McCarthy (op. cit) sticks to ? in the example. FROM is an optional clause that provides the URI of the dataset to use. Here, it is pointing to a local file, but it could also indicate the URL of a graph somewhere on the Web. Finally, the WHERE clause consists of a series of triple patterns, expressed using Turtle-based syntax[35]. These triples together comprise what is known as a *graph pattern.* The query attempts to match the triples of the graph pattern to the model. Each matching binding of the graph pattern's variables to the model's nodes becomes a *query solution,* and the values of the variables named in the SELECT clause become part of the query results. In the example, the first triple in the WHERE clause's graph pattern matches a node with a foaf:name property of "Jon Foobar," and binds it to the variable named contributor. In the bloggers.rdf model[36], contributor will match the foaf:Agent blank-node at the top of the graph. The graph pattern's second triple matches the object of the contributor's foaf:Weblog property. This is bound to the URL variable, forming a query solution.

It is worth mentioning that the query languages mentioned above focus only on a single format (in this case RDF). Berger et al. (2005) explain that the integration of data from different sources and in different formats becomes a daunting task that requires knowledge of several query languages, as well as overcoming the impedance mismatch between the query paradigms in the different languages. For instance, bibliography management applications already access (in varying combinations) book data from Amazon, Barnes & Noble, and other vendors, citation data from CiteSeer, PubMed, ACM's digital library, etc., as well as topic and researcher classifications in RDF format by crawling to and from syndication sites extracting keywords, abstracts, or tables of contents from DocBook representations of articles. Berger et al. (op. cit) argue that for such applications, Web query languages need to be more versatile, i.e., to be able to access data in different Web representation formats. They introduce a new query language called Xcerpt[37], which provides versatile access to data in different Web formats within the same query. Xcerpt is being further developed and refined at the

[35] http://www-128.ibm.com/developerworks/library/j-sparql/

[36] http://www-128.ibm.com/developerworks/xml/library/j-sparql/

[37] http://www.xcerpt.org/about/intro/

University of Munich and as part of the activities of the working group on "Reasoning-aware Querying" in the EU Network of Excellence REWERSE ("Reasoning on the Web with Rules and Semantics")[38].

2.2.6.2 Inference and Reasoning

Inference, one of the most important features of the Semantic Web, is to derive new knowledge from existing knowledge based on a generic rule. Reasoners are a type of application capable of processing the knowledge available in the Semantic Web by controlling overall execution of generic rules. Reasoners can be employed to check cardinality constraints, class membership, and create an inferred ontology model. One of the key features of ontologies that are described using OWL-DL is that they can be processed by description logic reasoners like Racer Pro[39], Pellet[40], Fact[41], and Jess[42]. Horridge (2004) demonstrates an example of inference using an OWL-DL ontology for a Pizza domain. Figure 9 shows a class called *CheesyPizza* which has asserted necessary and sufficient OWL class restrictions that specify that it is a type of class *Pizza* and has a relationship called *hasTopping* with a value of *CheeseTopping*.

Figure 9 : OWL class restrictions (Horridge 2004).

The asserted ontology model represented in Figure 10 shows that the classes *MargaritaPizza*, *AmercicanHotPizza*, *AmericanPizza*, and *SohoPizza* are all subclasses of the *NamePizza* class. There are, however, no subclasses of the *CheesyPizza* class.

---

38 http://rewerse.net/I4/

39 http://www.franz.com/products/racer/

40 http://www.mindswap.org/2003/pellet/

41 http://www.ontoknowledge.org/tools/fact.shtml

42 http://herzberg.ca.sandia.gov/jess/

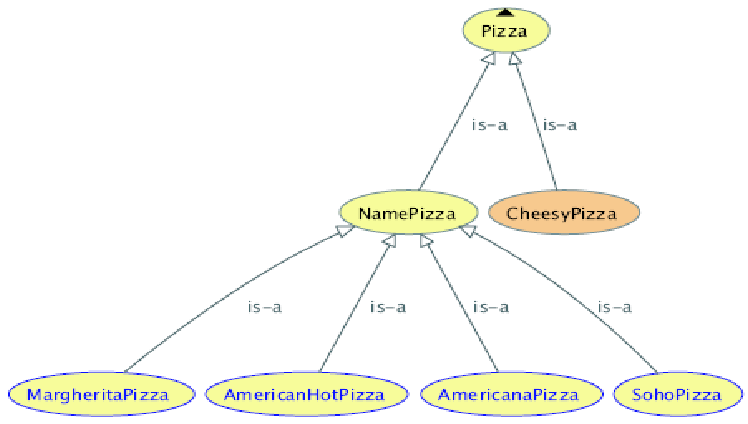


Figure 10: Static hierarchy (Horridge 2004).

Figure 11 demonstrates that with the activation a reasoner, an inferred ontology hierarchy is produced showing that the classes *AmericanHotPizza*, *SohoPizza, MargaritaPizza, AmericanaPizza* become inferred subclasses of *CheesyPizza*. The inference has occurred because of the class restrictions specified in Figure 9. This now means that all instances of the class *Pizza* that satisfy the restrictions specified for the *CheesyPizza* class, will be viewed as instances of *CheesyPizza*, and can be queried using an RDF query language in conjunction with a query application.

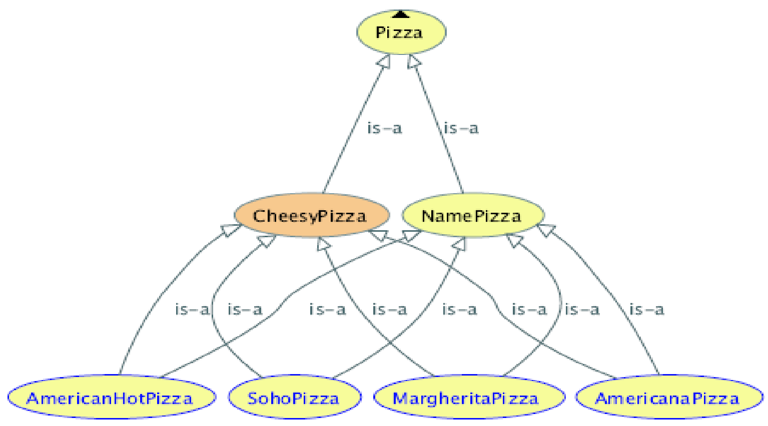


Figure 11: Inferred hierarchy (Horridge 2004).

Abrahams and Dai (2005b) demonstrate a similar example of inference, this time in the tourism domain, where class restrictions are used to infer the attractions associated with a particular resort based on the resort's star rating. In this example, a tourism customer searches a semantic accommodation portal for a 5 Star Hotel/Motel somewhere in Victoria (Australia) with a swimming pool, bar, restaurant and valet parking. Room facilities are to include pay TV and air-conditioning. The attractions hiking and surfing have also been selected in the search criteria. The customer is flexible about the precise location of the resort. Victoria (Australia) is the preferred state. The application, which is called *AcontoWeb*, has a forms-based GUI and in this instance, the query is presented to the system as illustrated in Figure 12.

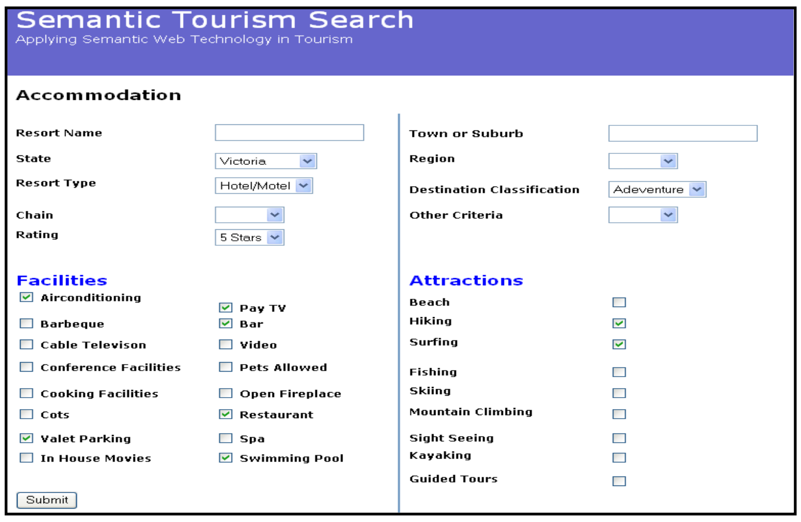


Figure 12: AcontoWeb GUI (query interface) (Abrahams & Dai 2005b).

Once the user presses submit, the query is processed by a Jena[43] supported middleware environment and a Racer reasoner. The ontology reasoning for the Figure 12 example query is shown in Figure 13.

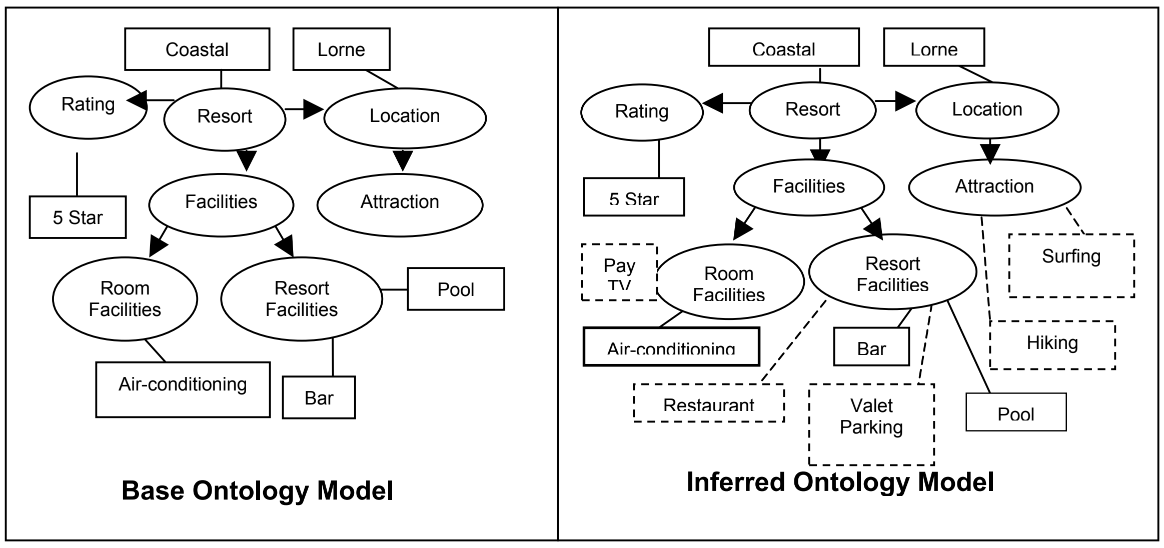


Figure 13: Ontology reasoning (Abrahams & Dai 2005b).

The query has returned a list of matching results shown in Figure 14. The results are displayed in an ordered of hierarchy of closest match to the user's request.

---

[43] http://jena.sourceforge.net/

Matching accommodation

Preferred Accommodation/Comprehensive listings

The Coastal	Hotel	Lorne
Sofitel Melbourne	Hotel	Melbourne
AAA Five Star Linden Gardens	B&B Traditional	Mount Dandenong
Amethyst Lodge	B&B Traditional	Dixons Creek
Arthurs Superb Views & Luxury Accommodation	B&B Traditional	Arthurs Seat
Aspect Villas	Holiday Unit	Halls Gap
Azimuth Country Estate	Cottage	Red Hill
Ballarat Goldfields Holiday Park	Caravan Park	Ballarat
BekSeas - Bed & Breakfast	Apartment	Warrnambool
BIG4 Melbourne Ashley Gardens	Caravan Park	Braybrook
Boroka Downs	Villa	Halls Gap
Buttercup Cottage	Cottage	Merrijig
Chateau Yering Historic House Hotel	Guest House	Yering

SEARCH AGAIN

Figure 14: Query results (Abrahams & Dai 2005b).

The Coastal hotel is returned as the closest match based on the inferred ontology model and a similarity measure. The Coastal does not explicitely state on their Web site that they have a restaurant, pay TV or valet parking, or that hiking and surfing are associated with the resort. These facts have been inferred.

2.2.6.3 Web Search Agents and Multi-agent Systems

The use of RDF and OWL tags in Web pages provides the opportunity for more advanced searching of Web content through the development of semantically enabled search engines. Several major companies including Microsoft have recently been investing in the development of a new breed of search engines called Web search agents. Web search agents do not perform like commercial search engines which use database lookups from a knowledge base. Instead, Web search agents can crawl the Web itself searching for RDF and OWL documents, while at the same time providing an interface to the user. They can be programmed to facilitate user queries including determining and executing a query plan, and can be designed to initiate middleware environment tasks. The applications are typically developed in a Java programming environment because of Java's powerful server side programming capability, and the fact that most middleware applications (see sub-section 2.2.7.2) can be readily interfaced with Java. Alesso et al.(2004d) contend that Microsoft's MSNBot program[44], which performs agent/robot like functions and searches

[44] http://search.msn.com/docs/siteowner.aspx

the Web to build an index of HTML links and documents, may pose a serious threat to Google. Figure 15 shows the typical work flow functionality of a Web search agent.

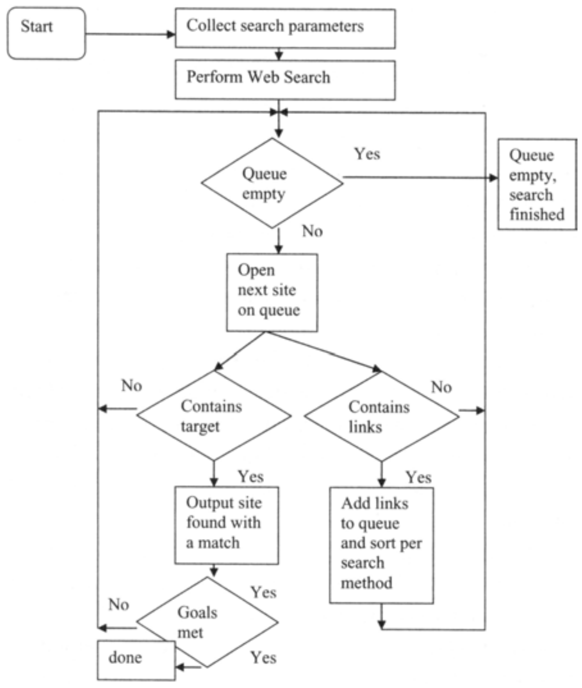


Figure 15: Web search agent basic flow (Alesso, P & Smith, C 2004d).

A multi-agent system (MAS) is a loosely coupled network of software agents that interact to solve problems that are beyond the individual capacities or knowledge of each problem solver[45]. Bloodsworth and Greenwood (2005) state that by placing Semantic Web technologies at the heart of a multi-agent system it is possible to create a system in which agent behaviour and internal representation are abstracted from coding. Each agent in the system uses this layer, in addition to instances, to form a knowledge base defining its behaviour. The ontology-layer is a mixture of domain specific and generic ontologies, which structures the behaviour of a multi-agent system. Bloodsworth and Greenwood (op. cit) believe that such a level of abstraction makes editing the behaviour of agents more convenient, requiring only the altering of domain specific ontologies without any major changes to the coding of the system. This ontology-centric approach encourages re-use, allowing the system to move from one problem domain to another by creating an ontology layer defining the new environment and system behaviour. These features make the future possibilities of such methods exciting.

[45] http://www.cs.cmu.edu/~softagents/multi.html

Comprehensive designs for a Semantic Web based multi-agent system were presented in Abrahams and Dai (2005a). In this environment individual agent behaviour is driven by intentions that are determined by problem solving logic coded into the agent. The agents interact to perform tasks such as: 1) crawling the Internet at regular intervals to search for RDF marked up documents consistent with the domain ontology; and 2) extracting RDF content and storing it in an RDF enabled database, which forms part of a Jena supported middleware environment maintained on a Web server. The GUI is accessed remotely by an end user searching for information in the same way as a conventional search engine. User requests are passed to the Web agents who, in turn, formulate a query plan. Inference is performed on ontology schema information and instance data by the activation of a reasoner, which is a component of the middleware. SPARQL queries are formulated and processed by the agents in conjunction with Jena and results displayed to the end user via the GUI. The multi-agent system is presently under development as part of the Phoenix[46] research program. The main theme of the PHOENIX project is applications integration through EAI (Enterprise Application Integration) processes and infrastructures to support real-time service oriented enterprise tasks. The high level architecture of the Phoenix multi-agent system is presented in Figure 16.

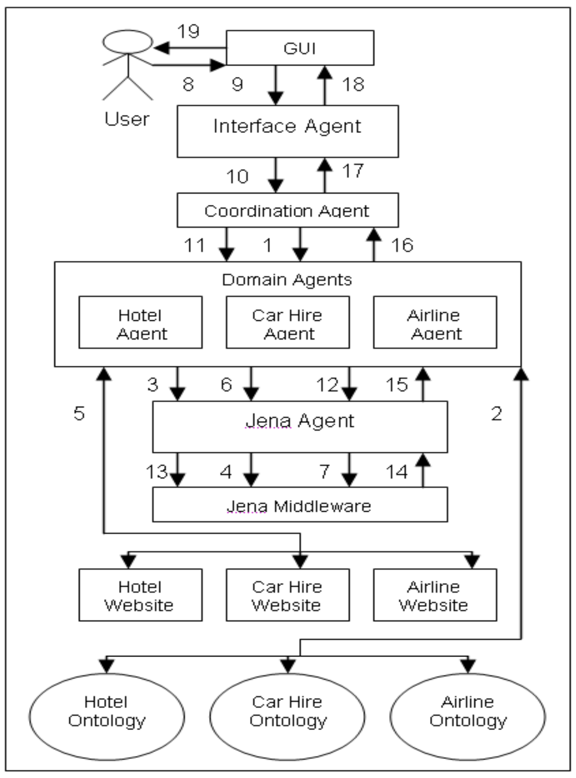


Figure 16: Multi-agent architecture (Dai & Abrahams 2005).

[46] http://www.staff.vu.edu.au/PHOENIX/phoenix/index1.htm

The numbers shown in figure 16 correspond to the following key processes as described by (Dai & Abrahams 2005):

1. Coordination agent instructs domain agents to crawl the Internet to update domain ontologies and search for RDF annotated Web sites.
2. Domain agents search for and download relevant domain ontologies from the Web.
3. The ontologies are sent by the domain agents to the Jena agent, which is responsible for interacting with the Jena middleware application.
4. Having established a connection to the Jena middleware, the Jena agent creates a Jena ontology model and saves the model using Jena's persistent storage capability linked to a backend database.
5. Domain agents crawl the Internet searching for and downloading Web pages with RDF markup containing a matching namespace to their domain specific ontology.
6. Domain agents extract the RDF annotations from the Web pages and send them to the Jena agent.
7. The Jena agent, having maintained a connection to Jena middleware, writes the extracted RDF markup into the relevant ontology model contained in the persistent storage database.
8. End user issues requests for a travel service via the GUI.
9. GUI accepts the user request, converts the request to an XML form and sends it to the interface agent.
10. Interface agent receives the user request and transforms the task descriptions into technical specifications which are then are passed to the Coordination agent.
11. Coordination agent divides tasks into subtasks, formulates a plan and allocates subtasks to domain agents.
12. Domain agents formulate a number of possible solutions to their specific tasks and convert the solutions into query specifications. The query specifications are each given a ranking based on best match to user request. Specifications are then sent to Jena agent.
13. Jena agent converts the query specifications into SPARQL query language format using parameters and predefined query templates. Jena agent also invokes the Racer reasoner to classify the ontology models which now contain both schema and instance data for each domain. Jena agent then initiates SPARQL queries over the inferred ontology model.
14. Jena agent retrieves the query results from the reasoner.

15. Results are sent back to the domain agents.
16. Domain agents sort the query results into their ordered hierarchy and send them to coordination agent.
17. Coordination agent confirms that a solution has been found. It determines how results are to be displayed (order and number of hits etc.) and sends the requirements and results to the interface agent.
18. Interface agent converts the results to HTML, formulates a page layout and passes results to the GUI.
19. GUI displays the results to the end user.

2.2.7 Semantic Web Application Development

This sub-section provides an overview of client-side (Webpage annotation), and server-side techniques for Semantic Web application development.

2.2.7.1 Client-Side Development (Webpage Annotation)

The first stage in the information item life cycle in a Semantic Web environment is the creation of information. An information item is generally created as a conceptual instance of an ontology class using an ontology based annotator such as Cohse[47], OntoMat[48] or Shoe Knowledge Annotator[49]. These applications allow the information provider to create RDF markups then associate the markup to a Webpage. To date there is no standard method for associating RDF with HTML. Palmer (2002) describes a number of possible annotation methods, including:

- Imbedding RDF in HTML – this involves placing the RDF markup somewhere that they can be readily extracted and not displayed by the browser. This may be done using the head tags or comment tags of the HTML document.
- Linking to external document – this is possibly the purest solution from an architectural point of view. The RDF annotations are stored on a separate RDF document somewhere on the Web. The original HTML document then contains a

[47] http://cohse.semanticWeb.org/software.html

[48] http://annotation.semanticWeb.org/ontomat/index.html

[49] http://annotation.semanticWeb.org/Members/lago/AnnotationTool.2003-08-25.5632

`<link>` to the annotation. This method has been subject to criticism since maintaining the metadata externally to the RDF is seen as inconvenient.

- Embed RDF as XHTML - this approach basically involves hacking up a small DTD (document type definition) using XHTML Modularization for a variant of XHTML, putting it on the Web, and then referencing it from your document. The main drawback is that the DTDs are large and relatively complex; this is not a viable approach for typical HTML authors.

Alternatively, Handschuh et al. (2003) propose an annotation framework where Web pages are generated from a database and the database owner cooperatively participates in the Semantic Web. In order to create metadata, the framework combines the presentation layer with the data description layer — in contrast to "conventional" annotation, which remains at the presentation layer. Therefore, the framework is referred to as deep annotation. Handschuh et al. (op. cit) argue that deep annotation should be considered particularly valid because; 1) Web pages generated from databases outnumber static Web pages; 2) annotation of Web pages may be a very intuitive way to create semantic data from a database and; 3) data from databases should not be materialized as RDF files, it should remain where it can be handled most efficiently— in its databases.

According to Gomes-Perez et al. (2004d), the most common approach to annotating Web documents is to embed the markup in the head or comment tags of an HTML file (see Figure 17) so that it can later be extracted by a Web crawler. This approach is used in the Cream (Handschuh, Staab & Maedche 2001) and AcontoWeb (Abrahams & Dai 2005b) projects.

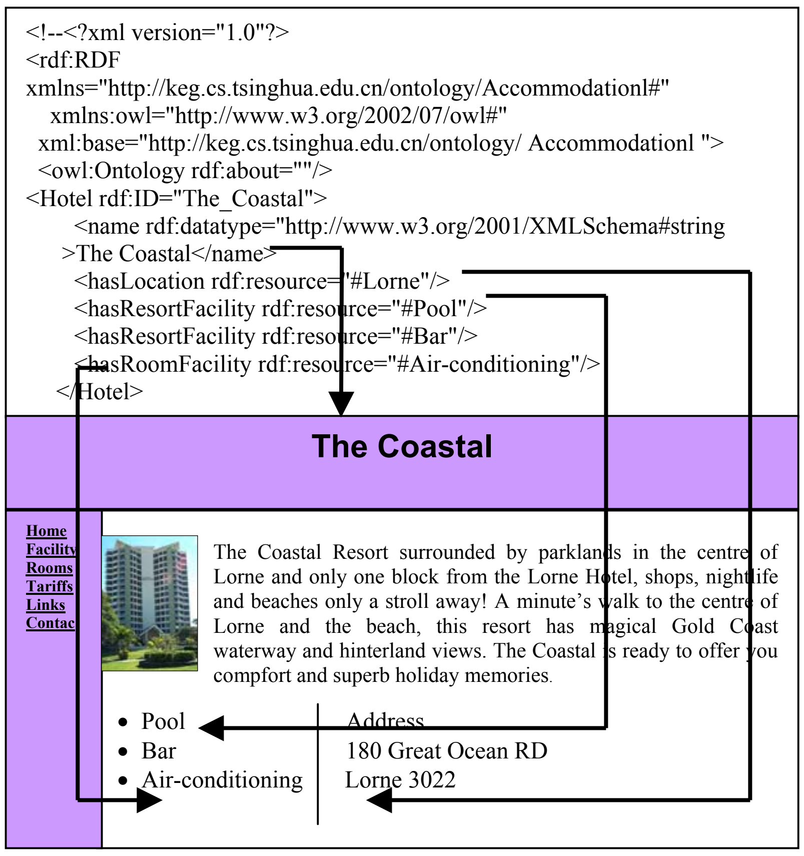


Figure 17: Annotated Webpage (Abrahams & Dai 2005b).

2.2.7.2 Server-Side Development

Sophisticated Semantic Web applications typically comprise more than one software module. Instead of coming up with proprietary solutions, developers should be able to rely on a generic infrastructure for application development in this context (Oberle et al. 2005, p. 1). Semantic middleware applications facilitate database backed RDF storage, retrieval, triple statement processing, inference via a reasoner, and query processing. Developers can access modules that perform the above tasks by interfacing with a middleware environment through an Application Programming Interface (API). There are many such middleware environments available today to assist Semantic Web application developers. An evaluation of the Sesame, RDF Suite, and Jena middleware environments was done by Oberle et al. (op. cit), who found that:

- Sesame[50] is a scalable, modular architecture for persistent storage and querying of RDF and RDF Schema. It supports two query languages (RQL and SeRQL), and can

[50] Sesame 1.2.4 available for download at: http://www.openrdf.org/

use main memory or PostgreSQL, MySQL and Oracle databases for storage. Oberle et al. (op. cit) note that the Sesame system has been successfully deployed as a proxy component for RDF support in KAON SERVER.

- RDFSuite[51] is a suite of tools for RDF management provided by the ICS-Forth institute, Greece. Among those tools is an RDF Schema specific database (RSSDB) that allows querying RDF using the RQL query language. The implementation of the system exploits the PostgreSQL object-relational DBMS. It uses a storage scheme that has been optimized for querying instances of RDFS-based ontologies. The database content itself can only be updated in a batch manner (dropping a database and uploading a file). Oberle et al. (op. cit) explain that, hence, it cannot cope with transactional updates (such as KAON's RDF Server).
- Jena[52] which was developed by Hewlett-Packard Research, UK, is a collection of Semantic Web tools including a persistent storage component, an RDF query language processor (SPARQL) and a DAML+OIL API. Oberle et al. (op. cit) explain that for persistent storage, the Berkley DB embedded database or any JDBC-compliant database may be used. Jena abstracts from storage in a similar way to the KAON APIs. However, transactional updating facilities have not been provided so far.

Table 3 contains a list of some popular middleware environments available today:

Developer	Product	Category
Administrator http://www.aidmistotor.nl/	Sesame 1.2.4	RDF(S) storage and retrieval, ontology based information presentation
FZI – AIFB http://kaon.semanticWeb.org/frontpage	KAON 1.2.7	Inference engine, knowledge management and tools
HP Labs http://jena.sourceforge.net/	Jena 2.3	Inference engine, knowledge management and tools
Intellidimension http://www.intellidimension.com/	RDF Gateway 2.2.3	RDF data management system
Kowari http://www.kowari.org/	Kowari Metastore 1.1	Metadata analysis and knowledge discovery, RDF storage
Ontoprise http://www/ontoprise.de/	Ontobroker 4.3	Inference middleware

Table 3: Semantic middleware environments.

51 RDFSuite available for download at: http://athena.ics.forth.gr:9090/RDF/

52 Jena version 2.3 available at: http://jena.sourceforge.net/

Ford (2004) contends that for some time the leading framework has been Jena[53]. Jena provides a programmatic environment for RDF, RDFS and OWL, including a rule-based inference engine. Jena is open source, and resulted from research conducted within the HP Labs Semantic Web Program[54]. The Jena Framework includes:

- A RDF API
- Reading and writing RDF in RDF/XML, N3 and N-Triples
- An OWL API
- In-memory and persistent storage
- SPARQL – a query language for RDF

2.2.7.3 Tools for Creating Semantic Portals

The task of building semantic portals can be made somewhat easier by using certain tools that provide a generic framework to assist with key development processes. SEAL (SEmantic portAL) (Stojanovic et al., 2001), is a system that exploits semantics for providing and accessing information at a portal as well as constructing and maintaining the portal. The SEAL architecture integrates a number of components that are also used in other applications (such as Ontobroker) and, more specifically, it contains navigation and query modules. The SEAL semantic modules include a large diversity of intelligent means for performing semantic ranking of concepts for querying and accessing Websites by crawling. The core modules, presented in Figure 18, have been extensively described in Stojanovic, et al. (op. cit).

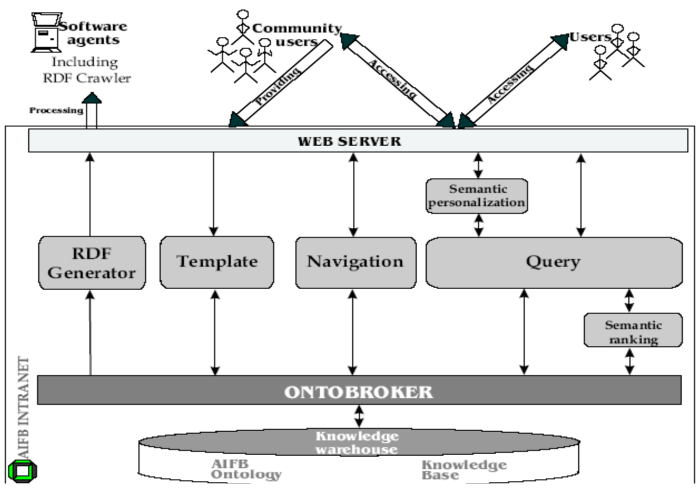


Figure 18: SEAL architecture (Stojanovic et al., 2001).

---

53 http://jena.sourceforge.net/

54 http://www.hpl.hp.com/semWeb/

Another tool that assists with creating semantic portals is Ontoviews (M¨akel¨a et al. 2004). Ontoviews provides developers with two important services; 1) a search engine based on the semantics of content; and 2) dynamic linking between pages based on semantic relations contained in the underlying knowledge base. The Ontoviews architecture consists of three main components:

1. Prolog-based logic server (Ontodella) – provides the system with reasoning services such as category generation and semantic recommendations.
2. Java-based multi-facet search engine (Ontogator) - defines and implements an RDF based query interface that separates view based search logic from the user interface. The interface is defined as an OWL ontology and can be used to query for category hierarchies of the ontology. It also facilitates keyword based searches.
3. User interface (OntoViews-C) – binds the previous two components together and is responsible for the user interfaces and interaction.

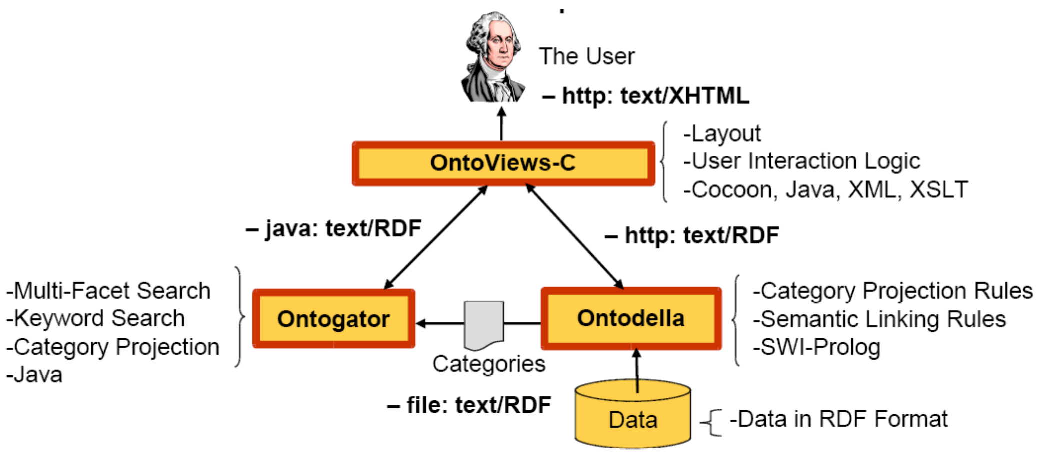


Figure 19: Ontoviews architecture (M¨akel¨a et al. 2004).

The Ontoviews search engine presents the end user with concepts for navigation in a hierarchical structure. The concepts, known as categories, are linked via semantic relations contained in the individual developer's ontology. Figure 20 shows a sample query from the Museum of Finland semantic portal[55] which was built using Ontoviews. With Museum of Finland, the content consists of collections of cultural artefacts and historical sites consolidated from several heterogeneous Finnish museum databases, annotated in RDF format using seven different ontologies. In the Figure 20 example, a search for 'esp' matches the category Spain ("Espanja" in Finish), and a list of

[55] http://www.museosuomi.fi/

semantically related categories are then displayed as hyperlinks. Searches may also be performed by navigating hyperlinks alone without using keywords.

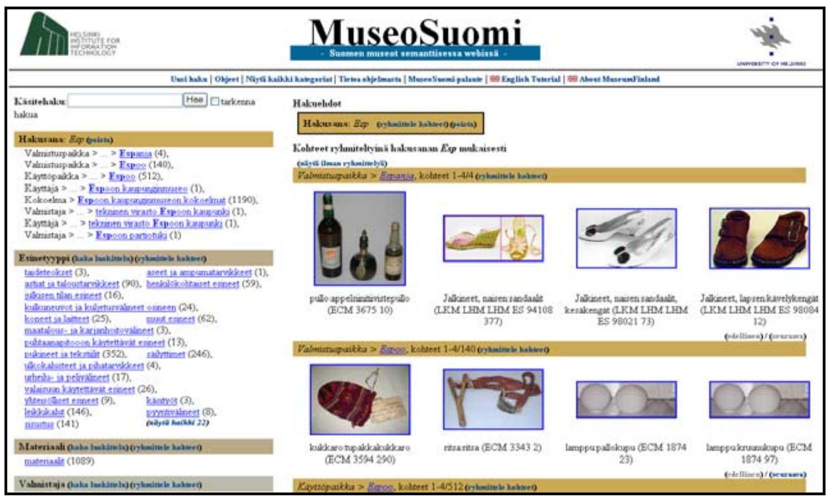


Figure 20: Museum of Finland sample query (M¨akel¨a et al. 2004).

A developer can use Ontoviews to create a semantic portal by setting up the components on a server, then adapting the system to their own data. This adoption requires a number of configuration steps. Rules describing how categories are generated and items connected to them for the view based search must first be created. The next step is to create rules describing how links are generated for the recommendations. The last step involves changing the layout of visual templates to suit the developer's needs. Ontoviews can greatly assist with creation of semantic portals by facilitating some of the key requirements of such systems. The concept based multi-facet search engine exploits the semantic relations in the underlying knowledge base providing the end user with a classification tree view containing semantic links. Ontoviews offers different user interfaces, functionality for different devices and adapts to a wide variety of semantic data.

2.2.8 Ontology Schema Integration

As mentioned in previous sections, the existence of Semantic Web standards enables information on the World Wide Web to be represented in a uniform way. This uniformity makes it easier to automatically process information in a homogenous environment, as well as information from other sources via Ontology merging and mapping techniques, thus facilitating federated data source queries. Ontology merging and mapping are defined by Noy & Musen (2002) as:

- Ontology merging - The process of generating a unique ontology from the original sources.
- Ontology mapping - Establishing different kinds of mappings (or links) between two ontologies. This sub-section focuses on ontology merging techniques.

2.2.8.1 Schema Integration Issues

Struckenschmidt & Harmelen (2005b) explain that even in almost completely homogenous environments such as relational databases, the exchange of information is a problem. This is because heterogeneity in the way information is structured and interpreted leads to conflicts when information from different sources makes it difficult to combine the information. Various attempts have been made to characterize the types of data conflicts that may occur. Dell'Erbra et al. (2005) identify two types of heterogeneity:

1. Semantic clashes: These address different interpretation or meaning of concepts. They include naming conventions as well as structural differences in the ontology.
2. Representational clashes: These relate to different markup syntaxes used, e.g. XML, RDF(S), OWL.

Wache (2003) provides a very comprehensive classification of data conflicts, categorized as either:

- Structural conflicts - the fact that the same objects and facts in the world can be described in different ways using structures provided by RDF, or
- Semantic conflicts – these occur due to the use of different encodings and conflicts due to a different conceptualization of the domain.

2.2.8.2 Schema Integration Process

Jakoniene (2003) suggests the following solution to the types of heterogeneity described above:

- The interrogation of ontologies to find places where they overlap.
- Relate concepts that are semantically close via equivalence and subsumption relations (aligning).
- Check the consistency, coherency and non-redundancy of the result.

Figure 21 shows an example two heterogonous ontology models representing library information.

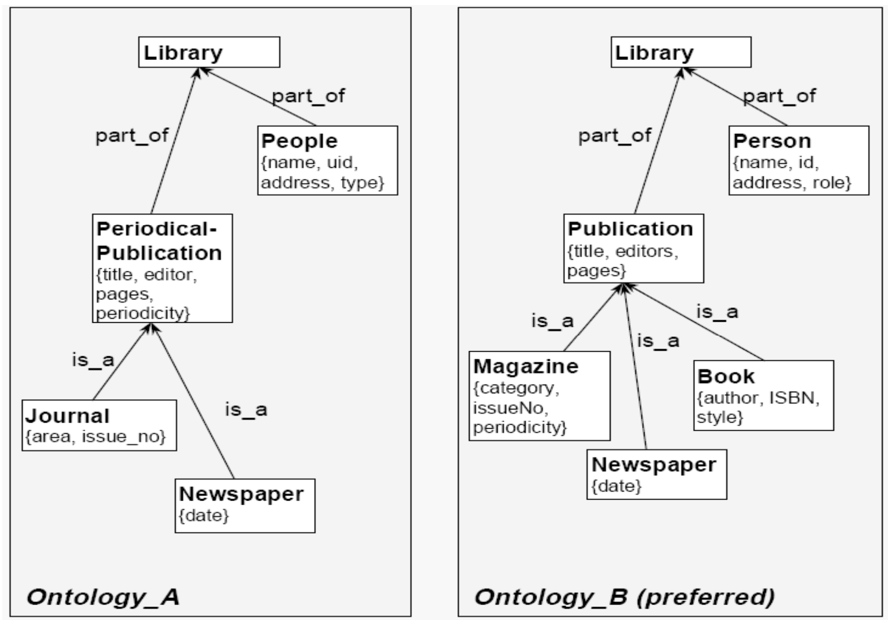


Figure 21: Ontologies to be merged (Jakoniene 2003).

Using Jakoniene's (op. cit) method, the ontologies in Figure 21 may now be merged as shown in Figure 22.

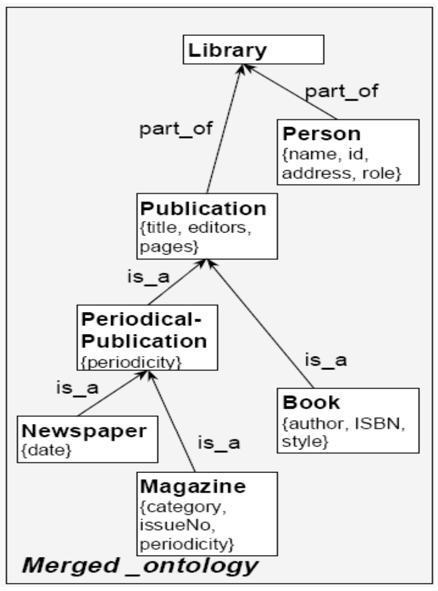


Figure 22: Merged ontology (Jakoniene 2003).

McGrath and Abrahams (2006a) demonstrate that integration is not always all that simple. They illustrate, as shown in Figure 23, a case of where information needs to be exchanged between two tourism and hospitality portals, focusing on hotels and, more specifically, on the relationships between employees and the departments in which they work. In this instance, the constraints *C11* and *C21* are contradictory: that is, in Ontology 1, an employee must be associated with one and only one department but, in Ontology 2, each employee can work in a number of departments but must play a specific *role* in each one[56].

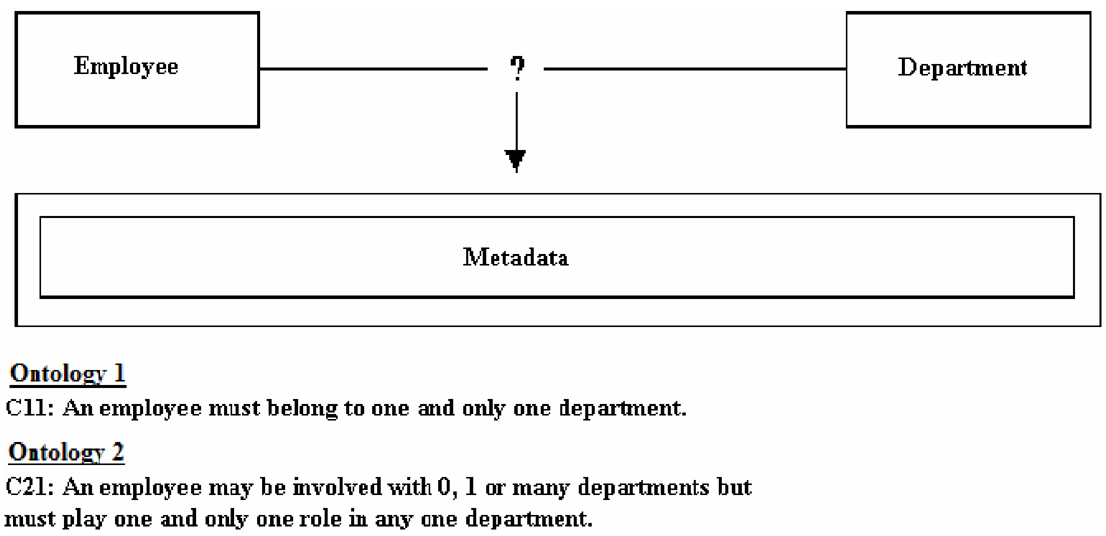


Figure 23: Example of a semantic conflict (McGrath & Abrahams, 2006a).

Following McGrath and Abrahams' (op. cit) example, assume now that employee instance data is required to be transferred between repositories corresponding to the two ontologies. First, where the direction is Ontology 2 to Ontology 1 (Case 1), all data associated with employees working in more than one department will be rejected (because constraint *C11* is breached). Alternatively, where the direction is Ontology 1 to Ontology 2 (Case 2), *all* data will be rejected (because there are no roles associated with any employee-department relationship and, hence, *C21* is breached).

To reconcile data here, the following three approaches might be adopted: 1) declare either ontology as the 'standard'[57] and amend code in all affected systems built around the (now) non-standard ontology; 2) add intelligence to the metadata (thereby creating a new meta-ontology) to perform any necessary reconciliation; or 3) establish a new standard

[56] For example, an employee could be a wine waiter with Food Services and a shift supervisor with Frontdesk Operations.

[57] Probably Ontology 2 – because it is richer and assumes Ontology 1.

ontology (as in 2) above) and embed the required intelligence in rules that map data from source systems to (and from) a form consistent with the new standard. A major benefit of option 3 is that source systems do not have to be touched and, essentially, this is the Harmo-TEN approach[58]. Informally stated, mapping rules developed for this particular example (assuming something close to Ontology 2 is adopted as the new standard) might be:

Case 1

if *the source data is defined by Ontology 2*
and *employee Ei works in the set of departments {D1,----,Dn}*
and *the principal_department of Ei is Dj*
and *the role of Ei in Dj is Rij*
then *Ei belongs to Dj with role Rij.*

Case 2

if *the source data is defined by Ontology 1*
and *employee Ei works in department Dj*
and *Rij is declared as the role of Ei in Dj*
then *Ei belongs to Dj with role Rij.*

In each of these cases, some user intervention is required: specifically, with Case 1, principal departments must be selected and, with Case 2, involvement roles must be declared. However, within limits, this approach is more efficient and less expensive than alternatives. In particular, each organization connected to the semantic portal is free to change its systems independently of other participating organizations. Where this occurs, any necessary changes are restricted to interfaces to the portal (i.e. the mapping rules) (McGrath & Abrahams, 2006a, p. 10).

2.2.9 Semantic Web Services

Web services add a new level of functionality to the current Web, transforming the Web from a distributed source of information to a distributed source of functionality. They provide a standard means of interoperating between different software applications,

[58] As detailed at: ENTER Workshop 2, "Harmo-TEN: A Cost Effective Solution for Information Exchange", ENTER 2006, Lausanne, Switzerland, 18-20 January, 2006.

running on a variety of platforms and/or frameworks (Lausen et al. 2003, p. 6). The W3C Web services activity statement[59] says that Web services are characterized by their great interoperability and extensibility, as well as their machine-processable descriptions thanks to the use of XML. They can combine in a loosely coupled way in order to achieve complex operations. Programs providing simple services can interact with each other in order to deliver sophisticated added-value services. Current Web service technologies, however, which are based on protocols UDDI, WSDL, and SOAP, offer limited service automation support. Alesso, and Smith (2004c) report that recent industrial efforts have focused primarily on Web service discovery and aspects of service execution through initiatives such as the Universal Description, Discovery, and Integration (UDDI) standard service registry and ebXML, an initiative of the United Nations and OASIS (Organization for the Advancement of Structured Information Standards) to standardize a framework for trading partner interchange.

With the new generation of Web markup languages like OWL and RDF, a number of initiatives have emerged with the aim of creating Semantic Web services. Burstein et al. (2005, p. 2) describe Semantic Web services as Web services in which Semantic Web ontologies ascribe meanings to published service descriptions, so that software systems representing prospective service clients can interpret and invoke them. Enriching Web services with semantic information allows automatic location, composition, innovation, and interoperation of services (Lausen et al. 2003). OWL-S is an OWL-based Web service ontology developed by the W3C, which supplies Web service providers with a core set of markup language constructs for describing the properties and capabilities of their Web services in unambiguous, computer-interpretable form. OWL-S has been designed to facilitate: 1) automatic Web service discovery; 2) automatic Web service innovation; 3) automatic Web service interpretation; and 4) automatic Web service execution monitoring.

Enriching Web services with semantic information allows automatic location, composition, invocation, and interoperation of services. Significant work has already been done in this decade on Semantic Web services, and a large body of relevant work exists from earlier decades in fields such as knowledge representation, planning, agent-

[59] http://www.w3.org/2002/ws/Activity

based systems, databases, programming languages, and software engineering. Nevertheless, many difficult research challenges remain, and much work is needed to adapt relevant existing technologies to the context of Web services and the Semantic Web, and to prepare the more mature languages, capabilities and architectures for widespread deployment. These challenges are discussed in more detail in sub-section 2.2.10.10.

2.2.10 Challenges and Future Trends

In spite of the big advantages that the Semantic Web promises, its success or failure will, as with the World Wide Web be determined to a large extent by easy access to, and availability of high-quality and diverse content. It is widely acknowledged in the AI literature that there are still many challenges to face if this is to happen. The following is a list of some the most widely recognized issues along with future trends that could possibly provide solutions:

2.2.10.1 Availability of Content

For the Semantic Web to succeed there needs to be a critical mass of metadata enriched documents; currently there is little available. The reality today is that most Web pages are rendered in HTML and this is likely to remain the case for some time. Benjamins et al. (2004) believe that existing Web content should be migrated to Semantic Web content, including static HTML pages, dynamic Web pages, and multimedia and Web services. From this viewpoint, annotation tools are critical to the success of the Semantic Web. Alesso and Smith (2004b) point out two limitations of existing annotation tools: 1) most of them annotate static pages only, and 2) many of them focus on creating new content. This leads to a situation where dynamic Web page content is not considered, and existing content may be excluded from the Semantic Web. Manual annotation therefore needs to be augmented with other means of creating metadata such as text mining and semi-automated annotation as described by Priebe et al. (2005), and Latent Semantic Indexing (LSI)[60].

[60] http://www.cs.utk.edu/~lsi/

LSI shows a lot of promise. It is a method that organizes existing information into a semantic structure that takes advantage of implicit higher order associations of words with text objects. The resulting structure from applying LSI reflects the major associative patterns in the data. This permits retrieval based on the "latent" semantic content of the existing Web documents, rather than just on keyword matches. LSI offers an application method that can be implemented immediately with existing Web documentation.

Extensive work has also been done on annotating dynamic Web content by researchers such as Song et al. (2004) and Stojanovic et al. (2002). Approaches tried so far include:

- Extracting the dynamic content from its source, annotating and storing it. The problem with this is the almost finite amount of static pages that can be generated from a dynamic site, including continuous updates, creations, and removals of pages when data changes in databases.
- Leave the content in the database and annotate the query that retrieves the concerned content. This option is less space-consuming and provides consistency in the annotations with respect to the underlying sources of information, since the content is dynamically annotated when retrieved (Alesso, P & Smith, C 2004b, p. 411).

2.2.10.2 Ontology Development and Availability

Benjamins et al. (2004) state that a major challenge for implementing the Semantic Web is creating common widely used ontologies on the provision of adequate infrastructure for ontology development, change management, and mapping. Due to the immaturity of the Semantic Web, there is a need to improve methodological and technological tasks for most activities associated with the ontology development process. Chebotko et al.(2004) supports this view and adds that because communities develop ontologies in their domains, with many experts in the same domain each having their own perspective, there is also a social challenge created. Chebotko et al. (op. cit) believe that it is essential for a domain to have a collaborative ontology development environment that will enable version control, proposal and release control, and coordination and collaboration support. The development of such an environment is a major technical challenge. Accessing existing ontologies is now becoming a little easier with the emergence of ontology library systems such as: DAML ontology library[61], Ontolingua ontology library[62], Protégé ontology library[63], SHOE ontology library[64], WebODE[65], and WebOnto ontology library[66].

To this stage, there are also no formal guidelines or techniques on how to model ontologies. A number of methods have been proposed but all have shortcomings. Gruber (1993b) proposed modelling ontologies using frames and first order logic. Rumbaugh et al. (1998) suggests that Unified Modelling Language (UML) might be a suitable technique, and Gomes-Perez et al. (2004c) demonstrate an approach which involves extending the Entity Relationship (ER) diagram. The problem with these modelling approaches is that they limit the kind of knowledge that can be modelled and implemented by the newer bread of highly expressive Web markup languages. With this in mind, Gomes-Perez et al. (op. cit) express the view that AI-based approaches such as Ontolingua, Loom, OCML, FLogic etc., are better candidates for representing ontologies than non-AI approaches such as UML and ER diagrams.

2.2.10.3 Ontology Versioning Issues

Ontology versioning support is necessary because changes to ontologies may cause incompatibilities, which means that a changed ontology cannot simply be used instead of the unchanged version (Klein & Fensel 2001). Because there are dependencies between data sources, applications and ontologies, changes to the latter will have far-reaching side effects. Qin (2005) explains that changes to an ontology may invalidate its data instances and dependent ontologies, thus detecting changes to data objects has become essential for data warehousing, knowledge archival applications, and search engines. Qin (op. cit) adds that another problem is that semantics can lead to data instances being inferred from changes to others, and that this may subsequently pose a threat to confidentiality (since

---

[61] http://www.daml.org/ontologies/

[62] http://www-ksl.stanford.edu/knowledge-sharing/ontologies/index.html

[63] http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary

[64] http://www.cs.umd.edu/projects/plus/SHOE/onts/index.html

[65] http://Webode.dia.fi.upm.es/Webode/login.html

[66] http://Webonto.open.ac.uk/

ontologies may enable the inference of sensitive information from unclassified information). It is therefore important to take into account inference relationships and carefully assign access permissions to eliminate undesired inference.

After examining the effects on compatibility in a number of example scenarios for ontology versioning, Klein and Fensel (2001) sketched some elements for a versioning framework for ontologies. These elements mainly focus on identification of, and referring to specific versions of ontologies. Klein and Fensel (op. cit) attempt to achieve "maximal use" of the available knowledge. This implies that it is not sufficient to find out whether a specific interpretation of an ontology on data is invalid, there is also a need to derive as much valid information as possible

2.2.10.4 Scalability of Systems

Alesso and Smith (2004b) state that once the Semantic Web content becomes widely available, the resultant complexity of related facts will require management in a scaleable manner including organizing, storing, and searching content. The storage and organization of Semantic Web pages includes the use of semantic indices to group content based on topics. According to Alesso and Smith (op. cit), semantic indices may be generated dynamically using ontological information and annotated documents. Benjamins et al. (2004) also see scalability as an issue to address. Like Alesso and Smith (op. cit), they say that a significant effort must be made to organize Semantic Web content, store it and provide the necessary mechanisms to find it. All these tasks must be performed and coordinated in a scalable manner, as these solutions should be prepared for the huge growth of the Semantic Web.

2.2.10.5 Visualization of Content

The design of semantically and graphically enriched interfaces for e-commerce and information retrieval and presentation is a challenging area of practical Web development. Benjamins et al. (2004) state that the intuitive visualization of Semantic Web content will become more and more important in solving the increasing amount of information overload, as users will demand easy recognition of relevant content for their purposes. New techniques must be explored that differ from the usual hypertext structure visualization of the current Web. Geroimenko & Chen (2006) have produced perhaps the most comprehensive and advanced work on visualization techniques to date. They describe many techniques that can be used today and associated issues including: ontology based and topic map visualizations, visual interfaces for retrieving, browsing and mapping semantic information, SVG/X3D as new visualization techniques for the Semantic Web, methods used to construct high quality metadata / metadata taxonomies, interface issues related to filtering and recommending on the Web, and semantic-oriented use of existing visualization methods.

2.2.10.6 Stability of Semantic Web Languages

It is important that open standards dominate the Semantic Web. Markup languages have so far developed in a layered fashion as demonstrated in sub-section 2.2.4.7. Tool support, also needs to be considered in relation to standardization of languages. Alesso & Smith (2004b) say that tool support is essential to making a significant step forward in the construction of the Semantic Web, but the tools are partly dependant on the Semantic Web languages themselves. Therefore, integration and interoperability will always be a concern. Standardization efforts have already produced W3C recommendations for RDF Schema and OWL. Standardization efforts are continuing for the provision of rule-based support on top of these languages.

2.2.10.7 The Challenge of Ontology Mapping, Alignment and Merging

Even in one domain, it is difficult to enforce a single ontology for each data source. Instead, it can be argued that people should have the full freedom to use their proprietary ontology to annotate their data sources (Chebotko, Lu & Fotouhi 2004). Then, if they are willing, provide additional mapping to a standard (central) domain ontology to support data interoperability and queries across data sets. As was demonstrated in sub-section 2.2.8, this mapping is a challenge because there might exist heterogeneities between ontologies: syntactic, schematic and semantic. The mapping process might include not only ontology alignment to make ontologies coherent, but also ontology merging to add new terms in a central ontology. Therefore, interoperability between different information sources is an important topic with regard to the use and efficient sharing of information across different applications and domains. While many interoperability problems caused by structural and semantic differences have been solved, the notion of semantic interoperability remains to a large extent unsolved. Struckenschmidt et al. (2005b) explain that this is mainly because problems on the semantic level occur due to the inherent context dependency of information that can only be understood in the context of their original source and purpose.

2.2.10.8 The Challenge of Ontology-Based Information Retrieval

Annotated data is not useful if one cannot search through it. One promise of the Semantic Web is high precision. Search engines now should exploit available semantics and ontology reasoning to return not only precise results, but also specify meaningful relationships between them. New opportunities also require new approaches to query refinement and user interface tactics (García & Sicilia 2003). But the major challenge is searching across data sets annotated using different ontologies. As previously noted, there can be several ontologies for one domain since each domain can be modelled by several ontologies, or a domain may require the usage of several ontologies. Chebotko et al. (2004) contend that as a result of this, not only is ontology mapping required, but user query mapping may also be needed. A possible solution is to develop more versatile query languages that are able to access data in different Web representation formats. Such an approach was previously discussed in sub-section 2.2.6.1, where Berger et al. (2005) presented the Xcerpt query language, which provides versatile access to data in different Web formats within the same query.

Providing natural language query processing also remains a challenge for Semantic Web application developers. Natural language interfaces are required to provide easy and intuitive access to information sources so that users can express their information needs in their own words. The difficulty is that the development of Natural Language Processing (NLP) tools requires computationally intensive algorithms relying on large amounts of background knowledge, making the tools highly domain-dependant and virtually inapplicable to new domains or applications. Bernstein et al. (2006) tackle the issue by introducing Ginseng[67], a guided input natural language search engine for the Semantic Web. Ginseng does not use any predefined vocabulary and does not try to interpret the queries (logically or syntactically). Instead, Ginseng “only knows” the vocabulary defined by the currently considered ontologies. All ontologies are stored in a Jena inferencing model (*OWL_MEM_RULE_INF*). The vocabulary is closed and the user

[67] http://www.ifi.unizh.ch/ddis/?id=332

has to follow it. Bernstein et al. (op. cit) explain that this can limit the user's possibilities in general but ensures that every query can be answered. The vocabulary grows with every additionally loaded ontology.

Ginseng allows users to query any OWL knowledge base using a guided input natural language that strongly resembles plain English. The user enters the query into a free form entry field. When the user starts typing, the system predicts the possible completions of what the user enters (similar to completion suggestions in UNIX shells), and presents the user with a choice popup box. While the user is in the middle of a word, the popup offers suggestions on how to complete the current word. The possible choices reduce as the user continues to type. Ginseng guides the user through the set of possible queries while avoiding ungrammatical queries. When a query is complete, Ginseng translates it into SPARQL statements, executes it against the existing ontology model, and displays the generated SPARQL query and the result(s) to the user.

2.2.10.9 Change Management Issues

As well as the many technical challenges of implementing the Semantic Web, there are also important change management issues to consider. Resistance to technical change has long been recognized as a major problem in the implementation of new information systems. Bernard (1990) discusses numerous cases of underlying tensions between the control of process and the control of workers during implementation of new computer systems in business. The study illustrates to varying degrees, "management resistance to change", or a failure to accept some of the social consequences which the new technical systems seem to promote. Schlesinger (1979), in a more general study (not specific to ICT systems), found the following four primary reasons why certain people resist change:

1) **Parochial self-interest** - some people are concerned with the implication of the change for themselves and how it may affect their own interests, rather than considering the effects for the success of the business.
2) **Misunderstanding** - communication problems such as inadequate information.
3) **Low tolerance to change** - certain people are very keen on security and stability in their work.

4) **Different assessments of the situation** - some employees may disagree on the reasons for the change and on the advantages and disadvantages.

An investigation that was undertaken as part of this research about attitudes towards the adoption of new online technologies among Australian tourism operators is presented in Chapter 5.

2.2.10.10 Challenges for Implementing Semantic Web Services

There are a number of challenges to face before the Semantic Web services discussed in sub-section 2.2.9 can be widely implemented. These challenges are identified by Alesso and Smith (2004b) as:

- Integration with the Web – SOAP Web services use the HTTP infrastructure. It is not possible to hyperlink SOAP Web service via HTML links or XSLT functions.
- Extension mechanism – SOAP provides an extension mechanism via headers.
- Overall understanding of modules and layering – SOAP provides a framework within which additional features can be added via headers, but there is little agreement on the specific categories of functionality.

2.2.10.11 Application Design Issues

Research by Reynolds et al. (2004) into the development of a semantic portal of a directory of UK environmental organizations, revealed that the design of such portals throws up the following challenges, with wider implications for the design of all types of Semantic Web applications:

- **Moderation and access control -** decentralized portal design enables an interesting security model. In Reynolds' (op. cit) test implementation, the aggregator will have a record of which source URL's are deemed authoritative for a given organization. Each organization can then impose its own access and validation rules governing the update of that data. Some central administration is needed to moderate this "white list" of acceptable information sources. A Semantic Web crawler approach, which supports dynamic addition of new sources is one possibility, but does not in itself address the problem of discovering "unsuitable" material.

- **Navigation -** the rich classification of portal items is only useful if interface complexity is controlled. Current experience suggests that a faceted browse approach modelled after the Flamenco project[68] offers a good balance between expressiveness and simplicity.
- **Provenance** - the ability to mix community extensions and annotations with an organization's own data is a powerful feature of ontologies. However, it is important that users navigating sites are able to clearly separate authoritative data from third party data, and in the latter case find where it came from in order to decide how much to trust it. This raises design issues for efficient recording of provenance, trust model issues (delegation and so forth), but also user interface issues of how to make the provenance of items clear.
- **Open-ended data model** - Reynolds (op. cit) wishes to support the open-ended nature of the RDF data model so that new properties and classes (whether authoritative or third party) can be incrementally added. Reynolds (op. cit) states that the visualization engine, though, needs to adapt to such changes without requiring new rendering templates to be created at each stage.

2.3 Tourism E-Commerce and the Semantic Web

This section describes the economic significance of the tourism industry. It also discusses applications and issues relating to tourism e-commerce and the use of advanced tourism ICT applications, as well as the recent emergence of tourism related Semantic Web initiatives.

2.3.1 World Tourism Industry

Tourism is a vital industry to the economies of most countries worldwide (developed or less developed). It represents a cross-sectoral industry, including many related economic sectors such as culture, sport or agriculture, where over 30 different industrial components have been identified that serve travellers (Werthner 2003, p. 1). The components include services such as accommodation, car hire, air travel, and guided

[68] http://bailando.sims.berkeley.edu/flamenco.html

tours. World tourism's economic significance is emphasized by the World Travel & Tourism Council, whose 2006 Travel and Tourism Economic Research Report[69] States that world travel and tourism:

- is expected to generate US $6,477.2 bn of economic activity (total demand) in 2006, growing (nominal terms) to US$12,118.6 bn by 2016. Total demand is expected to grow by 4.6% in 2006 and by 4.2% per annum, in real terms, between 2007 and 2016.
- is expected to contribute 3.6% to Gross Domestic Product (GDP) in 2006 (US$1,754.5 bn), rising in nominal terms to US$2,969.4 bn (again, 3.6% of total) by 2016. The travel and tourism economic contribution (percent of total) is expected to rise from 10.3% (US$4,963.8 bn) to 10.9% (US$8,971.6 bn) in this same period.
- employment is estimated at 234,305,000 jobs in 2006, 8.7% of total employment, or 1 in every 11.5 jobs. By 2016, this should total 279,347,000 jobs, 9.0% of total employment or 1 in every 11.1 jobs. The 76,729,000 Travel and Tourism Industry jobs account for 2.8% of total employment in 2006 and are forecast to total 89,485,000 jobs or 2.9% of the total by 2016.
- is expected to generate 11.8% of total exports (US$1,646.2 bn) in 2006, growing (nominal terms) to US$3,468.4 bn (10.9% of total) in 2016.
- is estimated at US$2,844.7 bn or 9.5% of total personal consumption in year 2006. By 2016, this should reach US$4,916.3 bn or 9.8% of total consumption. World business travel is estimated at US$672.5 bn in year 2006. By 2016, this should reach US$1,190.3 bn.
- capital investment is estimated at US$1,010.7 bn or 9.3% of total investment in year 2006. By 2016, this should reach US$2,059.8 bn or 9.6% of total.
- world operating expenditures in 2006 are expected to total US$300.2 bn or 3.8% of total government spending. In 2016, this spending is forecast to total US$480.9 bn, or 4.0% of total government spending.

The economic significance of world tourism is also highlighted by the World Tourism Organization (UNWTO)[70] who predicts that there will be one billion international arrivals

[69] World Travel & Tourism Council, who's 2006 Travel and Tourism Economic Research Report is available at: http://w-"-w.wttc.org/frameset2.htm

in the year 2010. Werthner (2003, p. 1) adds that tourism grows faster than the other economic sectors, and that this growth explains the industry's heterogeneity. Due to world tourism's SME structure, it has a huge importance for regional development. For example, in the EU there are around 1.3 million hotels and restaurants (9% of all enterprises). And 95 % of them are very small, i.e., 1-9 employees.

2.3.2 Australian Tourism Industry

The Australian tourism industry has a 2-tiered structure, with Tier I comprised of a small number of large players (e.g. airlines, hotel chains and the dominant tour operators) and Tier 2 made up of a much larger collection of small-to-medium tourism enterprises (SMTEs) (Sharma, Carson & DeLacy 2000). The industry is diffuse in character and dispersed across all regions of the country. It is characterized by a predominance of small businesses, with the Australian Bureau of Statistics (ABS)[71] suggesting that there are over 100,000 Australian SMEs contributing to the industry. Tourism is responsible for 4.7% of national GDP and employs 551,000 (fulltime equivalent) workers. This corresponds to approximately 6% of the total Australian workforce (CRC Tourism 2002).

Exports of tourism goods and services compare favourably with other Australian 'traditional' export products. Exports of tourism products for example, are greater than coal, iron, steel and non-ferrous metals, but less than food and live animals. The ABS reports that in 2003-04, the sectors which accounted for the largest share of tourism exports for international visitors were long distance passenger transportation (16%), shopping (including gifts and souvenirs) (16%), accommodation services (10%), takeaway and restaurant meals (15%), food products (8%) and fuel (7%). According to the ABS, inbound tourism accounted for $7.6 billion of total GDP in 2003-04, an increase of 5.1 %, since 2002-03, and that the inbound tourism industry share of GDP was 1.0% in 2003-04.

---

[70] http://www.world-tourism.org/

[71] Source: ABS Tourism Satellite Account, 5249.0, 2003-04: http://www.abs.gov.au/

2.3.3 Australian Tourism Accommodation Sector

The ABS December quarter 2005 survey of tourist accommodation (STA)[72] publication concluded that accommodation represents the largest economic sector in the Australian tourism market. The report provides a good indication of how significant the sector is. It includes research results for the following categories of establishments:

- Licensed hotels and resorts with facilities and 5 or more rooms.
- Motels, private hotels and guest houses with facilities and 5 or more rooms.
- Serviced apartments with 5 or more units.
- Caravan parks with 40 or more powered sites.
- Holiday flats, units and houses of letting entities with 15 or more rooms or units.
- Visitor hostels with 25 or more bed spaces.

The STA report shows that at the end of June 2004, there were 5,682 accommodation businesses operating in Australia, employing 91,399 persons. In 2003-04, accommodation businesses generated $8,095.9m in income, which represented an average of $1,424,800 per business. For this same period, total expenses incurred were $7,322.3m. The total industry value of these businesses was $4,165.9m, which equates to 0.5% of Australia's GDP for 2003-04. In 2003-04, the operating profit before tax for these businesses was $776.7m, resulting in an operating profit margin of 9.7%. During 2003-04 accommodation businesses incurred $1,120.6m in capital expenditure, with renovations and refurbishments accounting for 16.1% ($180m). The 5,682 accommodation businesses at the end of June 2004 operated 6,372 accommodation locations around Australia.

The largest contributor to accommodation types was motels with 2,396 locations which represented 37.6% of all locations. The second largest contributor was caravan parks with 19.7% (1,253 locations) of all locations. Serviced apartments and licensed hotels accounted for 9.1% (578 locations) and 8.4% (535 locations) of all locations respectively. New South Wales accounted for the highest share of business counts, income and employment, followed by Queensland and Victoria. New South Wales accounted for 32.8% (1,861) of all accommodation businesses, followed by Queensland (25.5% or

[72] Source: ABS (cat. no. 8635.0) available at: http://www.abs.gov.au/

1,450) and Victoria (22.5% or 1,279). New South Wales and Queensland accounted for 32% ($2,588.5m) and 30.7% ($2,488.3m) of all income, with Victoria contributing 16.3% ($1,323.4m) to total income. Employment in New South Wales comprised just under a third of all employment (30.9% or 28,234 persons), Queensland had 29.1% (26,553 persons) while Victoria contributed 17.1% (15,654 persons).

2.3.4 Tourism E-Commerce

Travel and tourism is an information-based business. For this reason, it was one of the first sectors to employ e-commerce applications, an example being the airline computerized reservation systems in the early 60s. According to Werthner (2003), travel and tourism has now grown to be the leading application field in business-to-consumer (b2c) e-commerce, representing nearly 50% of total b2c turnover. The industry and its product have specific features which explain this circumstance: the product is a confidence good, consumer decisions are solely based on information beforehand; and the industry is highly networked, based on world-wide cooperation of very different types of stakeholders (Werthner 2003, p. 1). Consequently, this industry depends on advanced IT applications, suggesting that it may provide a good example of what happens and will happen in the emerging e-markets with regards to structural changes and challenging application scenarios.

The increased use of the Internet for tourism related e-commerce has attracted considerable attention. For example, an analysis of the US travel market reported in (ATDW 2001) predicted that: in 2002, some 67% of travel customers would do some research online with 37% proceeding to the booking stage; and these US Web travellers would spend just under 30% of their travel budgets online, generating US$22.5 billion annually (McGrath & Abrahams, 2006b, p. 2). Furthermore, (Parker, 2003) predicts continued strong growth in the leading edge US online travel market. This is roughly consistent with more recent research such as Mills & Morrison (2003), who reported global online travel spending of US$6.9 billion in the first quarter of 2002, and PhoCusWright (2003) and Weber et al (2005), who report tourism-related businesses (and accommodation enterprises in particular) experiencing rapid growth in online sales.

Travel Recommender Systems (TRS) are now increasingly being used by tourism e-commerce Web sites to provide individually tailored travel advice to customers. Venturini & Ricci (2006) contend, however, that implementing decision support technologies in a real commercial tourism destination portal is challenging. First of all, the peculiar problems related to the tourism domain, which have been studied in the recent years in e-commerce and tourism research must be considered. To provide an effective and useful tool, one must tackle additional requirements arising from the technical and operational environment, which influence not only the software development and architectural issues, but also methodological aspects. Triplehop's TripMatcher[73] (used by www. ski-europe.com, among others) and VacationCoach's expert advice platform, Me-Print, used by travelocity[74], both try to mimic the interactivity observed in traditional counselling sessions with travel agents when users search for advice on a possible holiday destination (Ricci 2002).

2.3.5 Australian Tourism Online

In Australia, the larger Tier I organizations are generally fairly advanced in their use of online technologies. Smaller Tier 2 players have limited ICT infrastructures and knowledge, and have been relatively slow to embrace the potential marketing and business efficiency benefits offered by e-business applications (Morrison & King 2002). Internationally, the same gulf between large and small tourism enterprises has also been noted by Maedche and Staab (2002). McGrath et al. (2005a) report that perhaps, one of the most significant, relevant, Australian on-line tourism studies that has been undertaken was the Australian National Online Tourism Scoping Study, conducted by the 'Sustainable Tourism Cooperative Research Centre' (STCRC) during the late-1990s (STCRC 1999). McGrath, et al. (op. cit) summarized the major findings of the study as being that:

- The Australian tourism industry had generally achieved a comparable level of online development with international competitors.
- Larger enterprises and relevant government agencies were, in general, considerably more advanced in taking advantage of online technologies than SMTEs.

[73] http://www.oracle.com/triplehop/index.html

[74] http://www.travelocity.com/

- Despite the above, little validated data on the extent of online technology diffusion was available.
- Major impediments to online technology uptake among tourism enterprises included: poor online product coverage; marginalization of local destination product in the international online market space; online information overload; the lack of an adequate international legal framework for e-commerce; concerns over online transaction security; intermediaries being threatened by new technologies and associated role changes; and a lack of knowledge, skills, technical support, funds and time among SMTEs - especially in rural and regional areas.

The Roy Morgan May 2006 press release (article No. 492)[75] shows that Australia's tourism distribution channels and booking patterns have been radically redefined by the Internet. Travellers not only use the Internet as a means of pre-purchasing accommodation and travel tickets, but also as an important information source during the holiday planning stage. According to Roy Morgan, eighty percent of the Australian population 14 years and over have accessed the Internet at some point in their lives, with thirty-two per cent of Australians having made an online purchase. This represents an increase of 26% from 6% in the June 1999 quarter. The most purchased items over the Internet since mid-2001 are travel tickets and accommodation. In fact, figures show that 10% of the Internet users had purchased accommodation or travel tickets online during that period. Roy Morgan adds that the purchase of travel tickets or accommodation has grown from less than 1% in the June 2000 quarter. Books/Magazines/Newspapers were the next most popular product category at 7%.

Around mid 2002, Australians really started to embrace the Internet as the primary means of travel booking. This trend has continued up to the year ending March 2006, where 12% of Australian travellers aged 14 years and over were reported to have used the Internet in booking their last short domestic trip. Interestingly, travel agent bookings accounted for only 3%. For longer holidays, Roy Morgan research shows that discrepancies between bookings for domestic and international leisure trips appear. Using the Internet to book long domestic holidays has shown substantial but less rapid growth than shorter holiday bookings over time. Longer holiday bookings by Internet

[75] http://www.roymorgan.com/news/press-releases/2006/492/

overtook travel agents as the primary booking medium mid-2003, with the gap continuing to widen. For example, in the year ending March 2006, 21% of bookings for domestic leisure trips of three nights or more were booked via the Internet, compared to 12% for travel agents. It appears that human interaction is more likely required for international travel bookings, partly because of the lack of destination knowledge and complex pricing. In this market, travel agents are still the preferred booking method, even though numbers have fallen 6% since 2001 from 71% to 65%.

Roy Morgan also report that overall, the growth of international holiday bookings of three or more nights has been minimal, with just a 2% rise during the last five years, whereas the use of the Internet to book long overseas leisure trips has increased by 16% (25% cf 9%) during the same period. For the inbound international market, the Bureau of Tourism Research, International Visitor Survey[76] found that 26% of visitors to Australia the year ended 31 December 2003 used the Internet to gather information before their arrival in Australia, while 8% of these travellers made an Internet booking. Of the items booked on-line, 36% were for accommodation, 29% for international air travel, 13% for car and caravan rental, 12% for domestic air travel, 7% for organized tours, and 3% other. Roy Morgan Research (2003), indicated that Australian travel bookings over the Internet increased from less than 3% in the (financial) year to June 2001 to over 9% in the year to June 2003, and more recent research has suggested substantial continuation Internet growth in travel product purchases.

The Australian government is well-aware of the importance of online tourism to the national economy and, among various initiatives it has provided substantial support for the development of the Australian Tourism Data Warehouse (ATDW). A key objective of this highly-successful initiative is the capture and integration of national tourism-related information (e.g. accommodation, activities, events etc. data) for, among other uses, the development of advanced 'Destination Marketing Systems' (DMS) (McGrath & Moore, 2003). Another major online tourism data initiative is the Decipher project[77], which is an online data warehouse dedicated to providing the Australian tourism industry with the most recent and reliable tourism research and business intelligence. The Decipher

[76] Source: Bureau of Tourism Research, International Visitor Survey http://www/tourism.australia.com/home.asp

[77] http://www.decipher.biz/

Website is a one-stop shop for a comprehensive range of up-to-date tourism information from more than 100 qualified sources, and is a valuable tool for anyone involved with the tourism industry. Decipher was launched nationally on 10 February 2005.

Wotif.com[78] is an accommodation portal that offers a service distinct from those mentioned above. Wotif.com was launched in Brisbane, Australia in March 2000 and quickly became known as the online marketplace for hotels' distressed inventory. They pioneered selling discounted accommodation based on hoteliers' live and up-to-date inventory. By only selling a week ahead, they were able to offer great rates from the hotels. This innovative way of displaying room rates (our "hotel price matrix") added to Wotif.com's success. It gave travellers, and the hotels, an easy way to check all available prices, up-front. Customers could now see discounted pricing from a number of hotels on the one screen, and then simply book the room they wanted. Wotif.com also showed rates for the next 7 days, to give genuine last-minute prices.

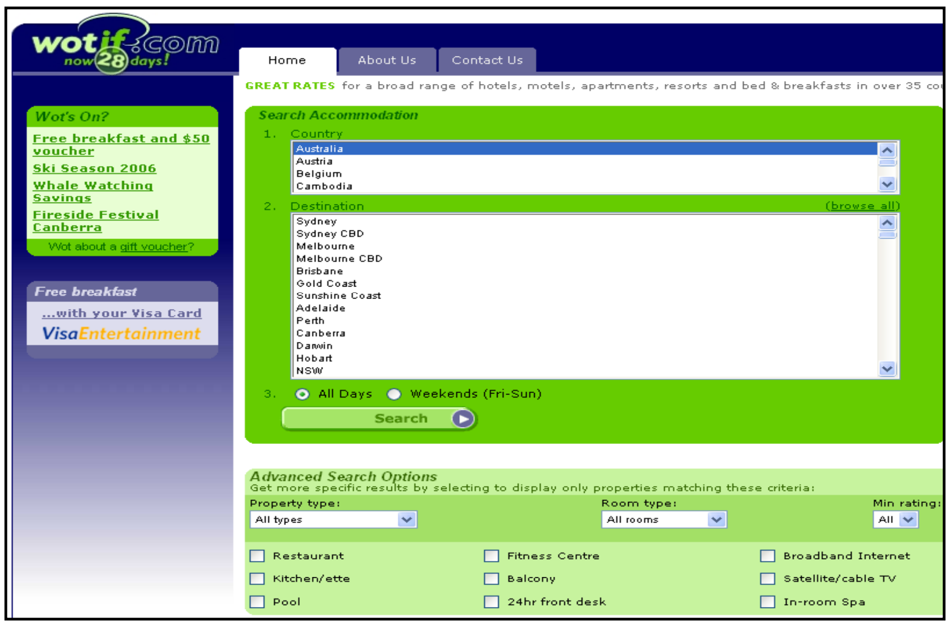


Figure 24: WhatIf.com accommodation portal.

2.3.6 Semantic Web in Tourism

Many researchers like Antoniou et al.(2005) and Bergamaschi et al.(2005) for example, believe that the tourism industry is a good candidate for the update of Semantic Web technology. Petrie (2006) supports this view on the basis that although tourism is just one

[78] http://info.wotif.com/about_our_history

application domain, researchers have naturally identified it as an ideal showcase because of its information heterogeneity, market fragmentation, and complex discovery and matchmaking tasks, including substitution and composition — all of which are limitations that Semantic Web technologies promise to overcome. The reasons for the industry being widely viewed as a suitable candidate for adoption are summarized by Maedche and Staab (2002), who describe the following industry characteristics:

- Its products are complex.
- A tourism product will perish if it not sold in time.
- The tourism industry depends on complex value creation chains involving a large number of participants (travel agencies, tour operators, hotels, etc.).

Antoniou et al. (2005) add that the current Internet poses a number of limitations to information processing, and the tasks of finding, extracting, and interpreting the information are left to human readers, which means that the need for semantically connecting the dispersed and isolated pieces of information seems to be very crucial. It is within this context that Maedche and Staab (op. cit) stress the need for (a) semantic search engines for tourism, (b) semantic based electronic markets, and (c) Semantic Web services for the tourist. Current Semantic Web based projects in the field of Tourism ICT include:

- Harmo-TEN (Dell'Erbra et al. 2005), formally known as Harmonise, is a major European Community initiative aimed at promoting tourism information systems interoperability through the adoption and use of a 'minimum tourism ontology'. The Harmo-TEN project and their approach is based on facilitating and simplifying mappings between data models based on different standards (or none). As part of their work, the Harmo-TEN team analysed existing tourism data standards and projects (Hopken 2002) and discovered: 1) more than 40 tourism-related data standards; 2) many different modelling approaches, languages and levels; and 3) while there is a fair amount of consistency between some of the major standards (e.g. the OTA and IFITT RMSIG reference models), there is also a high degree of semantic overlap and conflict. In addition, the Harmo-TEN team contends that most current tourism IT standards are low-level and that "--- harmonisation should be independent of the technical solution and should take place on a more abstract conceptual level" (Missikoff et al. 2003, p. 60). The Harmonisation process for integration has two phases[79]:

1) The *customisation phase* is based on the semantic mapping between the data owned by the user and the concepts in the Harmonise ontology. This phase is executed once when a new tourist organisation enters in the Harmonise network. The output is a set of Custom Reconciliation Rules which will be used during the Co-operation phase.

2) The *co-operation phase* aims to transform the user's data format in a representation suitable to be exchanged with any other user of the Harmonise network, based on the Custom Reconciliation Rules.

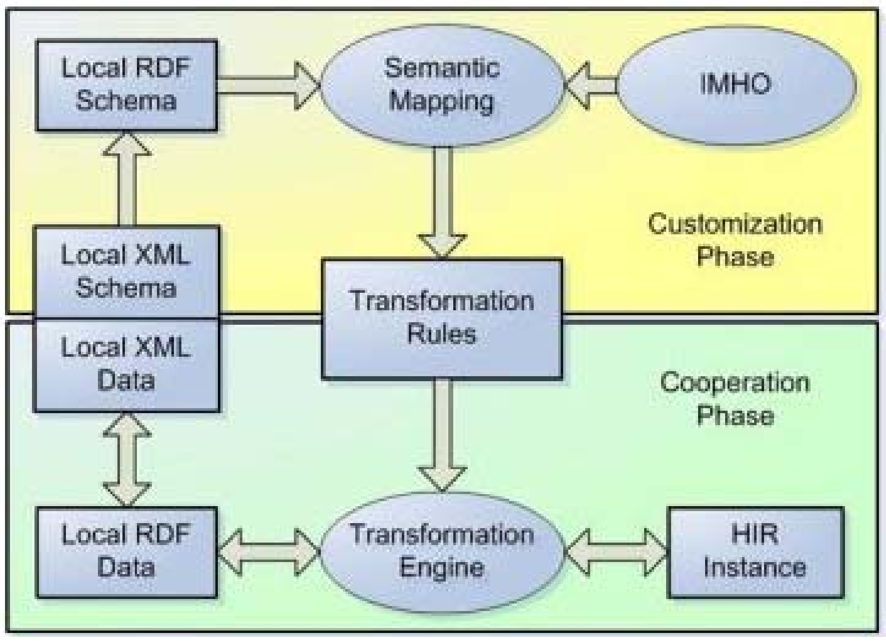


Figure 25: Harmo-TEN integration phases (Dell'Erbra et al., 2005).

- The SEED project (Cardoso, Jorge & Fernandes 2005) was started with the objective of developing a new way to implement dynamic packaging systems. To create dynamic packages, systems must integrate different tourism data sources. These data sources can have very different data formats and can be accessed by very different methods. To deal with heterogeneity, SEED use Semantic Web technology. By creating a semantic model of the tourism domain and associating this model with each one of the data sources, sources of information are more easily integrated.

[79] Source: http://www.harmo-ten.info/index.php?option=content&task=view&id=5&Itemid=29

- SATINE 6[80] stands for Semantic-based Interoperability Infrastructure for Integrating Web Service Platforms to Peer-to-Peer Networks. The SATINE Project has realized a secure semantic-based interoperability framework for exploiting Web service platforms in conjunction with Peer-to-Peer networks in the tourism industry. The aim of this project is to provide Semantic Web services on well-established service registries like UDDI or ebxmI to seamlessly interoperate with Web services on P2P Networks. Travel ontologies are being developed, and the semantics applied to Web services are designed based on standard specifications such as that of the Open Travel Alliance[81].
- The IM@GINE IT[82] project aims to develop one single access point, through which the end user can obtain location-based, inter-modal transport information, mapping and routing, navigation and other related services everywhere in Europe, anytime, taking into account the personal preferences of the user. A Key innovative feature of the project is the development of common transport and tourism ontologies for Semantic Web applications.
- Antoniou et al. (2005) have established a semantic brokering system that provides matchmaking of tourism product offerings and customer requirements by generating semantic representations of tourism data. The architecture of the broker consists of five main parts: (a) reasoning module; (b) control module; (c) semantic and syntactic validator; (d) RDF suite module; and (e) rule-query-RDF loader module. The system has been implemented as a prototype in a multi-agent environment.

2.4 Chapter 2 Summary

The chapter provided an introduction to the Semantic Web and discussed its background and potential. The need for the Semantic Web was shown to have has arisen because current Web technology presents serious limitations for searching, accessing, extracting, interpreting and processing information. In laying out a roadmap for its likely development including tools, languages, development techniques, the key elements of the

---

80 80 http://www.srdc.metu.edu.tr/Webpage/projects/satine/

81 http://www.opentravel.org/

82 http://pi.ijs.si/PiBrain.exe?Cm=Project&Project=IM@GINE+IT&Reference=508008

Semantic Web were discussed. These included knowledge representation, inference, ontologies, and semantic search. The chapter demonstrated how the Semantic Web can improve search engines by processing the underlying concepts associated with a Web page, rather than relying on keywords which is the major limitation of today's search engines. It was also demonstrated that information can be seamlessly integrated via the Semantic Web through ontology merging and alignment techniques.

A number of challenges associated with implementing the Semantic Web were widely reported in the literature, including scalability of systems, stability of Semantic Web markup languages, availability of Semantic Web content, ontology versioning and maintenance, and change management issues. Possible solutions were presented, such as those proposed for the major problem of availability of content, which could include annotation by means of creating metadata (through techniques such as text mining and semi-automated annotation).

The later part of the literature review focussed on the tousism industry. The economic significance of tourism was discussed along with various tourism ICT applications, which form part of world's the largest e-business sector. Finally, previous and ongoing Semantic Web initiatives in tourism were presented, including the Harmo-TEN project which is aimed at promoting tourism information systems interoperability through the adoption and use of a 'minimum tourism ontology'. In closing, the chapter provided an indepth analysys of other work relevant to tourism information integration and utilization within a Semantic Web context.

3 METHODOLOGY

3.1 Chapter 3 Overview

Chapter 3 outlines the research methodology. The chapter commences with a discussion about the research philosophy, based on the systems development approach to research described by Burstein (2002). The various research phases are then outlined, including a discussion of how these phases are interlinked from the initial research questions to the end proposition of a grounded hypothesis. The systems development research method of Nunamaker et al. (1990-1991) and cited in Burstein (2002p. 151) was applied to develop and test a prototype system for the purpose of generating new theory in the field of information systems. This development process and its contribution to meeting the research aims are described in detail, along with the evaluation methods used to test and validate the proposed theory. System evaluation included a comparative query experiment designed to demonstrate improved tourism information integration through the use of Semantic Web technologies. The design for a survey of tourism operators aimed at providing insight into attitudes towards the adoption of a new Internet technology is also presented, along with the secondary data used to support the survey findings. Research limitations and threats to external validity are also covered towards the end of the chapter.

3.2 Research Philosophy

The research was conducted using a systems development method to generate grounded theory. Burstein (2002 p.148) explains that systems development, as a research method, has been omitted from most *taxonomies* or classifications of information systems research methods, mainly due to the assumption that system development does not lie within the information systems research domain. According to Cerez-Kecmanovic (1994) and cited in Burstein (2002 p.148). Information systems research has been perceived by some as purely a social science thus ignoring its technological side. However, Burstein (2002) says that this view is changing as more researchers recognize that information systems involve an unavoidable technical component. Some prominent researchers such as Nunamaker and Chen (1990), Nunamaker et al. (1990-1991) and Parker et al. (1994), have debated extensively and justified the legitimacy of systems development as a valid research activity within the technical domain of information systems. The philosophy is that systems development may bridge the gap between the technological and the social sides of information systems research. Parker et al. (op. cit) contend that this aim can only be achieved by building an application of the proposed theory as an illustration of the 'technical' side of an information systems domain.

The existing taxonomy of research methods ((Neuman 1994); (Galliers 1991)), distinguish between basic and applied research. Basic research is directed towards 'theory building' and contributes to the advancement of the general knowledge of society. Burstein (op. cit) argues that to a certain extent, this kind of research can only be conducted after a field of study has reached a certain level of maturity and has all the parameters clearly defined to be generalized in the form of an appropriate theory: an established paradigm (Kuhn 1970). Alternatively, applied research targets a specific problem, which in the context of this research, relates to the introduction or functioning of an information system. In this respect applied research is closer to practice. The result of such research is intended to help practitioners to be better informed about their work environment and do their job better (Neuman 1994).

Building a theory involves discovery of new knowledge in the field of study and can be seen as rarely contributing directly to practice (Burstein, 2002, p. 149). Once a theory is proposed, however, it needs to be tested in the real world to show its validity, recognize its limitations, and make appropriate refinements according to new facts and observations made during its application. Burstein (op. cit) states that Information Systems still represents a relatively new discipline, resulting in a need and place for both types of research. She contends that in any large research project, there are identifiable elements of basic and applied research, usually one followed closely by the other.

According to Burstein (op. cit) testing can be conducted in more or less natural settings. Both interpretive and pseudo-scientific approaches can be applied. Interpretive studies represent a less-controlled mechanism of applied exploration, whereas experimentation requires a certain level of control over some of the variables under consideration. The experimentation approach assumes an ability to differentiate between controlled, independent and dependent variables. In the context of information systems research, the theory proposed may lead to the development of a prototype system that is intended to illustrate the theoretical framework (Burstein, 2002, p. 149). Thus, systems development becomes a natural, intermediate step linking basic and applied research.

Nunamaker et al. (op. cit) argue in their seminal paper on the role of systems development in information systems research that systems development represents a central part of a multi-methodological information systems research cycle. This extended structure, which is represented in Figure 26, allows multiple perspectives and flexible choices of methods to be considered in various stages of the research process. Thus, integrating a systems development component into the research cycle presents a complete, comprehensive and dynamic research process (Burstein, 2002, p. 149).

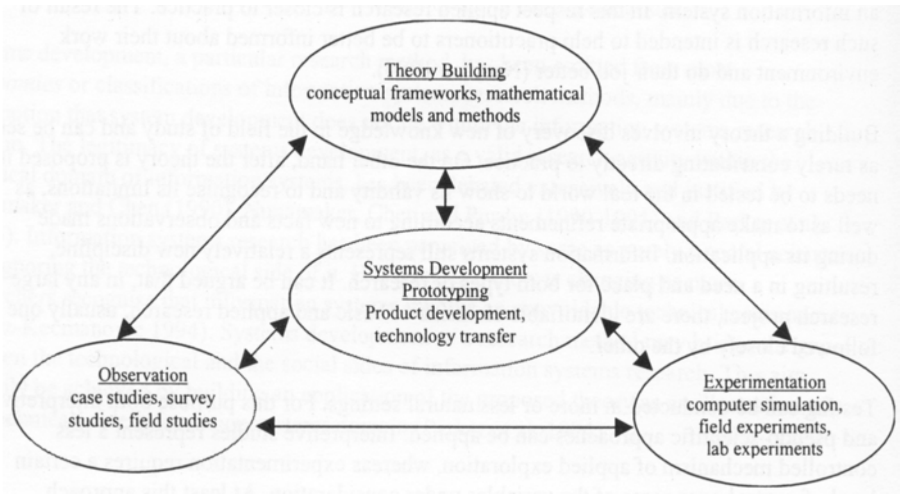


Figure 26: A multi-methodological approach (Nunamaker, Chen & Purden 1990-1991).

This multi-methodological approach was also applied to this thesis. The systems development process was augmented with survey type research (i.e. Tanner 2002) to provide a holistic view of technology, people, structure and processes in the topic area.

3.3 Research Phases

The research was conducted over a number of interlinked and sometimes concurrent phases (see Figure 27), commencing with an initial review of literature that identified knowledge gaps in the topic area and led to the initial research aims, to the end proposition of a grounded hypothesis.

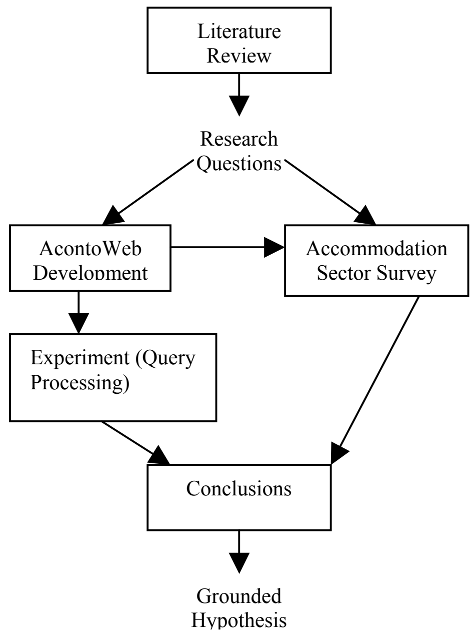


Figure 27: Research phases.

The aims were established in the first phase of the research after an initial review of available literature identified knowledge gaps in the topic area. The approach taken to meet the research aims was to build theory from available knowledge that could be used to predict the likely success or failure of a design, and then test the validity of this theory to demonstrate proof of concept. The research aims, as previously described in Chapter 1, were to:

- Provide an understanding of issues and problems involved in defining, establishing, capturing, integrating and using the heterogeneous, scattered and diverse supplier source data necessary for the development of Semantic Web based tourism applications.
- Specify a theoretical and conceptual solution to these data-related problems that addresses technical limitations with existing integration approaches and takes into account the critical social dimension.
- Develop a proof of concept DMS prototype (based on the conceptual model discussed earlier), restricted to matching tourism customers accommodation needs to suppliers' offerings. This prototype (titled *AcontoWeb)* will be 'ontology-driven'.
- Demonstrate the effectiveness of the DMS with regard to usability and value-adding potential for tourism industry customers and service providers – via a survey and experiment.

- Gain an insight into the attitudes towards the adoption of semantic technology by SMTEs and their requirements and preferences in relation to implementation and usability of such systems.
- Generate a grounded hypotheses that can be tested in further research.

A further (more in-depth) investigation of the topic area led to the construction of the major and minor research questions. The questions were designed to provide answers that would fill the knowledge gaps identified by the literature review, and thus, help build new theory and meet the research aims. The research questions, which were previously stated in Chapter 1, are:

Major Research Question

To what extent can the Semantic Web and related technologies assist with the creation, capture, integration, and utilization of accurate, consistent, timely, and up-to-date Web based tourism information?

Minor Research Questions

- What is the ease of ontology development, availability, and Website annotation?
- What level of ontology and Website annotation richness can be obtained?
- What is the maturity and ease of use of Semantic Web development tools?
- How robust are Semantic Web operational environments at present?
- How can the Semantic Web best be queried?
- What are the potential query results and accuracy?
- How do query results compare to that of conventional database systems?
- How useful is the Semantic Web and what are its limitations?
- Howe successfully can tourism information be integrated on the Semantic Web?
- What are the managerial issues faced in gaining user acceptance of Semantic Web technology in the tourism industry?

Having conducted the literature review and established the research aims and questions, the next phase was to develop a prototype that could be used to build and test new theory. The development was done by following the prototype systems development research process of Nunamaker et al. (1990-1991) as illustrated in Figure 28.

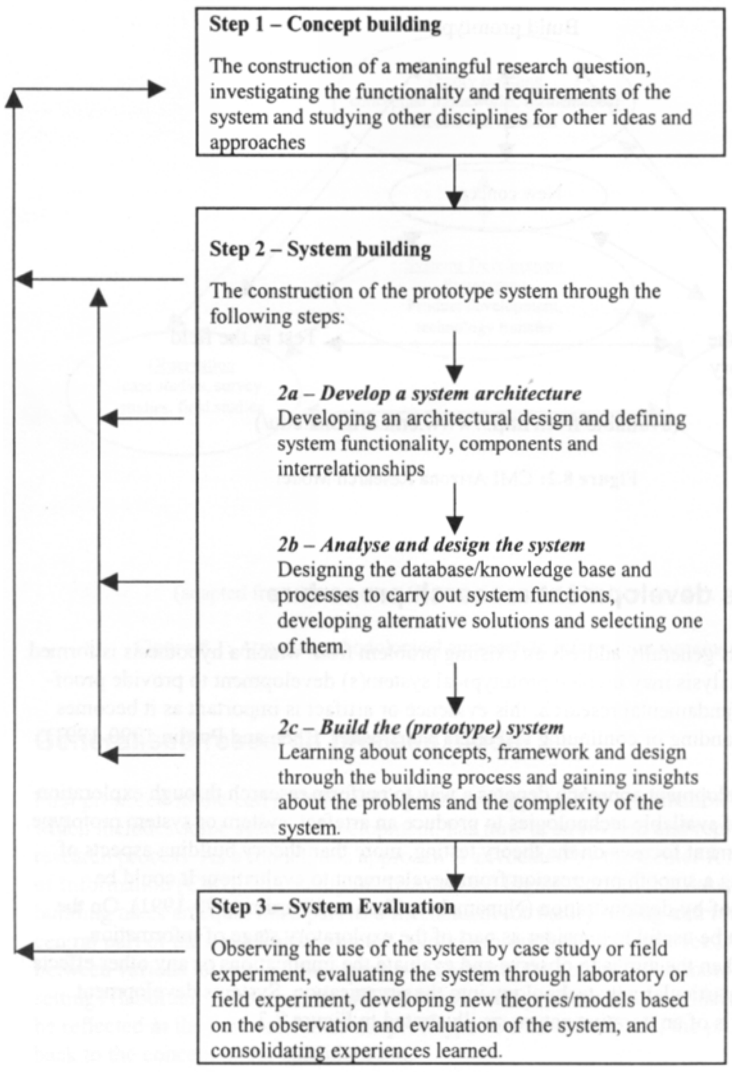


Figure 28: The systems development method (Nunamaker, Chen & Purden 1990-1991).

At the beginning of such a project, the implementation has to be justified as genuine research in terms of whether there is another existing system capable of demonstrating the features of the concepts under investigation (Burstein, 2002, p. 153). The review of literature identified some other Semantic Web initiatives in tourism. None of these projects, however, had attempted to make use of OWL semantics in an ontology model to improve information integration by inferring knowledge about the attractions associated with a resort based on the resort's location (and similar inferential processes). Nor did any of the projects investigated reclassify location types based on tourism market segments in the way AcontoWeb was designed to do. No other tourism related projects could be found that compared the complexity and subsequent ease of information integration of querying an ontology to that of a relational data model. The design of the AcontoWeb annotation tool was also unique, in that user input is accepted by the system, automatically transformed into RDF markup and imbedded into a Webpage. Other comparable RDF annotators such as OntoMat-Annotizer[83] require annotations to be manually dragged from ontology concepts into a Webpage.

From a technical point of view, probably the most unique aspect of the AcontoWeb design is its generic reasoning and SPARQL querying capability. The system was specified to allow any OWL DL (Description Logic) ontology to be loaded into a Jena supported backend, classified with a reasoner, and SPARQL queries run over the inferred version of the ontology. This design represents a significant advancement in presently available technology because it allows information to be reorganized to suit different user needs using a completely different navigation structure, while at the same time providing access to a SPARQL query engine that processes inferred knowledge. Other SPARQL query tools such as Semqueries[84] or the Protégé SPARQL[85] query tab only work on a static (base) ontology model. Finally, the research was unique because of the holistic approach taken, which included an investigation of managerial issues associated with adoption of the Semantic Web technology. Other tourism related Semantic Web projects focused mainly on technical issues.

The prototyping phase consisted of three major steps: concept development; system building; and system evaluation. As shown in Figure 28, the concept building stage involved some theory building, where the theory can be illustrated by a system. In the case of this research, theory identified in the literature suggested that the Semantic Web had the potential to improve upon current Internet technology by allowing Web authors to explicitly define their words and concepts, thus giving information well-defined meaning. It was theorized that this would allow software agents to analyse the Web on our behalf, making smart inferences that go beyond the simple linguistics performed by today's search engines, better enabling computers and people to work in cooperation (Berners-Lee, Hendler & Lassila 2001). The literature also suggested that the limitations of the current Internet had made the effective integration and utilization of Web based tourism information a difficult time consuming task (Staab 2005). The theory proposed

[83] http://annotation.semanticweb.org/ontomat/index.html

[84] http://semweb.krasu.ru/SemQueries/

[85] http://protege.stanford.edu/plugins/owl/sparql.html

by this research was that the use of Semantic Web technologies in tourism ICT applications could significantly improve the integration and utilization of Web-based tourism information.

The second step in the prototyping phase was system building. A system was developed using available technologies deemed capable of testing the validity of the proposed theory and illustrating a new theoretical framework, which in this case, was '*Semantic Web based tourism information integration via semi-automated annotation and intelligent querying'*. System building involved the design of the system architecture, the specification of the knowledge base (Accommodation ontology), and coding of the system. The major difference between this approach as a research method and conventional systems development is that the major emphasis is on the concept that the system has to illustrate, and not so much on the quality of the system implementation (Burstein, 2002, p. 153).

Evaluation of the prototype was needed to test the validity of the proposed theory and recognize its limitations, as well as to make appropriate refinements according to new facts and observations made during its application. The evaluation stage of the systems development method also differs from testing a commercial system. It has to be done from the perspective of the research questions set up during the concept-building stage, and the functionality of the system is very much a secondary issue (Burstein, 2002, p. 153). An interpretive evaluation approach was applied to answer the more general research questions such as 'How useful is the Semantic Web and what are its limitations?', and 'How robust are Semantic Web operational environments at present?' The answers to these questions resulted from general observations made of the Semantic Web technology throughout the development process. An experimentation approach was used to make technical observations that could answer questions such as 'What are the potential query results and accuracy?' and 'How do query results compare to that of conventional database systems?' A query evaluation model (see sub-section 3.4.1) provided the necessary control over variables used in the experiment. Figure 28 shows that systems development research is iterative, with results from the system evaluation used to refine the initial concept proposed.

The AcontoWeb development phase was accompanied by a survey phase as shown in Figure 27. Many researchers (see e.g. (El Sawy 2001)) have stressed the necessity to take a holistic view of technology, people, structure and processes in IT projects and, more specifically, Sharma et al. (2000, p. 151) have noted that as significant as DMS technological problems are, they may well pale into insignificance when compared with the managerial issues that need to be resolved. With the need to take a holistic approach emphasised, a survey of tourism operators was conducted to provide insight into the attitudes towards the adoption of a radical new Internet technology among tourism operators. The survey was intended to provide an answer to the question of 'What are the managerial issues faced in gaining user acceptance of Semantic Web technology in the tourism industry?' The survey was also aimed at providing an understanding of the usability requirements that tourism operators have for such a technology so that these preferences could be incorporated into the design of the prototype.

The conclusion phase of the research was where the overall findings were evaluated and presented. Answers are provided in the conclusion to both the major and minor research questions, along with the proposition of a grounded hypothesis that was formed through the exploration of knowledge undertaken throughout the research phases. The grounded theory expresses a viewpoint about the extent to which the Semantic Web and related technologies can contribute to the creation, capture integration, and utilization of accurate, consistent, timely, and up-to-date Web based tourism information.

3.4 Experimental Design

This section presents the design of the AcontoWeb query experiment comparing complexity of querying and subsequent ease of information integration of a semantic portal that uses a rich ontology for indexing purposes, to that of a conventional portal supported by a relational database.

3.4.1 Query Evaluation Model

A query evaluation model was formulated to assist with query complexity analysis. The model was created from a Business Information Systems perspective (i.e. without applying excessive mathematical formulas), and was designed to demonstrate in a practical manner, potential benefits for information integration that can be obtained by taking advantage of the semantics and reasoning capabilities of OWL ontologies. In description logics, theory is divided into two parts: 1) the TBox which contains intentional (terminological) knowledge through the declarations that describe general properties of concepts; and 2) the ABox which contains extensional (assertional) knowledge specific to the individuals of the domain. It is important to recognize that the notion of data complexity is viewed here in the same context as that prescribed by Vardi (1982), in which the premise is that an ABox can be naturally viewed as a relational database. Equally important is the notion that the query evaluation model was designed from a knowledge base centric view, meaning that the logical complexity of querying a particular data model itself was evaluated, rather than individual query representational languages like SQL or SPARQL.

3.4.2 Conjunctive Queries

In order to compare queries between a relational data model and an ontology, it is necessary to deal with unary and binary predicates in the query expressions that correspond to classes and relations from the ontology. Conjunctive queries (Kolaitis & Vardi 1998) are queries that are generalized so that they can be bound to different views of a particular domain. For objective comparisons to be made, it is important to initially have a generalized view of query expressions (not specific to any particular representational language) as shown in Figure 29.

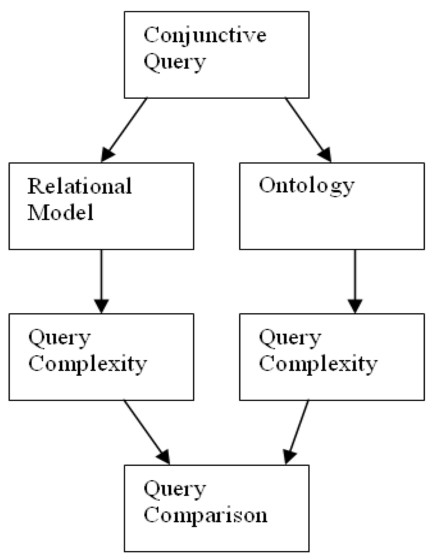


Figure 29: Conjunctive queries.

Using the general notion of terminological knowledge provided by Struckenschmidt et al. (2005a, p. 132), conjunctive queries in this thesis are defined in accordance with the following:

Definition 3.4.2.1 (Terminological Knowledge Base)

A terminological knowledge base T is a triple
T = [C, R, I]
where C is a set of class definitions, R is a set of relation definitions and I is a set of object definitions.

Further, the signature of a terminological knowledge base is defined as a triple *[CN, RN, IN],* where *CN* is the set of all names of classes defined in *C, RN* the set of all relation names and *IN* the set of all object names occurring in the knowledge base.

Definition 3.4.2.2 (Terminological Queries)

Let V be a set of variables disjoint from IN; then a conjunctive query Q over a knowledge base T is an expression of the form[86]

$$q_{1_i} \wedge \cdots \wedge q_{m_i}$$

where q_i *are query terms of the form* $x : c$ *or* $(x, y) : r$ *such that* $x, y \in V \cup \mathcal{IN}$, $C \in \mathcal{CN}$ *and* $R \in \mathcal{RN}$.

Figure 30 is a conjunctive query over two separate views (the relational model and an ontology shown in Appendices D and E) of the same accommodation domain. The query asks for *Accommodation* that lies in a *Destination* that has a *Museum*. The accommodation must have a *Hotel-Motel* category, and facilities are to include a *Swimming Pool*.

Q(X) ← Accommodation(X) ∧ hasAccommodaitonDestination(X,V) ∧
hasDestinationAttraction(V,W) ∧ hasCategory(X,Y) ∧ hasAccommodationFacility(X,Z) ∧
W = Museums ∧ Y = HotelMotel ∧ Z = SwimmingPool

Figure 30: Conjunctive query.

[86] Logic symbols used in this thesis are specified in Appendix C

Using a method proposed by Horrocks and Tessaris (2000), the Figure 30 conjunctive query can now be translated into an ontology concept expression. The idea proposed by Horrocks and Tessaris (op. cit), and demonstrated by Stuckenschmidt et al. (op. cit) in an Accommodation domain, is to translate the query into an equivalent concept expression, classify this new concept and use standard inference methods to check whether an object is an instance of the query. The approach relies on the fact that binary relations in a conjunctive query can translate into an existential restriction in a way that preserves consequence after a minor modification of the A-box. Details are given in the following theorem:

Theorem 3.4.2.1

Let *[C, R, A] be a description logic knowledge base with concept definitions C, relation definitions R and assertions A. Further, let R be a role,* C_I *concept names in C and a, b individual names in A. Given a new concept name* P_b *not appearing in C, then:*

$$\langle \mathcal{C}, \mathcal{R}, \mathcal{A} \rangle \models (a,b) : R \wedge b : C_1 \wedge \cdots \wedge b : C_k$$

If and only if:

$$\langle \mathcal{C}, \mathcal{R}, \mathcal{A} \cup \{b : P_b\} \rangle \models a : \exists R(P_b \sqcap C_1 \sqcap \cdots \sqcap C_k)$$

Dependencies between the variables that occur in the query expression make transformation of a complete query more difficult. Horrocks and Tessaris (op. cit) introduce the notion of a query graph to keep track of these dependencies during the transformation.

Definition 3.4.2.3 (Query Graph (Horrocks and Tessaris, (2000))

The graph induced by a query is a directed graph with a node for every variable and individual name in the query and a directed edge from node x to node y for every role term (x,y) : R in the query.

The correct transformation of a query to a concept expression depends on the relations between query variables, which is reflected by the query graph structure. While the approach of Horrocks and Tessaris (op. cit) is more general, in the model presented here, queries are restricted to where the query graph is a (directed) tree and its root node corresponds to the variable of interest (i.e. *Accommodation*). For this to work, none of the roles used in the query are allowed to be declared functional, and each constant may only appear once in a query. While using this simplification in their example, Stuckenschmidt et al. (op. cit) emphasize that the translation can be done for unions of conjunctive queries with an arbitrary number of result variables and a very expressive logical language for defining class expressions. The simplifying assumptions lead to a simple method for transforming a query graph into a concept expression.

Definition 3.4.2.4 (Query Roll-up (Horrocks and Tessaris, (2000))

The roll-up of a query Q with query tree G is a concept expression derived from Q by successively applying the following rule:

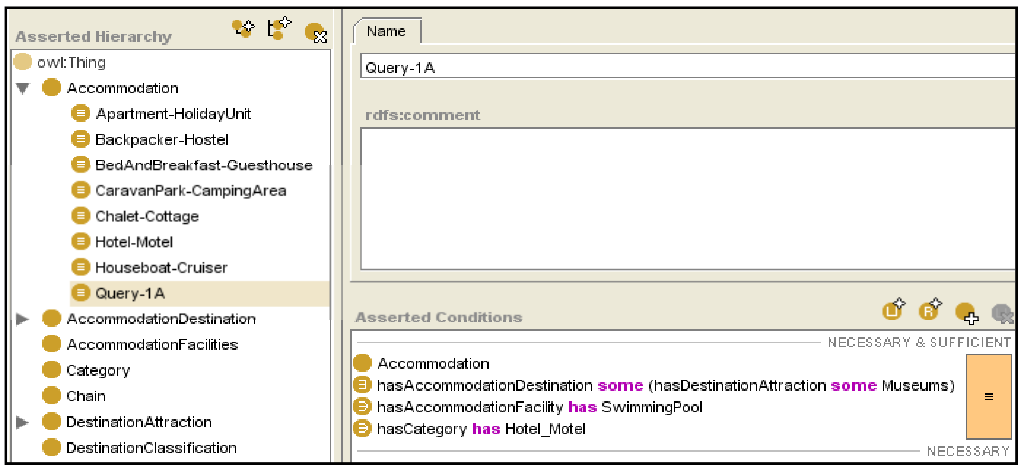
- *If G contains a leaf node y then the role term (x,y):R is rolled up according to theorem 3.4.2.1. The edge (x,y) is removed from G.*

The result of applying this translation technique to our conjunctive query example is shown in Figure 31 - as a concept expression asking for an *Accommodation* resort that lies in a destination that has a *Museum*, has a classification of *Hotel-Motel*, and facilities that include a *Swimming Pool*:

$$(\mathit{Accommodation} \sqcap (\exists\ \mathit{hasAccommodaitonDestination}.(\exists\ \mathit{hasDestinationAttraction}.\{\mathit{Museums}\})) \sqcap (\exists\ \mathit{hasCategory}.\{\mathit{Hotel\text{-}Motel}\} \sqcap (\exists\ \mathit{hasFacility}.\{\mathit{SwimmingPool}\}))$$

Figure 31: Conjunctive Query-1A represented as an ontology concept.

This concept can now be tested for instances in Protégé[87] by creating the query as concept *Query-1A* in the Accommodation ontology using NECESSARY & SUFFICIENT Asserted Conditions to specify query terms. The racer reasoner is then applied within Protégé to create the inferred hierarchy.



87

Figure 32: Concept Query-1A as an ontology concept in Protégé.

By right clicking on the concept *Query-1A*, then selecting "Compute individuals belonging to class", a list of instances matching the query concept is displayed.

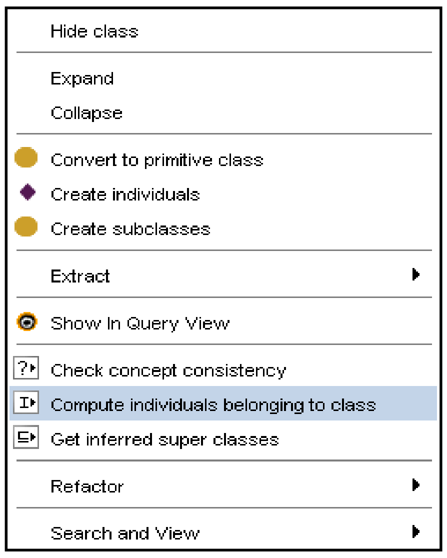


Figure 33: Computing class individuals.

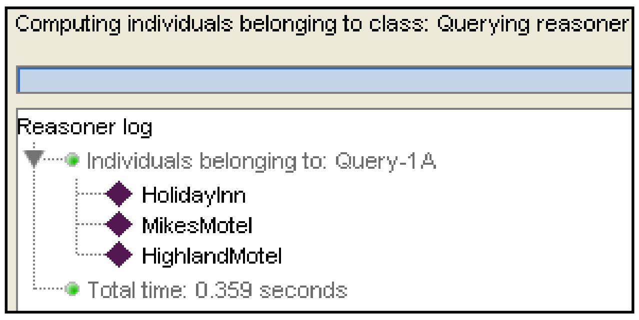


Figure 34: Query-1A results.

One of the main advantages of using an OWL ontology to model a domain is the ability to use semantic relations between objects via inference. Taking advantage of transitive properties and class restrictions in queries can reduce the number of variables, and therefore the number of value to variable assignments. In the Accommodation ontology, the transitive property *hasDestinationAttraction* provides a direct association between instances of the concept *Accommodation* and the instances of the concept *DestinationAttraction*. The direct association means that *Query-1A* can be shortened from that shown in Figure 30 by removing the *hasAccommodationDestination* predicate. With the use of OWL class restrictions, the reasoner is able re-classify the ontology so that resorts with a *Hotel-Motel* category value automatically become instances of the *Hotel-Motel* class. This allows the query to be initiated from the root class *Hotel-Motel*, which is a lower level more specific subclass of the *Accommodation* class. In this case *Query-1A* can be shortened to:

$$(Hotel\text{-}Motel \sqcap (\exists\ hasDestinationAttraction.\{Museums\})) \sqcap (\exists\ hasFacility.\{SwimmingPool\}))$$

Figure 35: Conjunctive Query-1B represented as an ontology concept.

The shortened version of *Query-1A* can now be created as a concept in Protégé. This time it is called *Query-1B*.

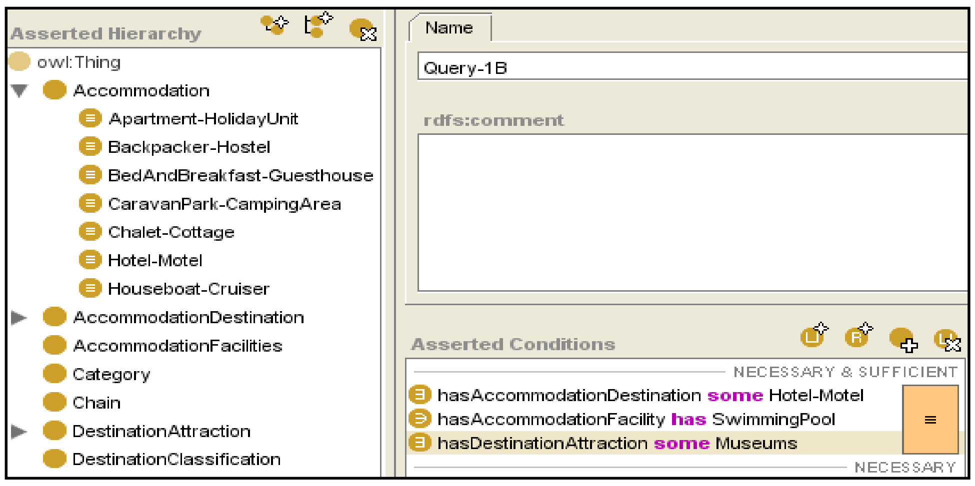


Figure 36: Query-1B as an ontology concept in Protégé.

Testing the new version of the query by applying the reasoner shows that the result for *Query-1B* is the same as the result of the original *Query-1A*.

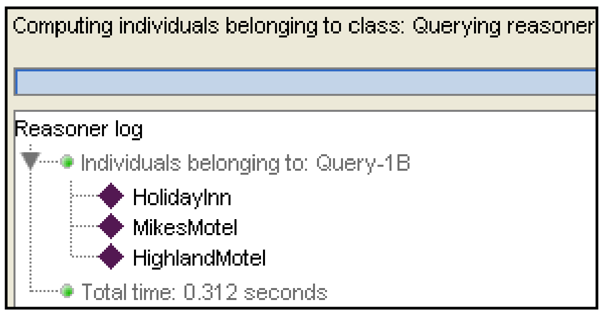


Figure 37: Concept Query-1B results.

In order to compare the complexity of Query-1A to Query-1B, the Horrocks and Tessaris (op. cit) transformation process first needs to be applied to concept *Query-1B* in reverse. The reversal allows the simplified *Query-1B* to be compared to *Query-1A* in the same format that the original conjunctive query was specified in Figure 30. To test the validity of this reverse process, one can take the result of the inverse transformation, and then re-apply the Horrock's & Tessaris (op. cit) method in its normal forward order (see Figure 38). If the result is the same as concept *Query-1B* (which in this case it is), then the reverse transformation is proven to be valid.

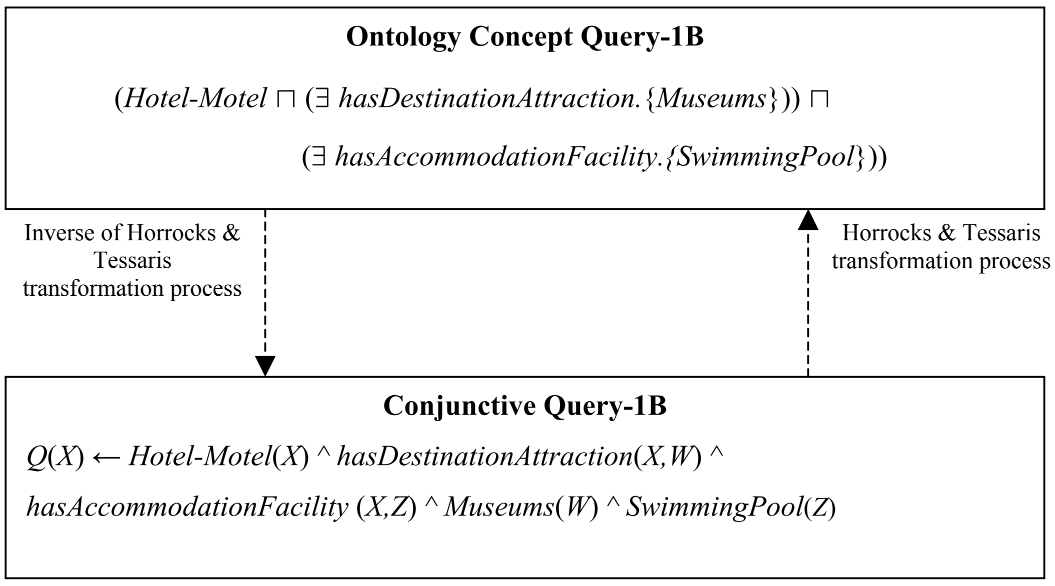


Figure 38: Inverse transformation of concept Query-1B.

The simplified *Query-1B* can now be compared for complexity to the original conjunctive query shown in Figure 30 (which was processed as *Query-1A*).

3.4.3 Measuring Query Complexity

Vardi (1982) describes three ways to measure the complexity of queries over a database. First, one can fix a specific query in the language and study the complexity of applying this query to arbitrary databases. The complexity is then given as a function of the size of the databases. This is often referred to as data complexity. Alternatively, one can fix a specific database and study the complexity of applying queries represented by arbitrary expressions in the language. The complexity is then given as a function of the length of the expressions. This is often referred to as expression complexity. Finally, one can study the complexity of applying queries represented by arbitrary expressions in the language to arbitrary databases. The complexity is then given as a function of the combined size of the expressions and the databases. This is often referred to as combined complexity.

Vardi (op. cit) in his seminal paper on evaluating and measuring the complexity of database query languages, contends that combined complexity is pretty close to expression complexity. In this research, queries represented by arbitrary expressions were evaluated against a specific data model (i.e. RACV's accommodation database). It was therefore query expression complexity that was evaluated. The following definition by Vardi (op. cit) was used as a basis for the complexity measure:

Definition 3.4.3.1 (Vardi 1982, p. 138)

Let φ be a sentence of size s (a sentence represents a query). φ has at most s variables. In order to evaluate φ on a database of size n, it suffices to cycle through at most n^s possible assignments of values from the database to the variables.

Using the above definition, query complexity can be defined using formal logic as:

$\exists\varphi\ ((\varphi \rightarrow s) \wedge (s \equiv n^s))$

In plain English the formula reads that: for some sentence φ, the sentence has a size s, and s equals the number of possible value to variable assignments from the database that may be assigned to φ.

Complexity of φ can therefore be expressed as the function:

$s \equiv n^s$

The value of s (query complexity) can then be obtained by simply calculating:

n^s

For easier visual comparison, the conjunctive queries *Query-1A* and *Query-1B* are placed together in Figure 39.

Conjunctive Query-1A
$Q(X) \leftarrow Accommodation(X)\ \wedge\ hasAccommodationDestination(X,V) \wedge hasDestinationAttraction(V,W) \wedge hasCategory(X,Y) \wedge hasAccommodationFacility(X,Z) \wedge W = Museums \wedge Y = Hotel\text{-}Motel \wedge Z = SwimmingPool$
Conjunctive Query-1B
$Q(X) \leftarrow Hotel\text{-}Motel(X) \wedge hasDestinationAttraction(X,W) \wedge hasAccommodationFacility(X,Z) \wedge W = Museums \wedge Z = SwimmingPool$

Figure 39: Conjunctive queries to be compared.

The following assumptions have been made for demonstration purposes:

- There are 100 Accommodation resorts with a Hotel-Motel category and a swimming pool facility.
- There are 50 Town-Suburb destinations that have a museum.
- There are 10 Accommodation resorts with a Hotel-Motel category and a swimming pool facility in destinations that have a museum.

For conjunctive *Query-1A*, the first variable *X* represents instances of Hotel-Motels. There are 100 accommodation resorts with a Hotel-Motel classification, meaning that the query commences with a value of 100 for X^s. Variable *V*, which represents destinations with the attraction *Museum*, has 50 possibilities, meaning that the value of V^s is 50. The value of n^s can be obtained by calculating the total possible combination of value to variable assignments for *X* and *V*. For conjunctive *Query-1A* this is X^s * $V^{s,}$ which is: 100 * 50 = 5000.

For conjunctive *Query-1B*, variable *X* directly represents instances of the *Hotel-Motel* class which is a subclass of *Accommodation*. The query can be initiated at this subclass level because the reasoner is able to re-classify instances of *Accommodation* with a *Hotel-Motel* category as instances of the *Hotel-Motel* class. Once again, variable *X* starts with 100 possible value assignments, because there are 100 accommodation resorts that are instances of the *Hotel-Motel* class. In *Query-1B* there are no value assignments to variable *V*. This is because the use of transitive property *hasDestinationAttraction* means that the *hasAccommodationDestination* property was able to be removed from the query when processed against the inferred ontology model. This effectively means that *Query-1B* was asking for *Accommodation* that *hasDestinationAttraction* with the constant value *Museum,* instead of asking for (as in *Query-1A*) *Accommodation* that *hasAccommodationDestination*, of which the variable *AccommodationDestination* has a *DestinationAttraction* with the constant value of *Museum*. The *Museum* clause in *Query-1B* is no longer bound to the *AccommodationDestination* variable *V*, but is now a direct constant value of the *Accommodation* variable *X*. Because there are 10 *Hotel-Motels* with the constant values of *Swimming Pool* facility and *DestinationAttraction Museum*, for *Query-1B,* the value of *X* has is now 10 and the value of n^s is therefore also 10.

The comparison of *Query-1A* and *Query-1B* is shown Table 4. Variables occurring in more than one query term in an *and* connected query evaluate similar to an equi-join in SQL (for relational databases) (Schaffert & Bry 2004, p. 6). Equivalent equi-joins are therefore included in the evaluation model for description analysis along with the number of query terms (represented by number of brackets clauses, e.g. (*V*)).

Measure	Results
Conjunctive Query-1A	
X^s	100
V^s	50
n^s	5000
Query terms	6
Equivalent equi joins	1
Ordinal complexity ranking	1
Conjunctive Query-1B	
X^s	10
V^s	0
n^s	10
Query terms	4
Equivalent equi joins	0
Ordinal complexity ranking	2

Table 4: Query evaluation model.

The evaluation model shows that *Query-1A* is more complex because there are 5000 total value to variable assignments compared to 10 for *Query-1B*. As a consequence, *Query-1A* had 1 more equivalent equi-join than *Query-1B*. *Query-1A* also had 6 query terms compared to 4 for *Query-1B*, meaning that the use of OWL semantics and a reasoner made the query easier to formulate. It is important to note that the evaluation model does not provide a finite statistical measure of the actual degree of a query's computational complexity. It cannot be said for instance, that *Query-1A* is 500 times more complex than *Query-1B* because the possible combination of value assignments is 5000 compared to 10. For that type of measure, a more in-depth mathematical analysis is required that must consider query optimization issues such as those covered by Calvanese et al. (2006), in which they characterize the LogSpace boundary of the problem, (i.e., finding maximally expressive DLs for which query answering can be done in LogSpace). This type of analysis, while acknowledged as relevant to measuring query complexity, lies outside the scope of the Business Information Systems nature of this research.

What the evaluation model presented here does provide is an ordinal query complexity ranking based on Vardi's (1982) prescribed evaluation measure. Using Vardi's (op. cit) theorem, Conjunctive *Query-1B*, which was only made possible because of the availability of OWL semantics and a reasoner, was less complex than the original *Query-1A.* This had the effect of eliminating an equivalent equi-join and reducing the number of query expressions, thereby making the query easier to formulate. The actual degree to which this computational complexity was reduced, however, can only be measured with a more in-depth mathematical analysis.

3.4.4 Experimental Queries

The comparative experiment was conducted with four conjunctive queries for accommodation resorts. Each query was initially run against a relational data model based on the structure of the RACV (AAA tourism) portal. The queries were then transformed using Horrocks and Tessaris' (2000) method and tested in the AcontoWeb semantic portal environment. Queries were tested in an ordered hierarchy based on the number of query terms, similar to the hierarchy established by Jansen (2000) in his study on the effect of query complexity on Web searching results. Starting from a basis of Level 1 though to Level 4, the queries used in the experiment are presented below:

- **Level 1 –** A basic query that searches for accommodation with certain constant values.

  Query 1 - A search for four star apartment/holiday units with a swimming pool, air-conditioning and conference facilities in Lorne Victoria.

- **Level 2** - A slightly larger query that searches for accommodation with constant values that lies in a location also containing constant values.

  Query 2 - A search for four star bed and breakfast/guesthouses with an open fireplace in a location that has surfing and bushwalking.

- **Level 3 –** An even larger query that searches for accommodation with constant values that lie in a location that can be classified as a certain type of location, based on the constant values of that location.

Query 3 - A search for three star caravan park/camping areas with barbeque and cooking facilities, that lie in a location classified as a backpacker location because of the associated attractions and accommodation resorts in the vicinity.

- **Level 4 –** The largest query of the experiment that searches for accommodation with constant values that lies in a location also containing constant values, and the location can be classified as a certain type of location, based on the constant values of that location.

  Query 4 - A search for a five star hotel/motel with conference facilities and a spa in an adventure destination somewhere in QLD with the attractions of beaches and guided tours.

The query experiment is documented in section 4.3 of the next chapter. It was anticipated before the experiment that the results of querying the data model of a semantic portal compared to that of a conventional portal would be identical. Query complexity, however, was expected to vary. The experiment was conducted by the researcher in July 2006. Complexity analysis was based on the original conjunctive format of the queries, rather than the SQL or SPARQL representations. The AcontoWeb front end implementation, SQL and SPARQL representations are not included in Chapter 4, but are provided in Appendices H, I and J for reference purposes.

3.5 Survey Design

This section outlines the design and objective of the tourism operator survey.

3.5.1 Sample Group

The principal purpose of the survey was to indicate the degree of interest among Australian accommodation enterprises in an advanced, new online technology. The survey was a ‘captive group’ survey with businesses randomly selected from the RACV accommodation portal. The RACV portal lists over 12,600 hotels, motels, guesthouses, B&Bs, cabins, holiday units, chalets, lodges and even houseboats Australia wide. The information was provided by AAA Tourism which is a subsidiary of Australian Motoring

Services[88] (AMS). AAA Tourism, in partnership with Australia's auto clubs manages the Australian STAR Rating Scheme, which provides consistent STAR Ratings for Australian accommodation listings. The scheme also publishes accommodation guides which are essential references when choosing accommodation or planning a holiday, and provides comprehensive and reliable information available online via the Auto Club Websites.

The survey was Web-based and created using Survey Solutions 6 software[89]. The link to the questionnaires was sent on February 16, 2005 to 4,632 eMail addresses taken from the Royal Automobile Club of Victoria (RACV) online accommodation component of the AAA Tourism Website. 600 messages were returned from expired or invalid addresses and, from messages received (plus a follow-up analysis of the address names of non-respondents), it was estimated that a further (approximate) 800 addresses from the original list did not belong to accommodation enterprises (but identified wineries, art galleries, skydiving operations etc.). The survey was left open for four weeks by which time 383 valid responses were received, giving a response rate of approximately 12%. This is quite reasonable for a Web-based survey of this type, but the sampling approach does contain some bias for which external validity implications are discussed in section 3.7. The final version of the survey contained 19 questions and is shown in Appendix F along with the message sent to subjects.

3.5.2 Pilot Survey

Business operators were contacted by telephone to request their participation in the pilot. Those willing to participate were sent a link to the survey. Twenty operators completed the pilot, with most contributing positive feedback about the survey design. The following suggestions were received.

1. The follow-up email requesting survey participation referred to tourism operators in general, rather than specifically to accommodation providers. This needed clarifying.
2. The meaning of Question 18, which asked how any new technology should be applied, was ambiguous and needed to be re-phrased.

---

[88] http://www.australianmotoringservices.com.au/

[89] http://www.mbaware.com/sursolforweb.html

3. It should be mentioned in the introductory email that the information obtained would be used purely for academic purposes.

The questionnaire was subsequently modified and re-sent to the pilot survey subjects who had raised initial concerns. Confirmation was then sought to ensure that the concerns had been addressed.

3.5.3 Survey Questions and Data Analysis

The survey was developed in accordance with the principles of good survey design as prescribed by Ticehurst and Veal (2000a). Namely steps, were taken in the wording of questions to:

- Avoid ambiguity.
- Simplify wording were possible.
- Avoid the use of jargon.
- Avoid leading questions.
- Ask only one question at a time (avoid multi-purpose questions).

The ordering of questions was also considered important with the following principles of Ticehurst and Veal (op. cit) adhered to:

- Start with easy questions.
- Start with 'relevant' questions.
- Leave sensitive questions to last.

Questions were carefully selected to ensure that the data requirements specified in the methodology concerning managerial issues were met. Questions were designed to gather the following information from tourism operators:

- Purpose of their business Website.
- Likelihood of overhauling business Website in the near future.
- Factors that would encourage or discourage overhauling of business Website.
- Creator and maintainer of business Website.
- Preferences and needs for any new business Website.
- Likelihood of adopting a new Internet technology.
- Factors that would encourage or discourage adoption of a new Internet technology.

The survey was created, and conducted, and the results were analysed by the researcher. Results were graphed and descriptive analysis applied to document the findings with simple frequency distributions produced for the responses to each question. More in-depth statistical methods such as factor analysis were not required for data analysis. The survey was not intended for example, to provide a definitive answer as to the proportion of members of a certain demographic that definitely would or would not use the Semantic Web technology. The survey simply aimed to provide a general indication of attitudes towards adoption of the Semantic Web among tourism operators to accompany the results of the technical experiment.

3.6 Secondary Data and Analysis

The secondary data was part of a research project sponsored and funded by the Australian Sustainable Tourism Cooperative Research Centre (STCRC), of which a detailed account of findings was reported in McGrath et al. (2005c). The project commenced in January 2004, ran for 12 months and involved seven researchers from four Australian universities. The major objective was to produce a National Information Architecture for the Australian Tourism Industry, and one of the three central project tasks involved a series of interviews conducted with over 40 key stakeholders within the local tourism industry. The objective here was to identify major industry information and information systems gaps and needs.

One of the major outcomes of the interviews was that there appeared to be an urgent need for a survey of small-to-medium tourism enterprises (SMTEs), addressing their take-up of IT and, particularly, the extent to which they were coming online (and utilizing the various online technologies). It was recommended that the survey should address the extent of front-office, back-office and online system take-up; online system functions covered (purely informational or bookings as well); plus levels of data accuracy, currency, robustness and timeliness. This particular recommendation is still under consideration by the STCRC Executive. Fortunately, the interviews did correspond to a large extent with data obtained from the survey about the willingness of Australian SMTEs to adopt a novel and very advanced online technology. Although not ideal, the interviews allow for a comparative follow-up analysis at of least some of the ‘impressionistic’ survey findings.

3.7 Research Limitations and Threats to External Validity

External validity refers to the generalize-ability of a study. For instance, can it be concluded that the results of a particular study (which was done in a specific place, with certain types of people, and at a specific time) might be generalized to another context (for instance, another place, with slightly different people, at a slightly later time)? Where this occurs, a survey may reveal significant results within a sample group but that these results may not be generalized to the population at large. The research was undertaken in full awareness of the threat to external validity. Limitations were therefore noted in the sections of the thesis that they relate to.

The fist limitation was recognized in the aims in section in 1.4, where it was noted that the focus of this study was on information integration *via the Semantic Web*. Thus, while acknowledging the importance of integration theory in areas such as integration methodologies, data mapping algorithms and approaches, data integration in the absence of commonly-accepted international standards, and the implications of information loss during data mappings, a systematic evaluation of all types of possible model differences[90] was not undertaken. A rigorous investigation of this is beyond the scope of the study because the integration investigation here is purely from a Web-based perspective (i.e integration of online tourism information). The issues mentioned above, however, have been identified as a promising area for further research that indeed could build upon the framework established here.

The next limitation is that even though the thesis addressed integration and utilization of tourism information as a whole, the data collection (experiment and survey) focused solely on the accommodation sector of this domain. As noted in Chapter 1, this was done after consideration of the available resources and the large scale of the tourism industry itself. It was decided that it would be more informative from a research perspective to focus the data collection on a specific tourism sector. Accommodation services represent the largest single economic sector of the Australian tourism industry, and as such, an investigation here was considered likely to provide good insight into the tourism industry at large. Consideration was also given to the fact that the technical experiment was

[90] Using for example, the metadata categorization scheme presented by Hsu, C. 1996, 'Enterprise integration and modelling: the meta database approach, kluwer', Norwell, MA.

domain independent because it was analysing the Semantic Web technology itself. Findings for this part of the research are therefore likely to apply to any domain in which the technology was implemented.

It is noted in sub-section 4.2.1.2 of the AcontoWeb SRS that only Websites annotated consistent with the accommodation ontology are included as part of the system. Because of time and resource constraints, cross-portal integration techniques were not used. Such techniques are recognized as important and are described in sub-section 2.2.8, but fall outside the scope of the system. It should be recognized that to this stage AcontoWeb is a prototype developed to demonstrate proof of concept. While the system is fully functional, it has yet been refined to a commercial state, and due to limited availability of resources the knowledge base was only populated with sufficient resort instance data to undertake the technical experiment. The system does not automatically extract Metadata from annotated Webpages and place it into the knowledge base. The system does, however, contain an RDF extractor capable of extracting and viewing RDF markup (consistent with the accommodation ontology) from Web pages. This was considered adequate to demonstrate proof of concept.

Finally, as indicated in sub-section 3.5.1, the survey sample contained some bias. Geographically the respondents' distribution was slightly biased towards the Australian state of Victoria. Specifically, 24.0% of the sample enterprises were based in Victoria compared with an actual figure of 21.4% (ABS 2002, p. 13). More significantly, the number of responses from WA, the ACT and NT were very low (7, 8 and 13 respectively). 31.6% of respondents were hotel/motel operators and 27.2% were B&B/guesthouse operators. Most enterprises (57.7%) were rated at the 4-4.5 Star level, 30.5% were 3-3.5 Star operations and only 4.4% were rated at 2.5 Star or less. This is not representative as, according to the ABS (ABS 2002, p. 18), only 23.3% of Australian accommodation establishments are rated at 4-5 Star, 53.5% are 3 Star establishments and 9.2% are rated at the 1-2 Star level (14.0% are ungraded). Consequently, the results should be treated with caution when applied to the 76.7% of all 4,348 Australian accommodation establishments rated up to the 3 Star quality level or ungraded (ABS 2002, p. 18).

In spite of these limitations, the research is still considered to be reasonably valid externally and capable of meeting the underlying research objectives. Specifically, the technical experiment was based on existing theory supported by the literature and the AcontoWeb system has been developed to a standard sufficiently capable of demonstrating proof of concept. The survey of tourism operators, although not perfect, was also sufficiently revealing to enable interested parties to ascertain whether or not there is a degree of Interest in the adoption of Semantic Web technology within the tourism domain.

3.8 Chapter 3 Summary

The chapter outlined the research philosophy which was based on a systems development approach. The chapter also explained the research phases and how the phases were linked from the initial research questions to the end proposition of a grounded hypothesis. The methodology used to develop and evaluate the AcontoWeb system was presented, along with a detailed explanation of how this development component contributed to meeting the research aims by demonstrating benefits to tourism information integration and utilization through the use of Semantic Web technologies.

The design of the tourism operator survey showed that questions were carefully chosen to assist with answering both the major research question and the formation of a grounded hypothesis. The survey aimed to indicate the degree of interest among tourism operators for the uptake of an advanced new Internet technology. The pilot survey provided valuable feedback about the wording and structure of the survey, and was used to improve the final version that was sent to tourism operators. The secondary data and its relevance were also discussed. This data was obtained from interviews conducted by the STCRC in the year 2004 about the uptake of advanced ICT applications among key tourism stakeholders. It closely matched the information sought in the tourism operator survey, and provided a good follow up analysis of the survey findings.

Finally, the chapter summarized the research limitations and possible threats to external validity. Limitations included the fact that the integration issues investigated were in the context of a Web environment and did not encompass some broader integration issues. It was noted that the data collection focused solely on the accommodation sector of the Australian tourism industry, even though the thesis was investigating issues related to online tourism at large. The AcontoWeb system was described as capable of demonstrating proof of concept, but not yet of a commercial standard. Survey results also contained some bias and need to be treated with caution. Despite these limitations, it was concluded that the research remained external valid, and importantly, was still capable of meeting the underlying research objectives.

4 ACONTOWEB

4.1 Chapter 4 Overview

Chapter 4 presents the AcontoWeb software design and query experiment. The software requirement specification (SRS) describes the system's functional requirements, interfaces, screen designs, Semantic Web components, and includes a usability guide that demonstrates typical system processes. The AcontoWeb query experiment compared the complexity and subsequent ease of information integration of querying the underlying data model of a semantic portal, where information is indexed using a rich domain ontology, to that of a conventional portal where information is indexed to a flat keyword list backed by a relational database. An evaluation model based on Vardi's (1982) prescribed methods was used for query complexity analysis.

4.2 Software Requirement Specification (SRS)

This section contains the software requirement specification (SRS) for the AcontoWeb system. AcontoWeb was designed and modelled using the structured systems analysis and design methodology (SSA&D) described by Donaldson Dewitz (1996). SSA&D focuses on systems functions, where the primary strategy is functional decomposition (in which high level functions are successively decomposed into more detailed functions). The approach emphasizes process modelling, thus the system is viewed from a process-driven perspective (Donaldson Dewitz 1996, p. 12).

4.2.1 SRS Introduction

The SRS commences with a statement of purpose. The scope and an overall description of the system are then outlined. Functional requirements are also specified, including an event list and data flow diagrams, which show how actors are likely to interact with the system and what the associated data flows would be. Interface requirements are specified, including hardware and software interfaces, as well as the system's Semantic Web components. Finally, the screen designs are presented, followed by a usability guide that describes typical system processing from an end-user perspective.

4.2.1.1 Statement of Purpose

The purpose of AcontoWeb is to create a system that provides a tangible benefit over existing accommodation Web portals by allowing tourism customers to search the underlying concepts of a Website, thus producing results that more closely match the customer's needs. This is achieved by using Semantic Web technology to infer knowledge about resorts and seamlessly integrating that knowledge so that it can be used by a tourism customer when searching for suitable accommodation.

4.2.1.2 Scope of the System

The scope of the system is limited to the annotation and querying of Australian accommodation Websites. Only Websites annotated consistent with the accommodation ontology employed are included in the system. Cross-portal integration, as discussed in sub-section 2.2.8, is not supported. Such techniques are recognized as important for the integration of accommodation information, but fall outside the scope of what the system aims to demonstrate. Webpage annotations are conceptually consistent instance data of the accommodation ontology, and are queried by the GUI using a database lookup from a Jena[91] backend knowledge base. Although conceptually consistent with the ontology, the instance data was manually captured from Web pages and inserted into the Jena knowledge base. The process of automatically capturing annotations from the Web had not been completed at the time of writing. The annotation tool, however, does contain an RDF extractor. This demonstrates that Webpage annotations are readily extractable from within Web pages, and is therefore sufficient to demonstrate proof of concept.

4.2.1.3 Overall Description

The AcontoWeb architecture (see Figure 40) is designed to support convenient annotation and intelligent querying of Semantic Web resources. Annotation software is used by a Web site owner to generate RDF markup describing the content of their Web site. The RDF markup is essentially instance data that conforms to an OWL accommodation ontology, and is imbedded by an annotation tool into readily extractable comment tags contained in a HTML file. Query functions are facilitated by a Jena based

[91] http://jena.sourceforge.net/

SPARQL query engine that uses a Pellet[92] reasoner and the OWL ontology to infer knowledge about the accommodation domain. The query facility is accessed remotely via a Web-based GUI and provides the end-user with a number of search options. Once a query is submitted, a list of matching results is displayed to the end-user. The annotation tool contains an FTP client to allow a Website owner (or, perhaps more likely, a contracted IT professional) to download their Website, annotate it then, upload it back to the host server. The annotation tool also contains an RDF extractor to allow the Website owner to readily extract and view RDF metadata imbedded in a Webpage.

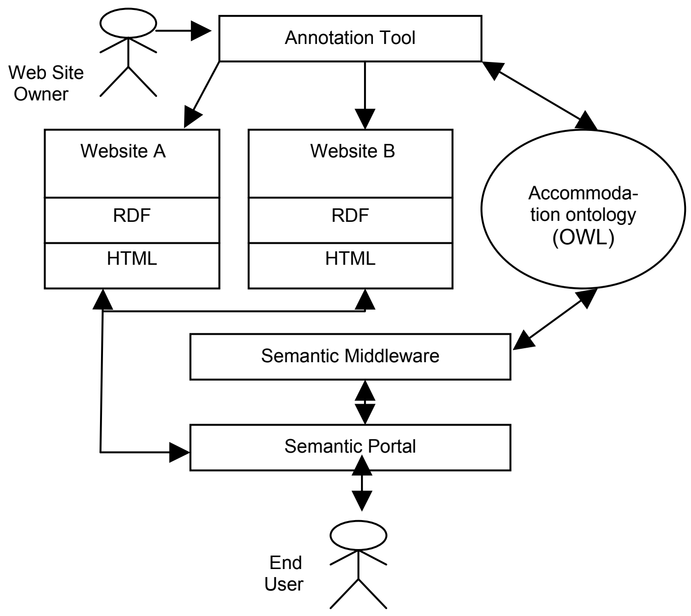


Figure 40: AcontoWeb architecture.

To allow for a precise and measured comparison of queries between a conventional portal and a semantic portal, the data structure of the RACV accommodation[93] (AAA tourism) portal was captured and remodelled using the relational modelling theory of Codd (1970). The data was then physically replicated in an Access database, as well as an OWL ontology (see Appendices D and E) by following the Methontology framework (see Appendix A). Two extra fields that were not part of RACV portal were added to both the database and the ontology. 'Destination Attractions' were added to demonstrate that with the use of transitive properties, attractions can automatically be inferred to be associated with a particular resort based on the resort's location. 'Destination

92 http://www.mindswap.org/2003/pellet/

93 http://www.accommodationguide.com.au/searchgateway.asp?sit=2&aid=1

Classification' was added to demonstrate that by using OWL class restrictions, a location could automatically be inferred as a particular type of location based on the attractions and types of accommodation that are in the vicinity.

The criteria used to specify location types was based upon Tourism Victoria's[94] 2004 marketing segment classifications and was obtained from Tourism Research Australia[95] and Roy Morgan Research[96], and is provided in Appendix B. Locations were classified according to five major market segment types. Research indicates for example, that the Backpacker market segment includes tourists that prefer the attractions of nightclubs, pubs, aquariums, zoos, wildlife-parks, national parks and state forests, museums, art galleries, and like to stay at backpacker hostels. These location features were modelled into the accommodation ontology using class restrictions, so that resorts in locations containing these features were automatically assigned a Backpacker destination classification in the inferred ontology. Market segments included in the accommodation domain were:

- Adventure tourism.
- Backpacker tourism.
- Caravan and camping tourism.
- Cultural tourism.
- Food and wine tourism.

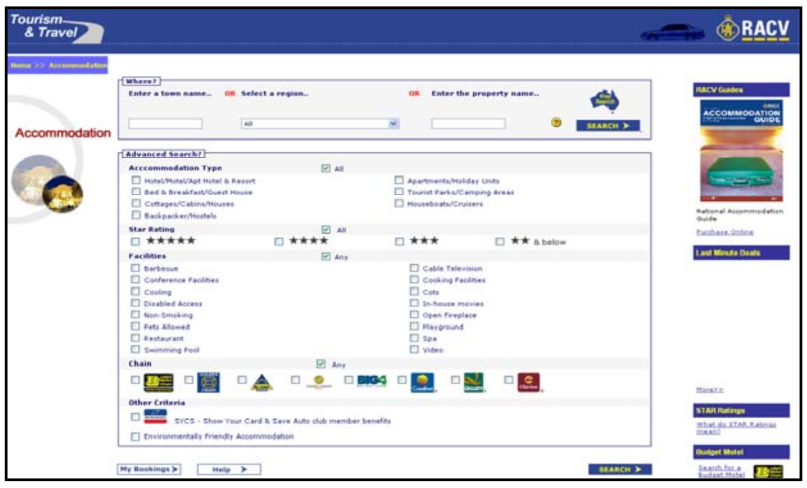


Figure 41: RACV accommodation portal.

[94] http://www.tourismvictoria.com.au/index.php

[95] Source: National and International Visitor Surveys, year ending December 2004, Tourism Research Australia

[96] Holiday Tracking Survey, year ending December 2004, Roy Morgan Research

4.2.1.4 Product Perspective

The AcontoWeb portal contains information about holiday units, flats, houseboats, cottages, hotels, motels, guest houses, chalets, apartments, and self-catering accommodation, each listing all the facilities and local attractions available at resorts. All accommodation has previously been investigated by RACV field-workers and has been given an official star rating (up to 5 stars for the most elaborate and luxurious stays). Many of the businesses use their star-rating for promotional purposes.

4.2.1.5 Development Team

Coding of the AcontoWeb annotation tool was done by the researcher. To improve usability, the software was designed considering the requirements of Web site owners as established by the survey of tourism operators. Coding for the query component of the semantic portal required specialized Java programming skills. Funding was therefore sought (and subsequently obtained) from the School of Information Systems at Victoria University to outsource this part of the development to a specialist Java programmer. The project became part of a university funded collaborative research scheme, with the aim of refining AcontoWeb to a commercial standard. Designs for AcontoWeb were validated with publication of Abrahams & Dai (2005a) and Dai & Abrahams (2005) in the proceedings of the 2005 International Joint Conference on Web Intelligence and Intelligent Agent Technology held at Compiegne University France.

4.2.2 Functional Requirements

This sub-section details the functional requirements of the system.

4.2.2.1 Event List

The event list shows the events that are initiated by user interaction with the system and the resulting data flows that occur.

Event	Data Flow
1. Accommodation provider downloads Website	Host Details/Website
2. Accommodation provider selects ontology	Ontology Request/Ontology
3. Accommodation provider creates new Website annotation	Accommodation Details/Website
4. Accommodation provider edits existing Website annotation	Amended Accommodation details/ Website
5. Accommodation provider deletes Website annotation	Withdrawn Accommodation Details/ Website
6. Accommodation provider uploads Website	Host Details/Website
7. Accommodation provider extracts RDF metadata	URL/RDF metadata
8. Customer searches for accommodation online	Search Criteria/Search Results
9. Customer requests new search	New Search Request/New Search Screen
10. Customer selects accommodation	Website Selection/Annotated Website

Table 5: Event list.

4.2.2.2 Data Flow Diagrams

<u>Context Diagram</u>

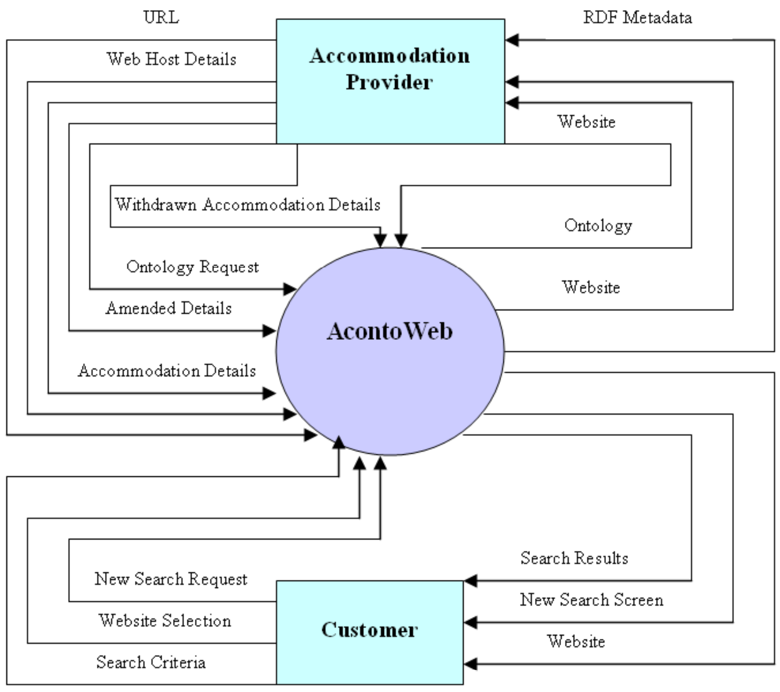


Figure 42: Context diagram.

Level 0 - Subsystem Data Flow Diagrams

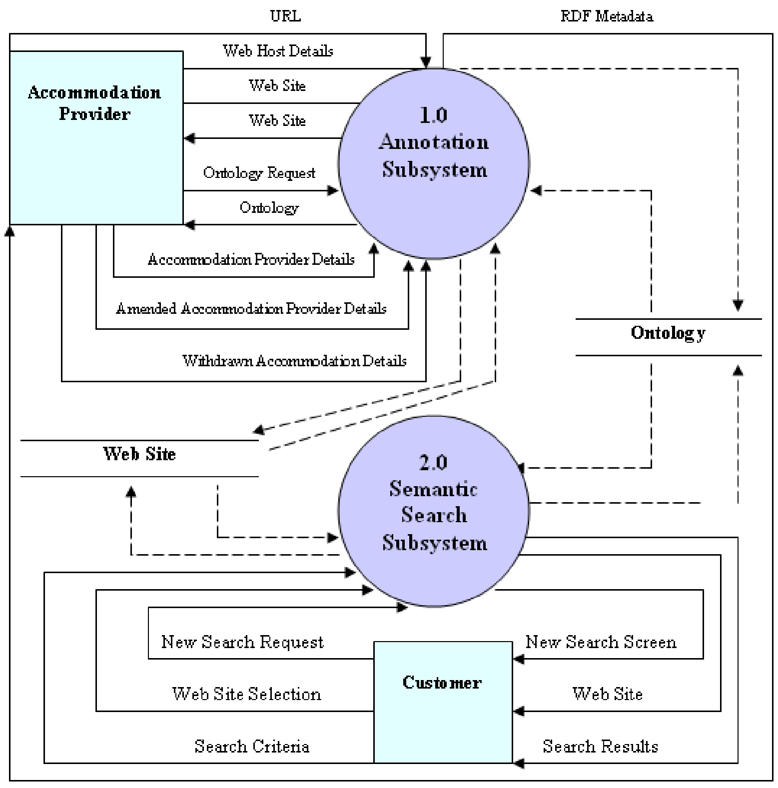


Figure 43: Subsystems data flow diagram.

Level 1 - Component Data Flow Diagrams

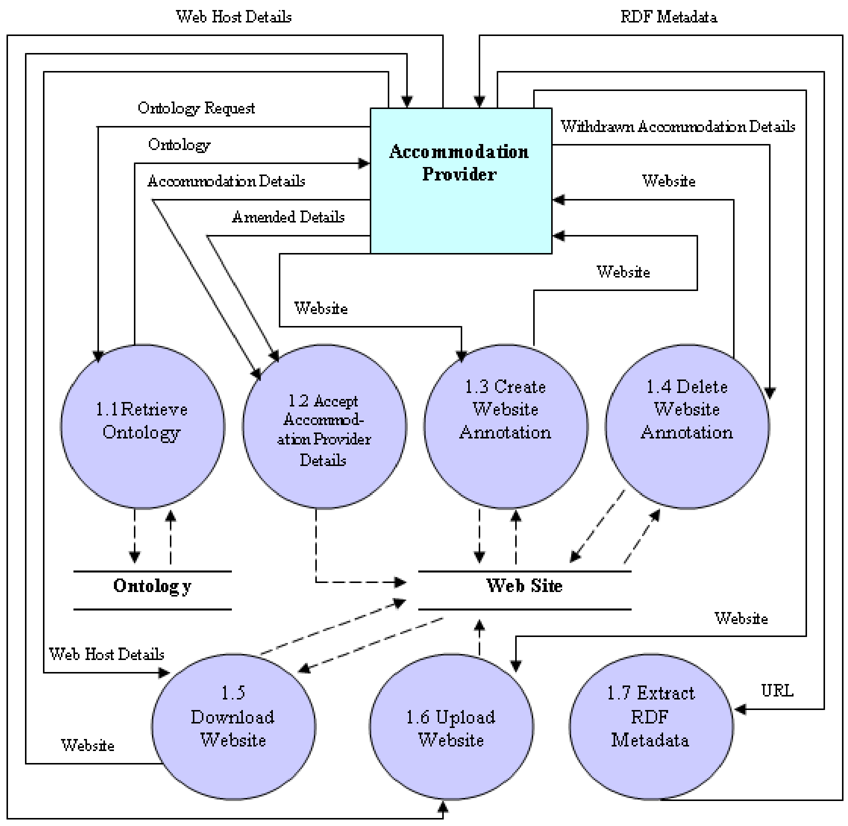


Figure 44: Annotation subsystem level 1 data flow diagram.

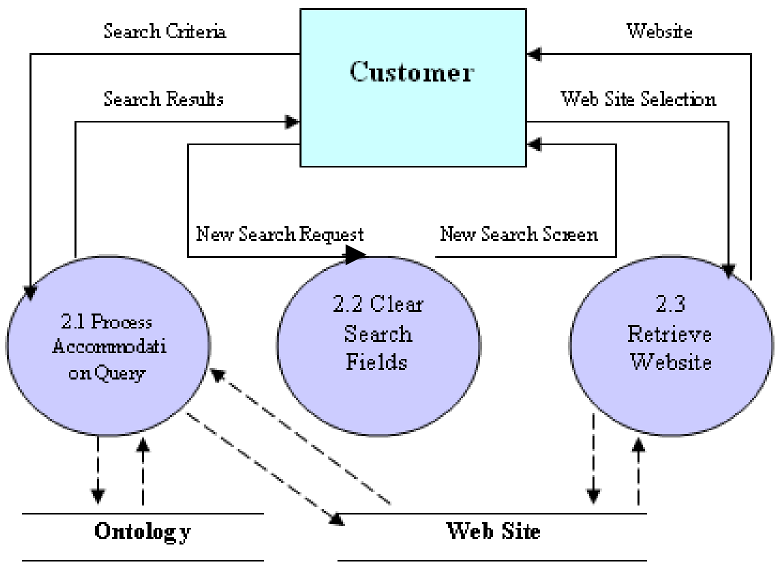


Figure 45: Semantic Search subsystem level 1 data flow diagram.

4.2.2.3 Interface Requirements

The Interfaces for AcontoWeb combine interactive computing software and Java compatible Web browsers, as well as the availability of workstations connected to the Internet to facilitate wide-scale distribution and access to the annotation and portal components.

User Interfaces

User interfaces for the Annotation subsystem are Windows forms. For the Semantic Search subsystem, user interfaces are in the form of a Java-capable browser.

Hardware Interface

A work station connected to the Internet plus mouse and mouse pad.

Software Interface

A Java-capable Web browser with access to the Internet, the Java Development Kit (JDK) from Sun Microsystems or Integrated Development environment (IDE), and a text editor for preparing HTML files.

4.2.2.4 Semantic Web Components

The Semantic Web application resides on a server computer and has three major components:

Jena Components

- Jena middleware application - Jena communicates with the custom servlet and the relational database. Jena is responsible for managing the reasoning system, the queries from the custom servlet and communicating with the relational database. The Pellet[97] reasoner was used as a reasoning system.
- Relational database - the database is MySQL and holds the Accommodation ontology which contains tourism data and rules. The database is managed totally by Jena. The ER diagram and data dictionary for this database are therefore not included in the SRS.

Server Component

- A Server computer running the Tomcat servlet container. Tomcat is listening for http requests on for example http://www.SomeTourismSite.com.au/8080/. Tomcat is embedded with a custom Java servlet. Tomcat and the custom servlet are responsible for picking up the choices from the Web page presented to the user. The servlet also has the job of displaying and returning the query results to the user. Figure 46 is a diagram of the proposed server architecture:

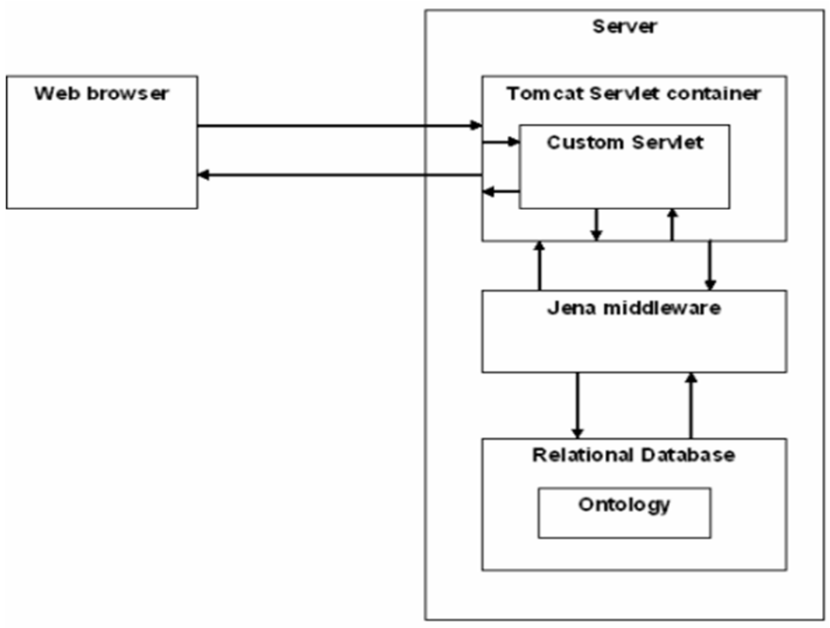


Figure 46: Server architecture.

[97] http://www.mindswap.org/2003/pellet/

4.2.2.5 Screen Designs

<u>Annotation Subsystem Screen Designs</u>

1. Screen Hierarchy Chart

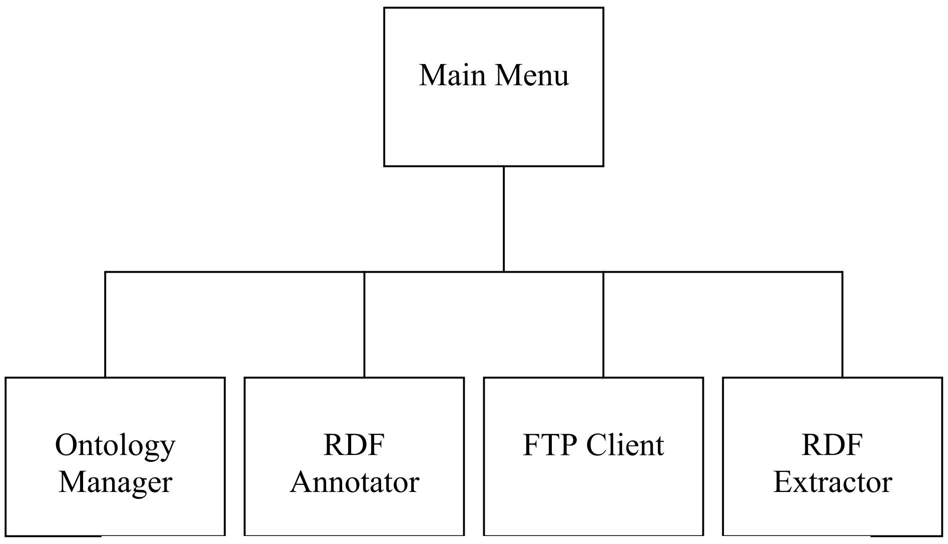


Figure 47: Annotation subsystem screen hierarchy.

2. Screens Layout

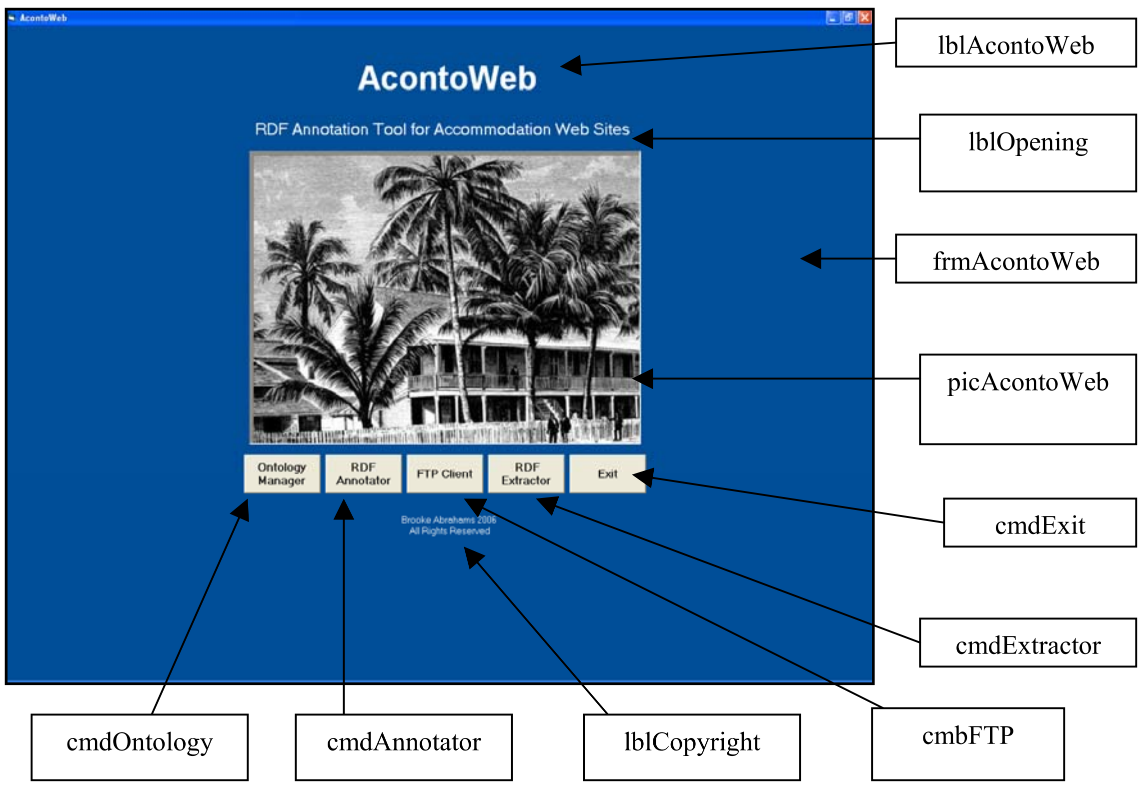


Figure 48: Main Menu screen layout.

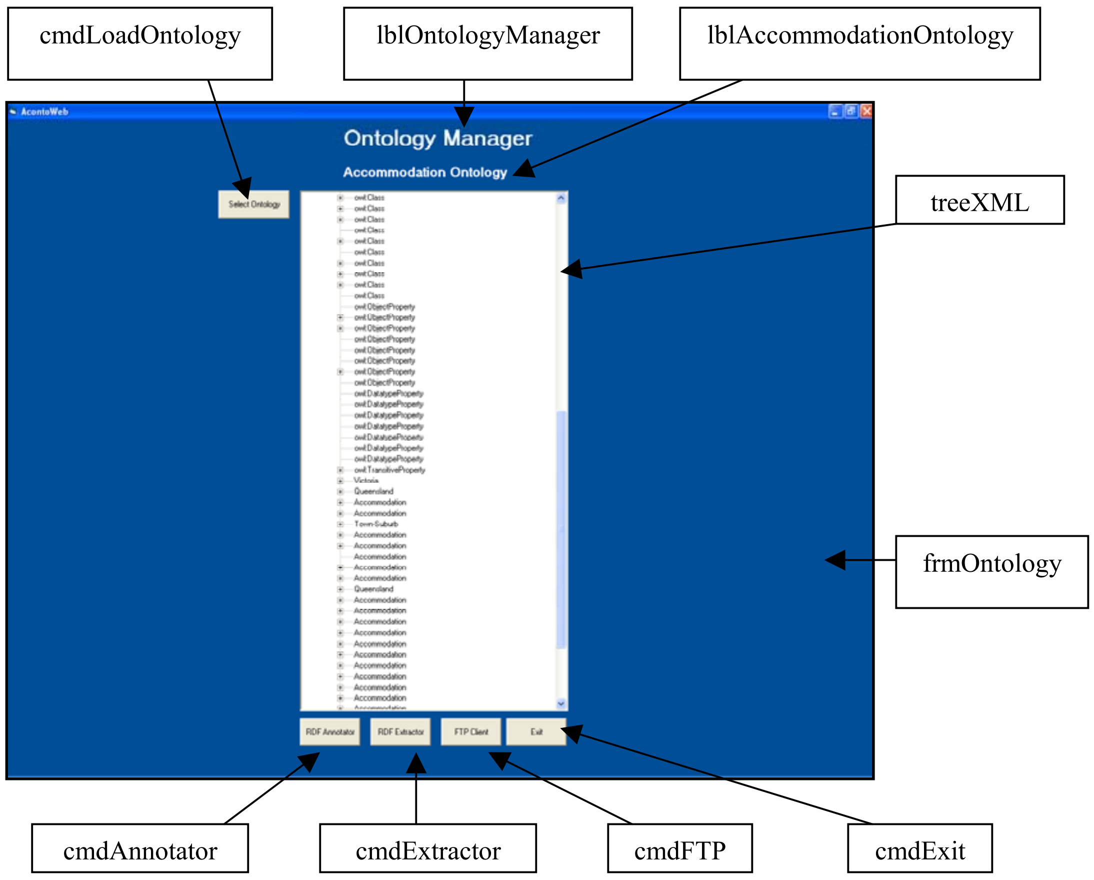


Figure 49: Ontology Manager screen layout.

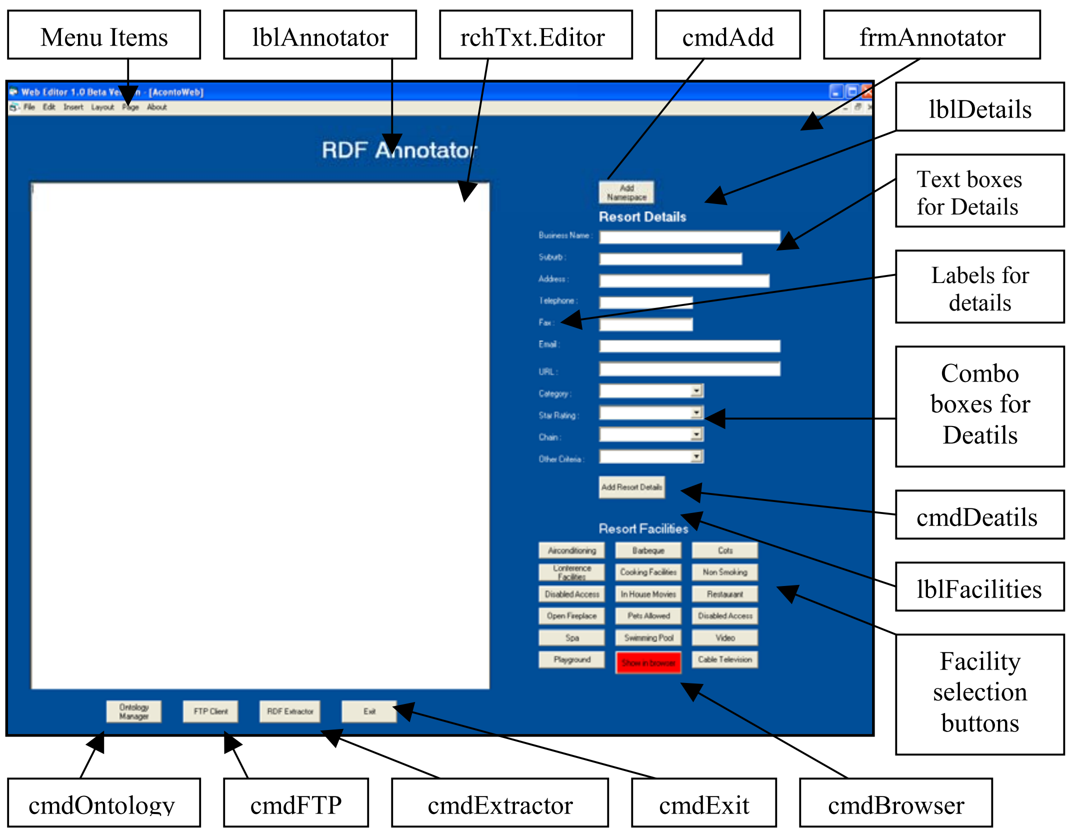


Figure 50: RDF Annotator screen layout.

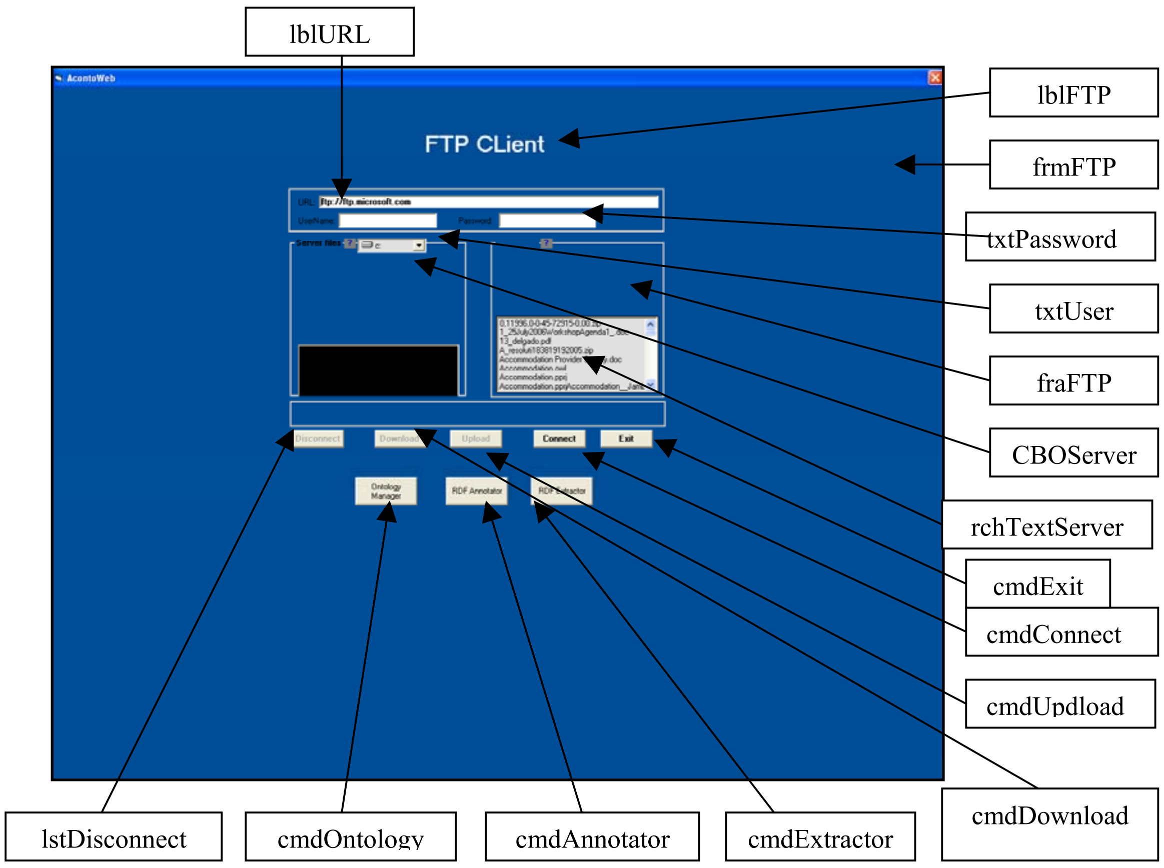


Figure 51: FTP Client screen layout.

Semantic Search Subsystem Screen Designs

1. Screens Hierarchy

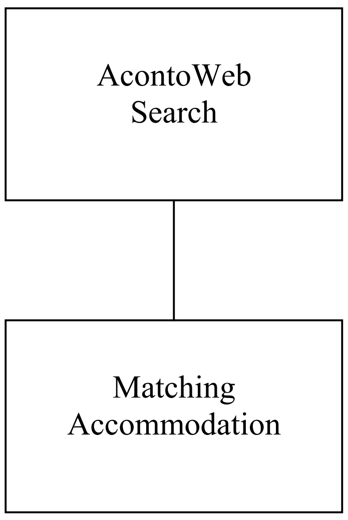


Figure 52: Semantic Search subsystem screen hierarchy.

2. Screens Layout

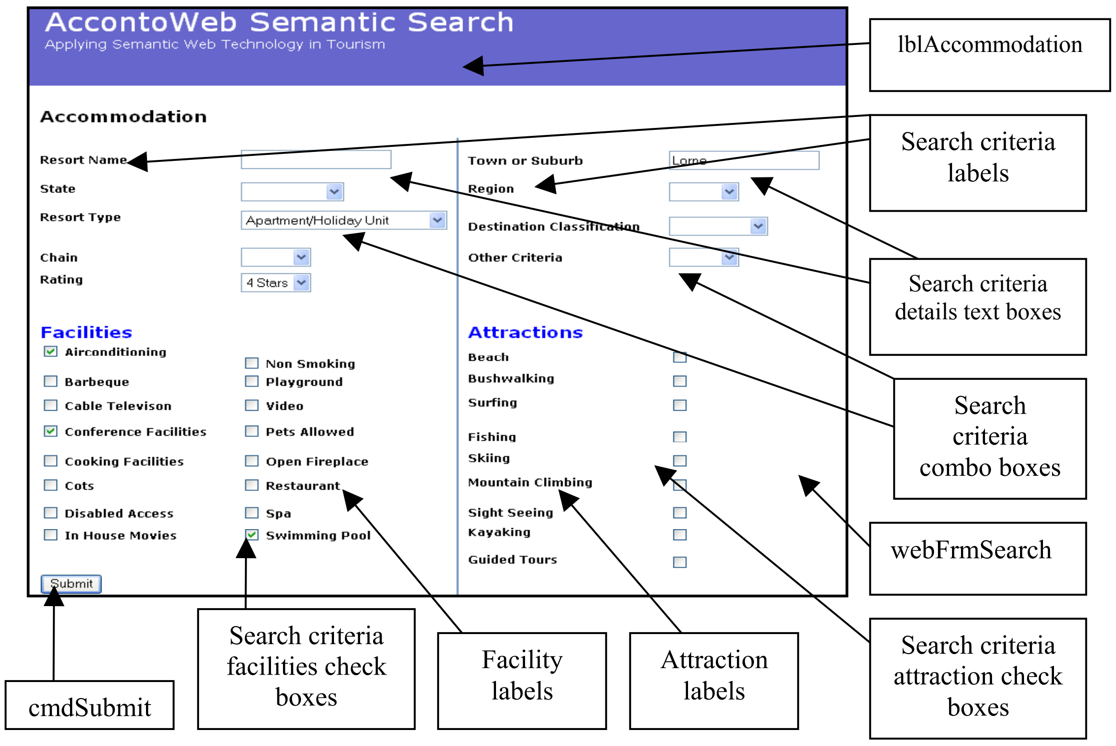


Figure 53: Semantic Search screen layout.

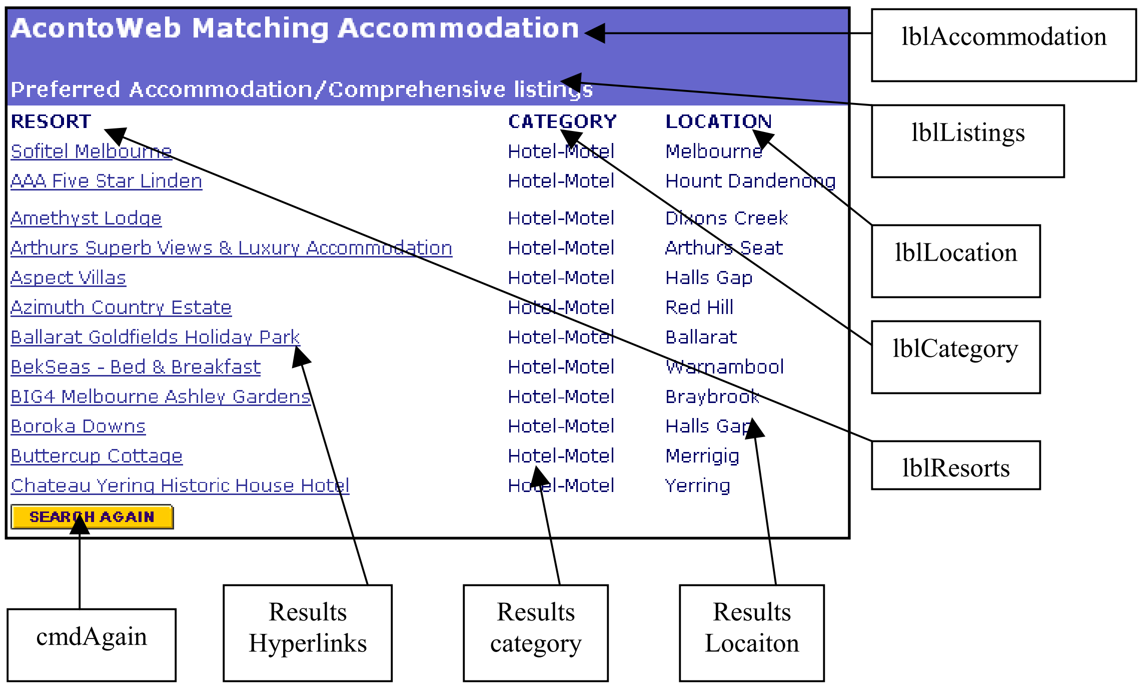


Figure 54: Results screen layout.

4.2.2.6 System Usability

This sub-section presents the typical course of the system from an end user perspective.

Annotation Subsystem Typical System Processing

1. Website owner opens AcontoWeb annotation tool and selects **FTP Client** from Main menu.

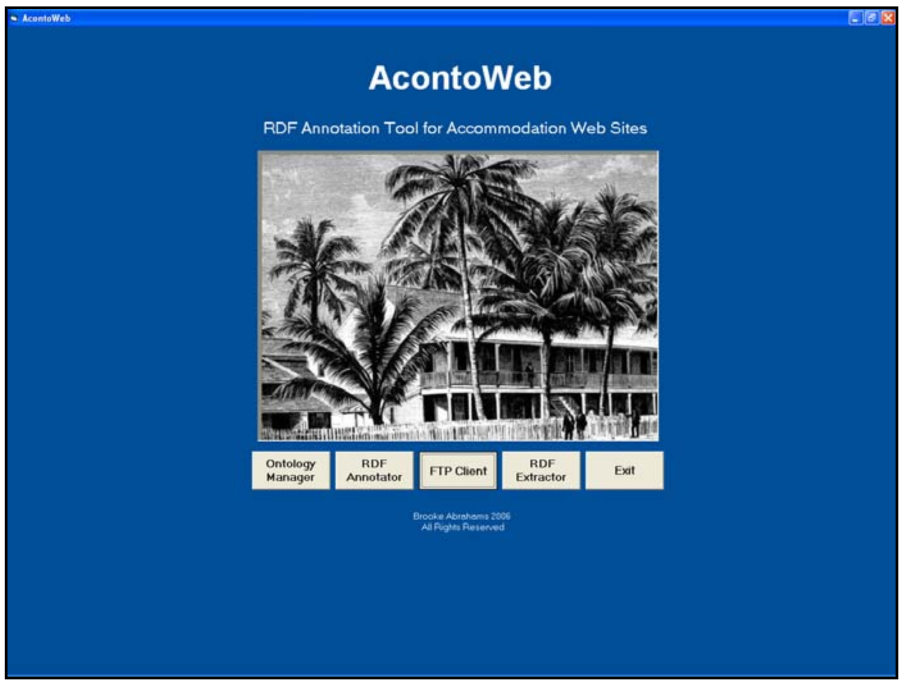


Figure 55: AcontoWeb Main Menu.

2. Website owner enters Web host **URL**, Web space **Username** and **Password**, then downloads their Web site from host server to local C drive.

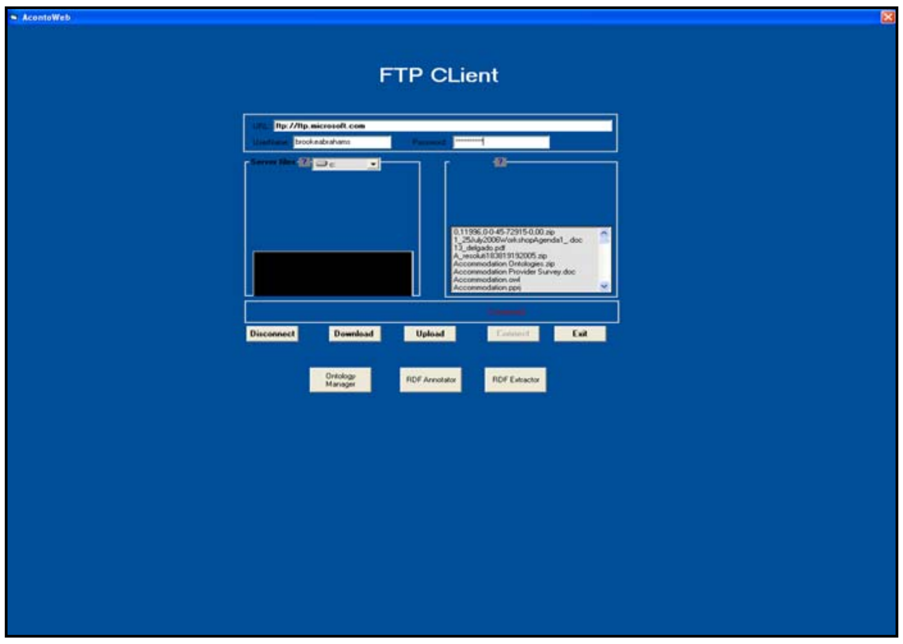


Figure 56: FTP Client.

3. Website owner opens **Ontology Manager**, chooses **Select Ontology** then opens an Accommodation ontology from file.

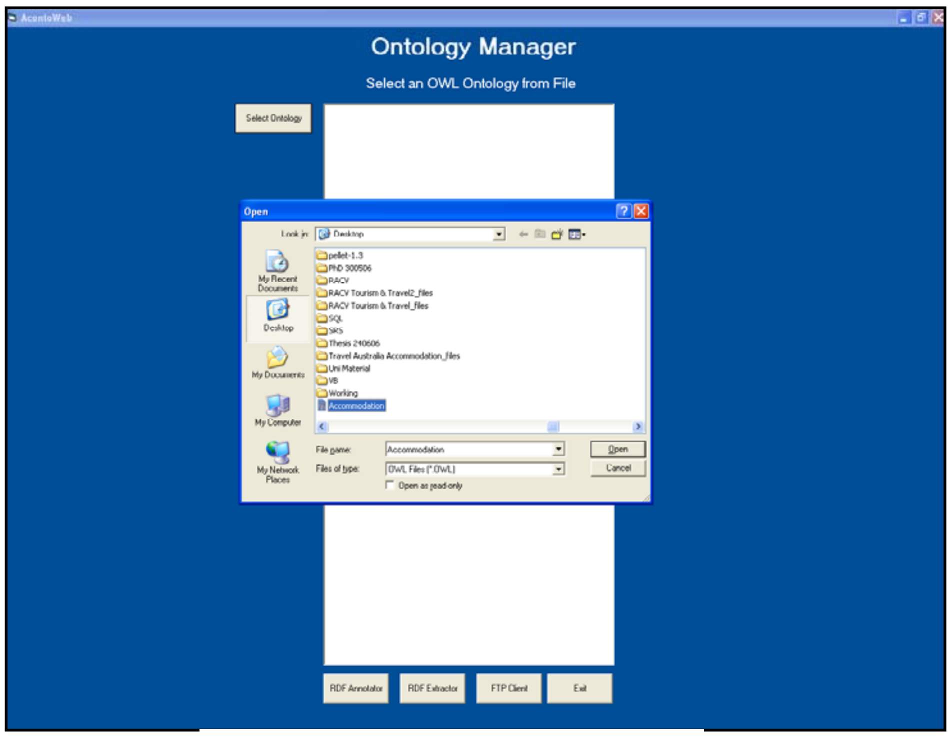


Figure 57: Selecting ontology.

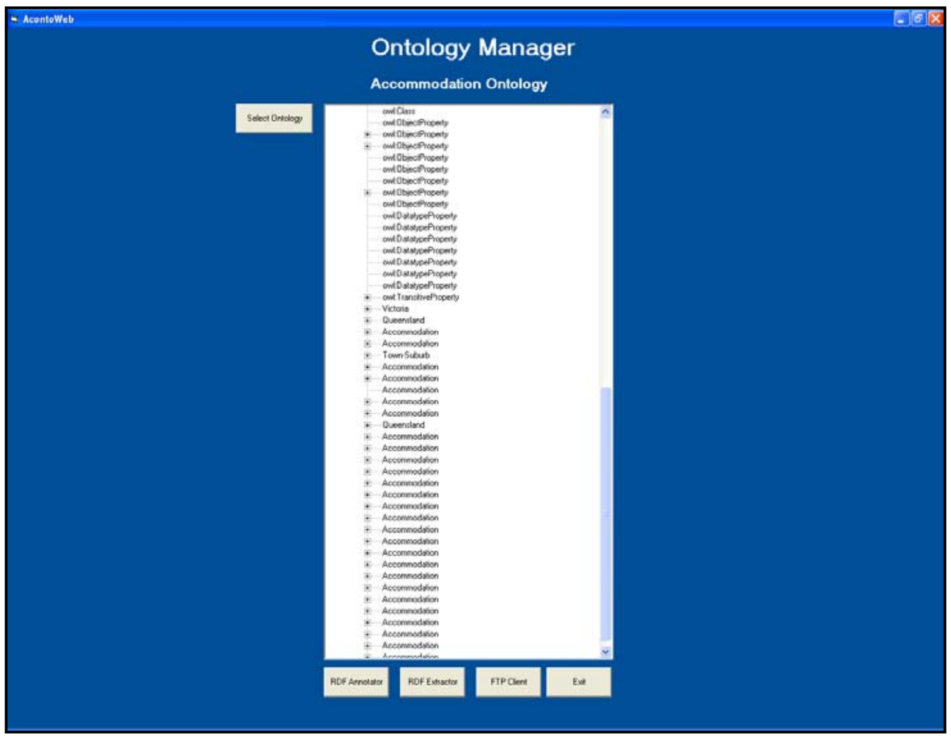


Figure 58: Ontology view.

Website owner opens the **RDF Annotator** then selects downloaded Webpage from C drive.

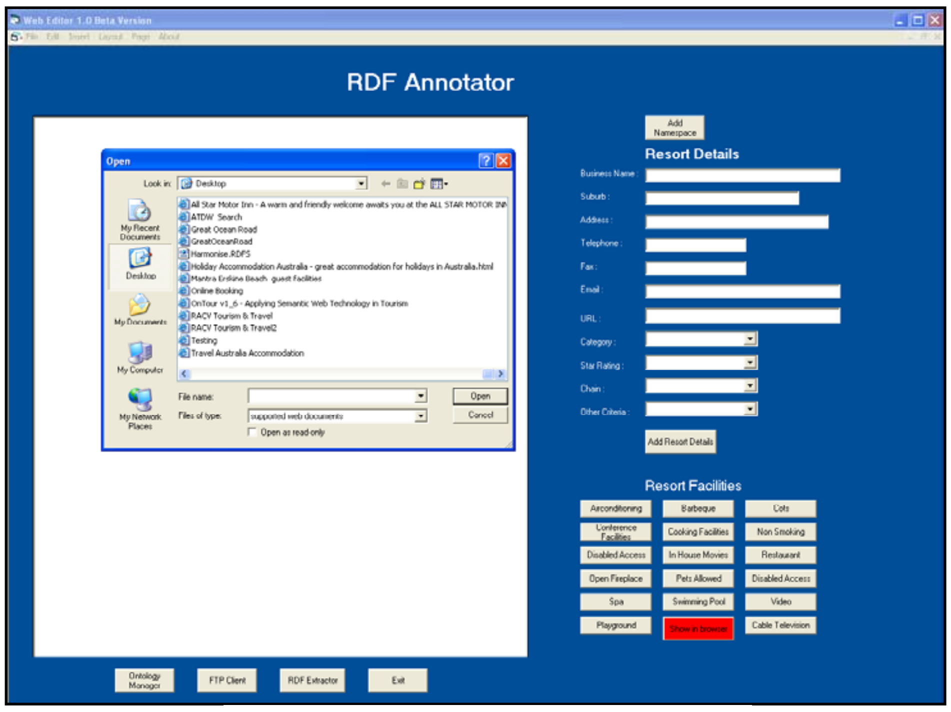


Figure 59: Downloading Webpage.

3. Website owner selects **Add Namespace**, enters **Resorts Details**, and selects **Resort Facilities**. Annotation is imbedded in Webpage. Website owner saves annotated Webpage.

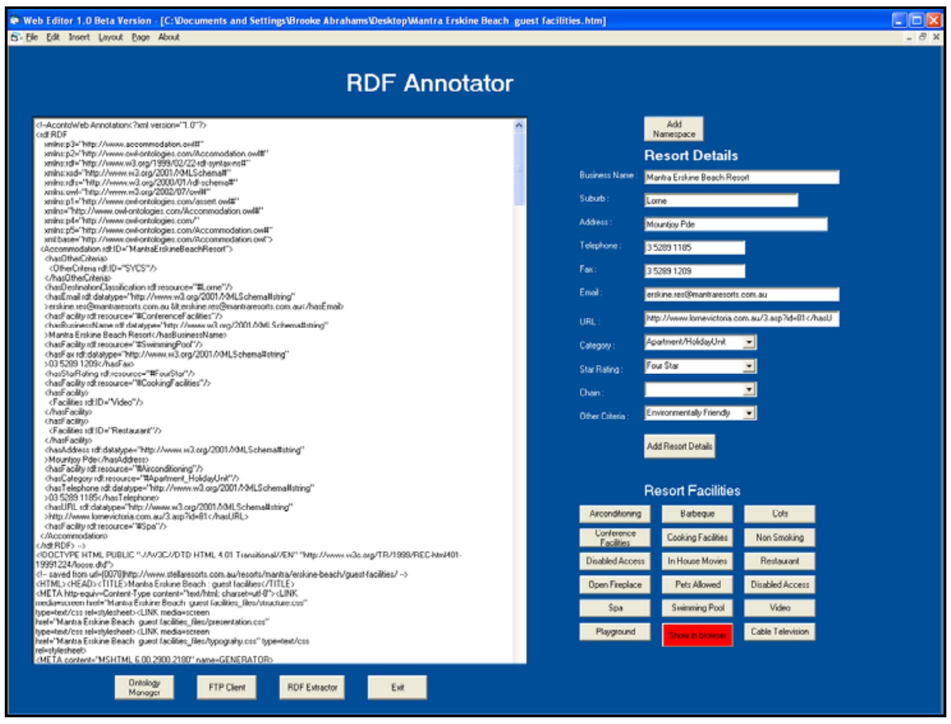


Figure 60: RDF Annotator.

4. Website owner presses **Show in browser** button to display Webpage.

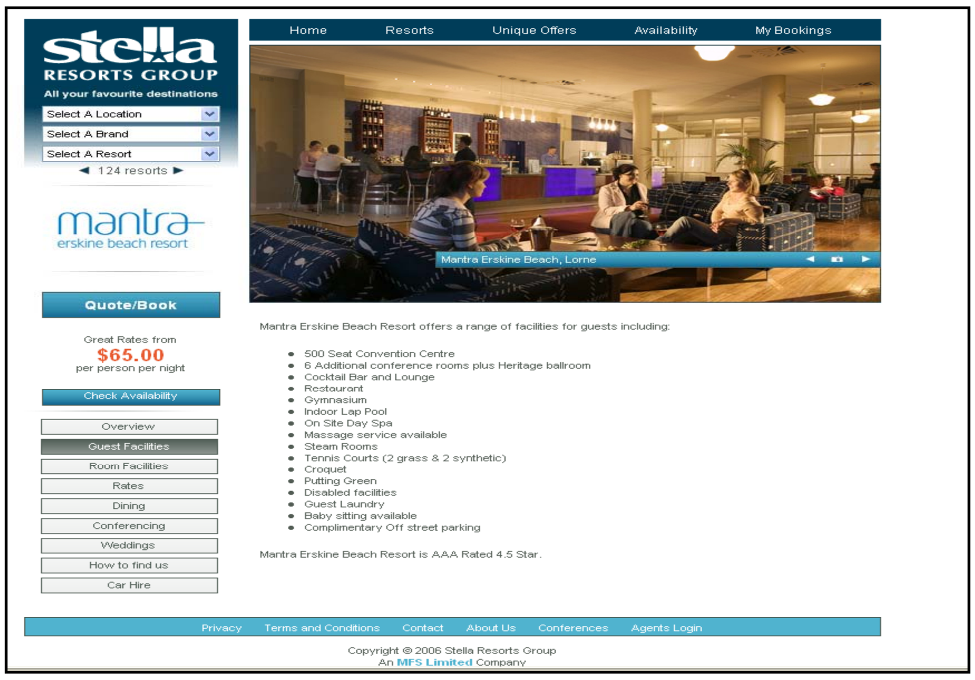


Figure 61: Webpage view.

5. Website owner then returns to **FTP Client** and uploads Webpage back to host server.

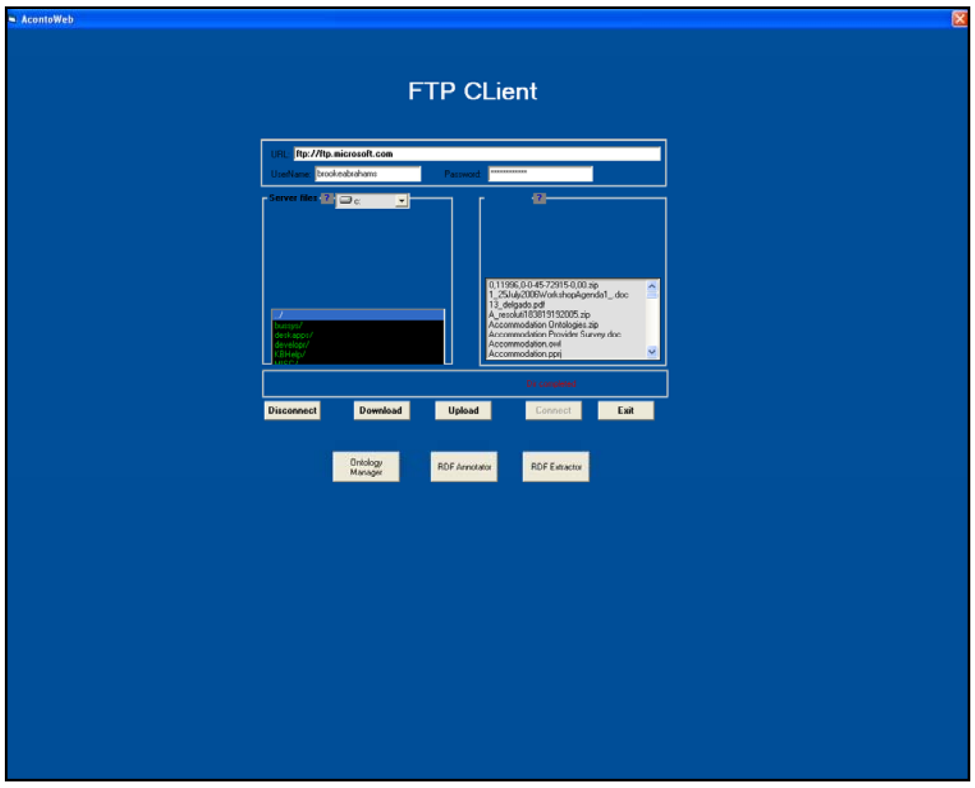


Figure 62: Uploading Webpage.

7. Website owner opens **RDF extractor**, enters the **URL** of an RDF annotated Webpage and presses **Navigate**. HTML source code, RDF metadata and Web page are displayed.

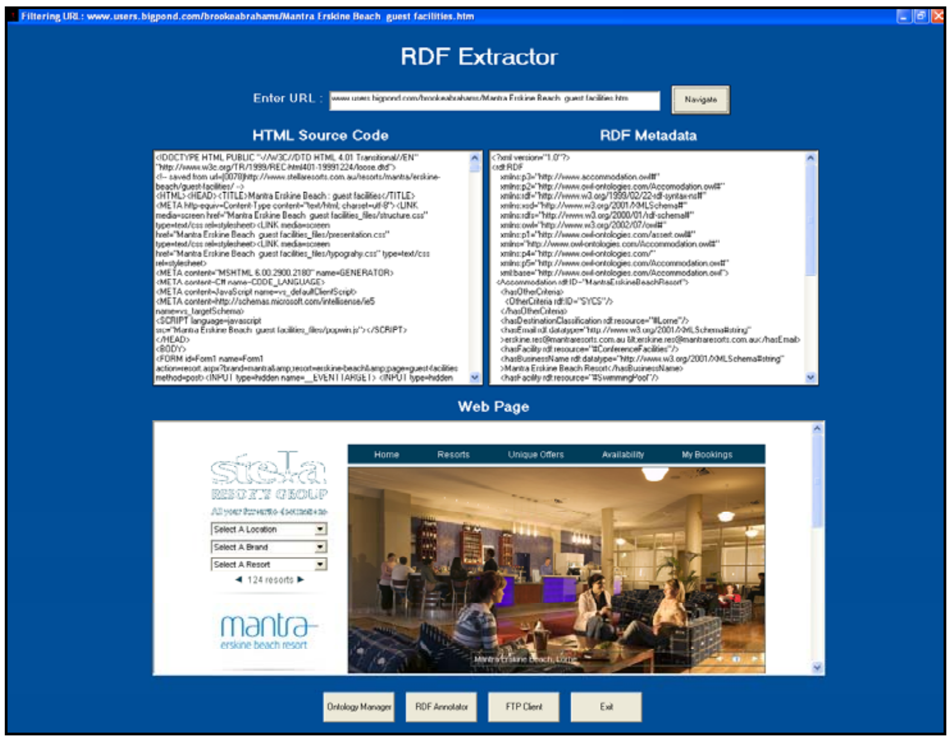


Figure 63: Extracting RDF metadata.

Semantic Search Subsystem Typical System Processing

1. Tourism customer accesses AcontoWeb portal, selects preferred accommodation options then presses **Submit**.

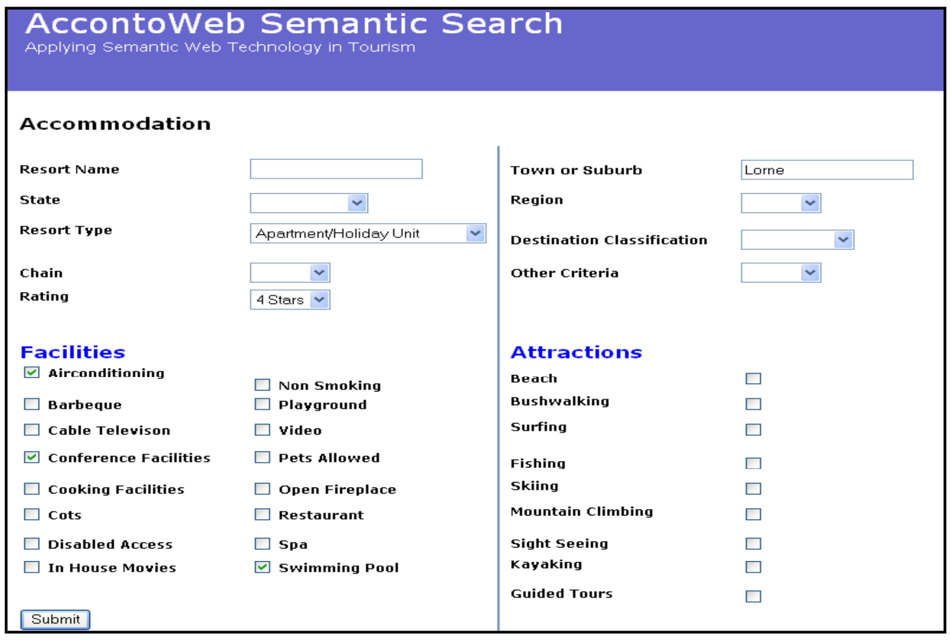


Figure 64: Performing accommodation search.

2. A list of matching accommodation is displayed to the tourism customer.

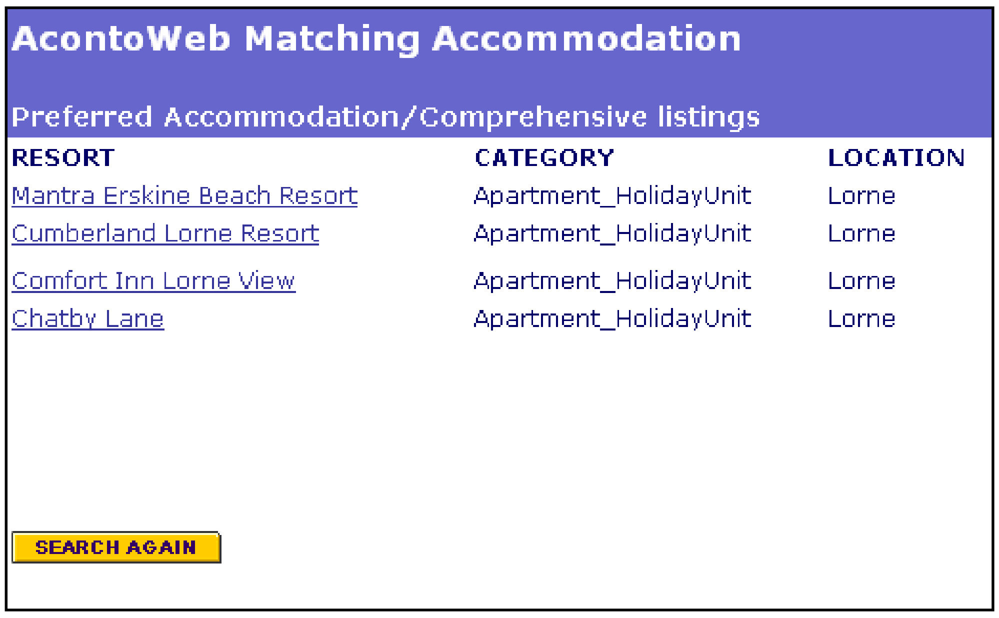


Figure 65: Accommodation search results.

4.3 AcontoWeb Experiment

The AcontoWeb experiment compared the complexity of querying the data model of a semantic portal, where information is indexed using a rich domain ontology, to that of a conventional portal that uses a flat keyword list backed by a relational database. Four sample queries represented in clausal form were compared in an ordered hierarchy similar to that established by Jansen (2000) in his study on the effect of query complexity on Web searching results. Each query was first tested in an Access database (using SQL) that replicated the data structure of the RACV accommodation portal. The queries were then transformed to an ontology-consistent form using the transformation method of Horrocks and Tessaris (2000) which was demonstrated in sub-section 3.4.2. The queries were shortened by using inferred knowledge in the ontology model and tested in Protégé[98] using the Racer[99] reasoner, then transformed back to a clausal form so that they could be compared for complexity to the original version implemented in the relational database environment. An evaluation model based on Vardi's (1982) prescribed methods query complexity analysis was used for the evaluation.

[98] http://protege.stanford.edu/

[99] http://www.sts.tu-harburg.de/~r.f.moeller/racer/

All four sample queries were also implemented in AcontoWeb to demonstrate that the semantic portal works. The AcontoWeb user input form (the GUI) and results page for each query is shown in Appendix H. The SQL and SPARQL representations are shown in Appendixes I and J. It is important to remember that it was the underlying logic (represented in clausal form) of querying the two data models that was being compared in the experiment, rather than any specific query representational language.

Query 1

Retrieve all Apartment-Holiday units in Lorne with a four star rating and swimming pool, airconditioning and conference facilities.

Query 1 Assumptions

- There are 4 four star Apartment-Holiday Units in Lorne with swimming pool, airconditioning and conference facilities.

Conjunctive Query-1A Represented in Clausal Form

$$Q(X) \leftarrow \mathit{Accommodation}(X) \wedge \mathit{hasAccommodationDestination}(X,A) \wedge \mathit{hasStarRating}\ (X,B) \wedge \mathit{hasCategory}(X,C) \wedge \mathit{hasAccommodationFacility}(X,D) \wedge \mathit{hasAccommodationFacility}(X,E) \wedge \mathit{hasAccommodationFacility}(X,F) \wedge A = \mathit{Lorne} \wedge B = \mathit{FourStar} \wedge C = \mathit{Apartment\text{-}HolidayUnit} \wedge D = \mathit{SwimmingPool} \wedge E = \mathit{Airconditioning} \wedge F = \mathit{ConferenceFacilities}$$

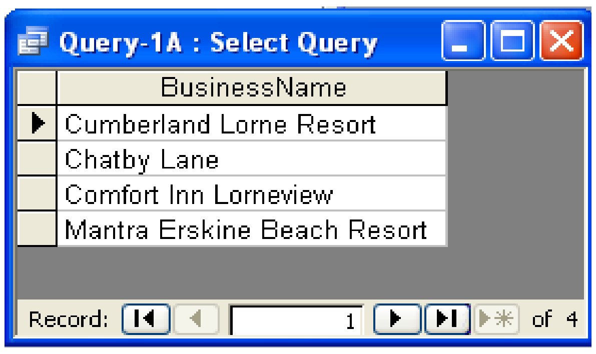


Figure 66: Conjunctive Query-1A results in Access.

<u>Conjunctive Query-1B Represented in Clausal Form</u>

$Q(X) \leftarrow Apartment\text{-}HolidayUnit(X) \wedge hasAccommodationDestination(X,A) \wedge hasStarRating(X,B) \wedge hasCategory(X,C) \wedge hasAccommodationFacility(X,D) \wedge hasAccommodationFacility(X,E) \wedge hasAccommodationFacility(X,F) \wedge A = Lorne \wedge B = FourStar \wedge D = SwimmingPool \wedge E = Airconditioning \wedge F = ConferenceFacilities$

<u>Conjunctive Query -1B Represented as an Ontology Concept</u>

$$(Apartment\text{-}HolidayUnit \sqcap (\exists\ hasAccommodationDestination.\ \{Lorne\}) \sqcap$$
$$(\exists\ hasStarRating.\{FourStar\}) \sqcap$$
$$(\exists\ hasAccommodationFacility.\{SwimmingPool\}) \sqcap$$
$$(\exists\ hasAccommodationFacility.\{Airconditioning\}) \sqcap$$
$$(\exists\ hasAccommodationFacility.\{ConferenceFacilities\}))$$

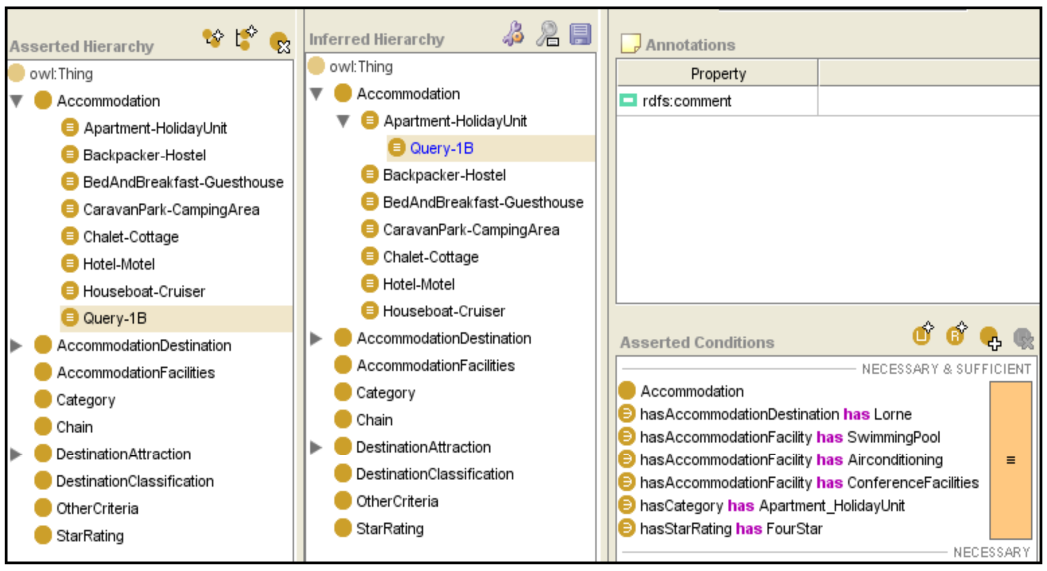


Figure 67: Conjunctive Query-1B as an ontology concept.

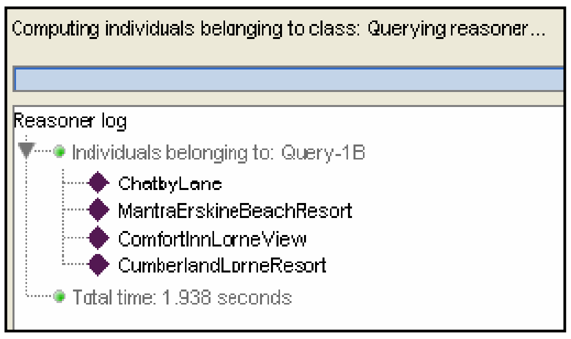


Figure 68: Conjunctive Query-1B results in Racer.

Query 1 Complexity Evaluation

Conjunctive Query- 1A
$Q(X) \leftarrow Accommodation(X) \wedge hasAccommodationDestination(X,A) \wedge hasStarRating(X,B) \wedge hasCategory(X,C) \wedge hasAccommodationFacility(X,D) \wedge hasAccommodationFacility(X,E) \wedge hasFacility(X,F) \wedge A = Lorne \wedge B = FourStar \wedge C = Apartment\text{-}HolidayUnit \wedge D = SwimmingPool \wedge E = Airconditioning \wedge F = ConferenceFacilities$
Conjunctive Query- 1B
$Q(X) \leftarrow Apartment\text{-}HolidayUnit(X) \wedge hasAccommodationDestination(X,A) \wedge hasStarRating(X,B) \wedge hasFacility(X,D) \wedge hasAccommodationFacility(X,E) \wedge hasAccommodationFacility(X,F) \wedge A = Lorne \wedge B = FourStar \wedge D = SwimmingPool \wedge E = Airconditioning \wedge F = ConferenceFacilities$

Figure 69: Versions of Query 1 to be compared.

Measure	Results
Conjunctive Query-1A	
X^s	4
n^s	4
Query terms	8
Equivalent equi joins	0
Ordinal complexity ranking	1
Conjunctive Query-1B	
X^s	4
n^s	4
Query terms	7
Equivalent equi joins	0
Ordinal complexity ranking	1

Table 6: Query evaluation model applied to Query 1.

Query 1 Evaluation Summary

For Query 1, the use of OWL semantics and a reasoner made no difference to the query's complexity when analysed in accordance with Vardi's (1982) theorem. The two conjunctive queries contained just one variable, which in both cases was *X* (i.e. *Accommodation* for *Query-1A*, and *Apartment-HolidayUnits* for *Query-1B*). For both queries, *X* had a value of 4. The only advantage gained by using an ontology compared to a relational database in this case, was that necessary and sufficient class restrictions (asserted conditions, see Figure 70) in the ontology meant that any resort with an *Apartment-HolidayUnit* classification was automatically reclassified as an instance of the *Apartment-HolidayUnit* class, which is a lower-level more specific sub-class of the *Accommodation* class. The result of this was that *Query-1B* was searching directly for instances of the *Apartment-HolidayUnit* class, rather than searching for instances of the class *Accommodation* with the *hasClassification* property value *of Apartment-HolidayUnit*. Thus, the number of query terms was able to be reduced by one from 8 in conjunctive *Query-1A* to 7 in conjunctive *Query-1B*. This demonstrates that *Query 1* was slightly easier to formulate in the Semantic Web environment than the relational database environment.

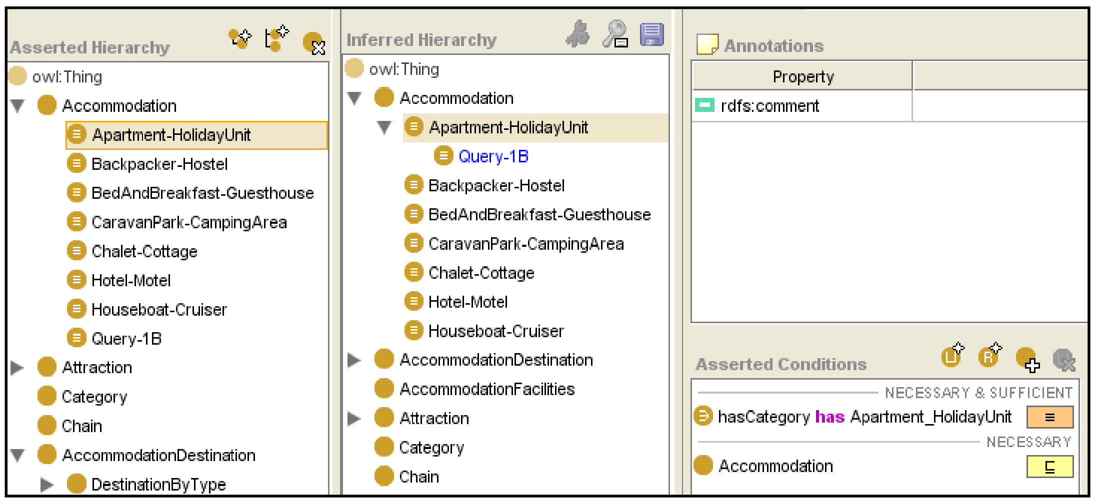


Figure 70: Apartment-Holiday unit classification.

Seamless integration occurred between the Mantra Beach and the Cumberland resorts who each use different naming conventions to describe resort conference facilities (see Figure 71). Because the Websites of both hotels are annotated with the RDF instance *ConferenceFacilities* from the Accommodation ontology, a search for the underlying concept *ConferenceFacilities* returned results for both resorts even though Mantra Beach resort use the term “Convention Centre” while the Cumberland resort uses the term “Conference Centre”. This demonstrates that in a Semantic Web environment, searching the underlying concepts of a Webpage can automatically (to some extent at least) integrate information. The Web page annotation for the Mantra Beach and Cumberland resorts is provided in Appendix K.

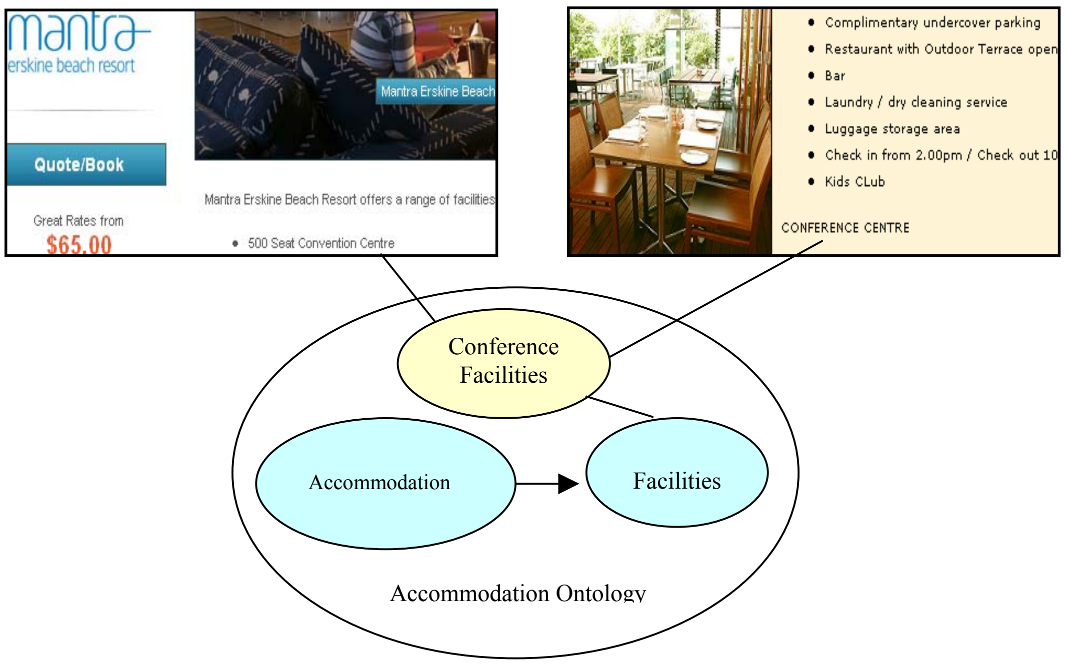


Figure 71: Seamless information integration.

Query 2

Retrieve all Bed and Breakfast_Guesthouses in Victoria with a four star rating, open fire facility, and destination attractions of surfing and bushwalking.

Query 2 Assumptions

- There are 100 four star apartment-holiday units in Victoria with an open fire facility.
- 20 destinations in Victoria have the attraction surfing.
- 20 destinations in Victoria have the attraction bushwalking.
- There are 5 four star apartment-holiday units in Victoria with an open fire facility in a destination that has the attractions of surfing and bushwalking.

Conjunctive Query-2A Represented in Clausal Form

$$Q(X) \leftarrow Accommodation(X) \wedge hasAccommodationDestination(X,A) \wedge hasAccommodationDestination(X,V) \wedge hasDestinationAttraction(V,B) \wedge hasDestinationAttraction(V,C) \wedge hasStarRating(X,D) \wedge hasCategory(X,E) \wedge hasAccommodationFacility(X,F) \wedge A = Victoria \wedge B = Surfing \wedge C = Bushwalking \wedge D = FourStar(D) \wedge F = OpenFireplace \wedge E = BedAndBreakfast\_Guesthouse$$

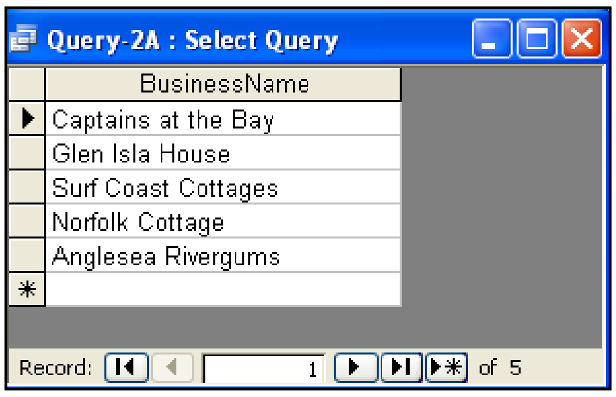


Figure 72: Conjunctive Query-2A results in Access.

<u>Conjunctive Query-2B</u>

Q(X) ← BedAndBreakfast_Guesthouse (X) hasAccommodationDestination(X,A) ∧

hasDestinationAttraction(X,B) ∧ hasDestinationAttraction(X,C) hasStarRating (X,D) ∧ hasFacility(X,F)∧

A = Victoria ∧ B = Surfing ∧ C = Bushwalking ∧ D = FourStar ∧ F = OpenFireplave

<u>Conjunctive Query -2B Represented as an Ontology Concept</u>

$(BedAndBreakfast\_Guesthouse \sqcap (\exists\ hasAccommodationDestination.\{Victoria\}) \sqcap$

$(\exists\ hasStarRating.\{FourStar\}) \sqcap$

$(\exists\ hasDestinationAttraction.\{Surfing\}) \sqcap$

$(\exists\ hasAccommodationFacility.\{Bushwalking\}) \sqcap$

$(\exists\ hasAccommodationFacility.\{OpenFireplave\}))$

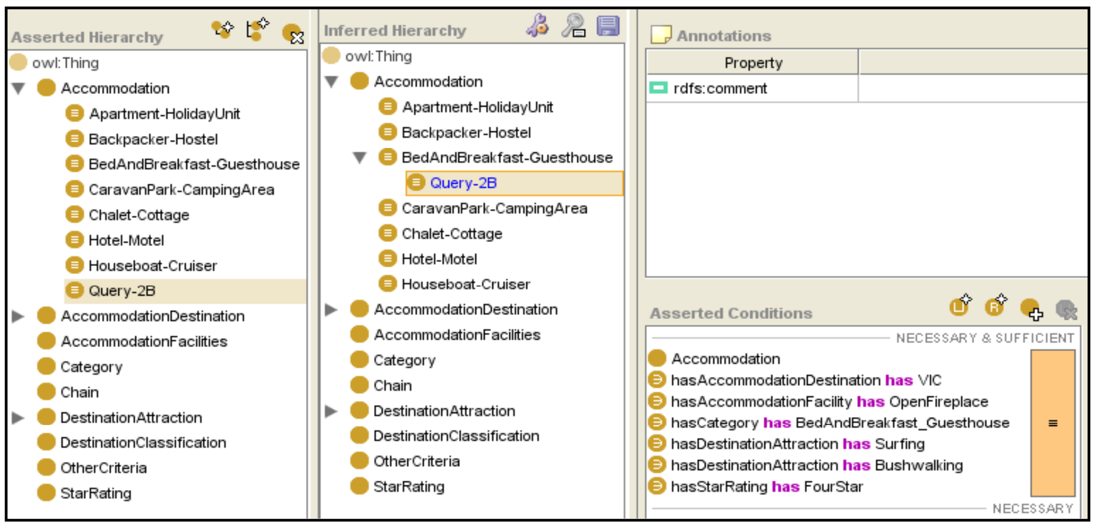


Figure 73: Conjunctive Query-2B as a Protégé ontology concept.

Query Comparison

Conjunctive Query-2A
Q(X) ← Accommodation(X) ∧ hasAccommodationDestination(X,A) ∧ hasAccommodationDestination(X,V) ∧ hasDestinationAttraction(V,B) ∧ hasDestinationAttraction(V,C) ∧ hasStarRating (X,D) ∧ hasCategory(X,E) ∧ hasAccommodationFacility(X,F) ∧ A = Victoria ∧ B = Surfing ∧ C = Bushwalking ∧ D = FourStar ∧ F = OpenFireplace ∧ E = BedAndBreakfast_Guesthouse
Conjunctive Query-2B
Q(X) ← BedAndBreakfast_Guesthouse (X) hasAccommodationDestination(X,A) ∧ hasDestinationAttraction(X,B) ∧ hasDestinationAttraction(X,C) hasStarRating (X,D) ∧ hasFacility(X,F) ∧ A = Victoria ∧ B = Surfing ∧ C = Bushwalking ∧ D = FourStar ∧ F = OpenFireplave

Figure 74: Versions of Query 2 to be compared

Measure	Results
Conjunctive Query-2A	
X^s	100
V^s (20*20)	400
Query terms	9
n^s(100*400)	40,000
Equivalent equi joins	1
Ordinal complexity ranking	1
Conjunctive Query-2B	
X^s	5
V^s	0
Query terms	7
n^s	10
Equivalent equi joins	0
Ordinal Complexity Ranking	2

s in Racer

Table 7: Query evaluation model applied to Query 2.

Query 2 Evaluation Summary

For query 2, the use of OWL semantics and a reasoner substantially reduced the query's complexity. Conjunctive *Query-1B* was searching directly for resorts with the destination attractions of surfing and bushwalking, as opposed to Conjunctive *Query-1A* which was searching for resorts with a destination that had the attractions of surfing and bushwalking. The use of transitive property *hasDestinationAttraction* (See Figure 75) in the accommodation ontology eliminated the equivalent of an equi-join for a relational model, and also reduced the number of query terms. Like Query 1, Query 2 was shortened since Conjunctive *Query-2B* was searching for a direct instance of the class

BedAndBreakfast_Guesthouse, as opposed to Conjunctive *Query-2A,* which was searching for an instance of the class *Accommodation* with the *hasCategory* property value of *BedAndBreakfast_Guesthouse*. As was the case with Query 1, this made Query 2 easier to formulate.

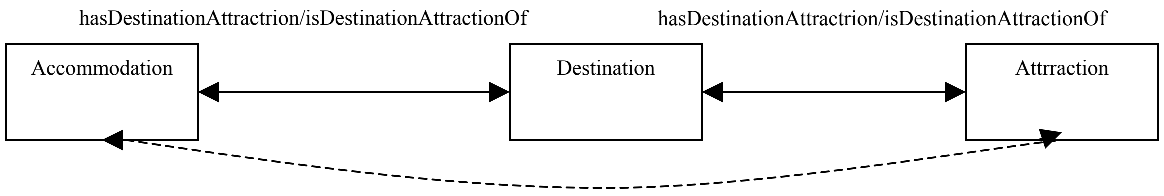


Figure 75: Use of a transitive property to reduce query complexity.

Query 3

Retrieve all three star rating CaravanPark_CampingAreas in NSW with cooking and barbeque facilities in a backpacker destination.

<u>Query 3 Assumptions</u>

- There are 10 three star rated CaravanPark_CampingAreas with cooking and barbeque facilities in NSW.
- There are 5 destinations in NSW with a backpacker classification
- There are 4 CaravanPark_CampingAreas with cooking and barbeque facilities in a backpacker destination in NSW

<u>Conjunctive Query-3A Represented in Clausal Form</u>

Q(X) ← Accommodation(X) ∧ hasAccommodationDestination(X, A)∧ hasDestinationClassification(X, B)∧ hasStarRating (X,C) ∧ hasCategory(X,D)∧ hasAccommodationFacility(X, E)∧ hasAccommodationFacility (X, F)∧ A = NSW∧ B = Backpackers ∧ C = ThreeStar ∧ D = CaravanPark_CampingArea ∧ E = CookingFacilities ∧ F = Barbeque

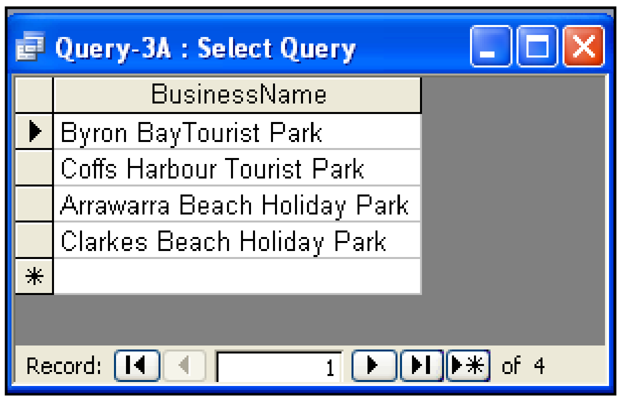


Figure 76: Conjunctive Query-3A results in Access.

<u>Conjunctive Query-3B Represented in Clausal Form</u>

Q(X) ← CaravanPark_CampingArea(X) ∧ hasAccommodationDestination(X,A) ∧ hasDestinationClassification(X,B) ∧ hasStarRating(X,C) ∧ hasAccommodationFacility(X,E) ∧ hasAccommodationFacility(X,F) ∧ A = NSW ∧ B = Backpackers ∧ C = ThreeStar ∧ E = CookingFacilities ∧ F = Barbeque

<u>Conjunctive Query -3B Represented as an Ontology Concept</u>

$$(CaravanPark\_CampingArea \sqcap (\exists\ hasAccommodationDestination.\{NSW\}) \sqcap$$
$$(\exists\ hasStarRating.\{ThreeStar\}) \sqcap$$
$$(\exists\ hasCategory.\{CaravanPark\_CampingAreas\}) \sqcap$$
$$(\exists\ hasDestinationClassification.\{Backpackers\}) \sqcap$$
$$(\exists\ hasAccommodationFacility.\{CookingFacilities\}) \sqcap$$
$$(\exists\ hasAccommodationFacility.\{Barbeque\}))$$

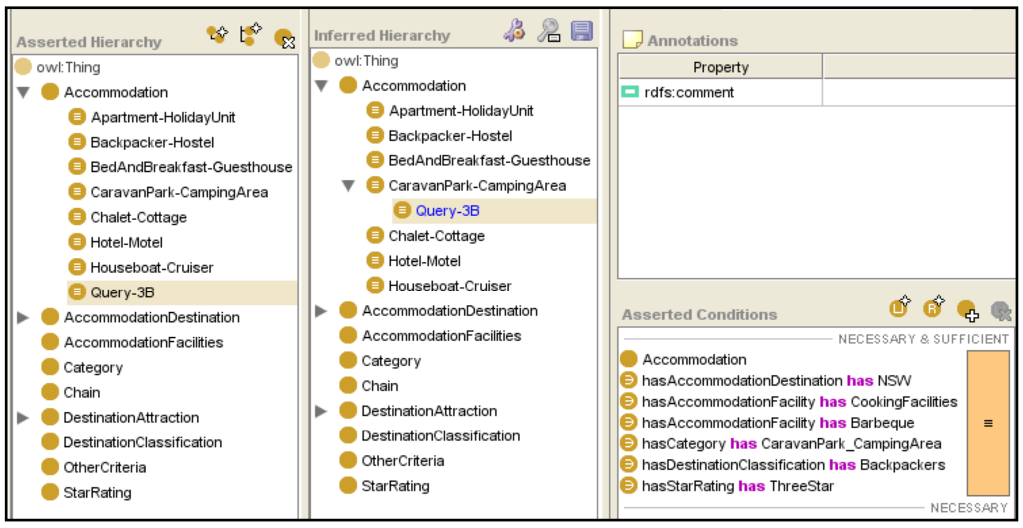


Figure 77: Conjunctive Query-3B as an Protégé ontology concept.

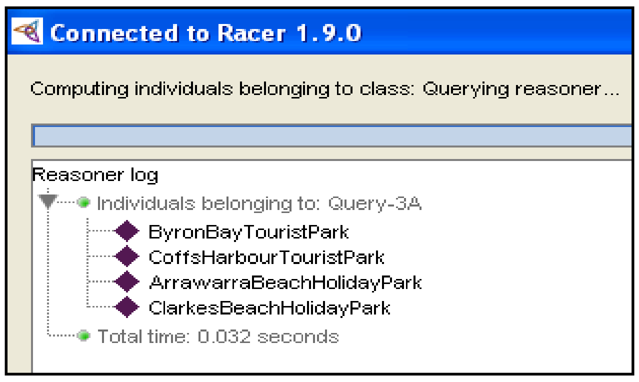


Figure 78: Conjunctive Query-3B results in Racer.

Query Comparison

Conjunctive Query-3A *Q(X) ← Accommodation(X) ∧ hasAccommodationDestination(X,A) ∧ hasAccommodationDestination(X,B) ∧ hasDestinationClassification(B,C) ∧ hasStarRating (X,D) ∧ hasCategory(X,E) ∧ hasAccommodationFacility(X,F) ∧ hasAccommodationFacility(X,G) ∧ A = NSW ∧ C = Backpackers ∧ D = ThreeStar ∧ E = CaravanPark_CampingAreas ∧ F= CookingFacilities ∧ G = Barbeque*
Conjunctive Query-3B *Q(X) ← CaravanPark_CampingArea(X) ∧ hasAccommodationDestination (X,A) ∧ hasDestinationClassification(X,C) ∧ hasStarRating(X,D) ∧ hasAccommodationFacility(X,F) ∧ hasAccommodationFacility(X,G) ∧ A = NSW ∧ C = Backpackers ∧ D = ThreeStar ∧ F = CookingFacilities ∧ G = Barbeque*

Figure 79: Versions of Query 3 to be compared.

Measure	Results
Conjunctive Query-3A	
X^s	10
B^s	5
Query terms	9
n^s(10*5)	50
Equivalent equi joins	1
Ordinal complexity ranking	1
Conjunctive Query-3B	
X^s	4
B^s	0
Query terms	7
n^s	4
Equivalent equi joins	0
Ordinal Complexity Ranking	2

Table 8: Query evaluation model applied to Query 3.

<u>Query 3 Evaluation Summary</u>

Again, *Query-3B* was considerably less complex than *Query-3A*. The use of OWL semantics shortened the query from 9 query terms to 7. The number of value to variable assignments was also reduced from 10 to 4. This was achieved through the use of class restrictions in the ontology, shown in Figure 80, which specified the location characteristics for Backpacker destinations. In doing so, Query-3B was able to search directly for accommodation resorts with a *Backpacker* destination classification, as opposed to searching resorts in a destination of which the destination has a *Backpacker* classification. The OWL semantics were used to infer a resort's location classification, thus integrating information about location characteristics with information relating to specific resorts. This form of inference was not possible in the static relational model. The location classification needed to be asserted in the relational model and manually associated with each resort.

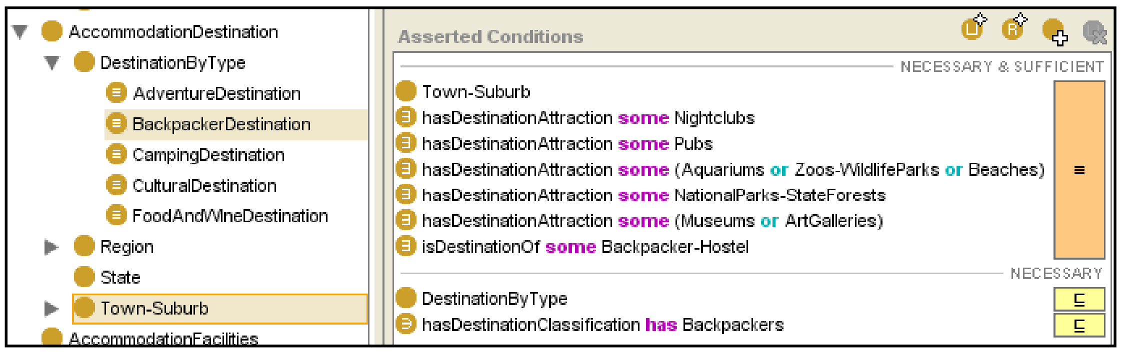


Figure 80: Class restrictions for specifying Backpacker destinations.

Query 4

Retrieve all Hotel-Motels with a five star rating in adventure destinations in QLD, with conference facilities, a spa, and destination attractions of beaches and guided tours.

<u>Query 4 Assumptions</u>

- There are there are 50 Hotel-Motels with a five star rating in QLD, with conference facilities and a spa.
- 10 destinations in QLD have a beach attraction.
- 10 destinations in QLD have a guided tour attraction.
- There are 5 adventure destinations in QLD.

- There are there are 5 Hotel-Motels with conference facilities and a spa with a five star rating in an adventure destination in QLD with the attractions of a beach and guided tours.

<u>Conjunctive Query-4A Represented in Clausal Form</u>

$Q(X) \leftarrow Accommodation(X) \wedge hasAccommodationDestination(X,A) \wedge$

$hasAccommodationDestination(X,B) \wedge hasDestinationClassification(B,C) \wedge hasStarRating\ (X,D) \wedge$

$hasCategory(X,E) \wedge hasAccommodationFacility(X,F) \wedge hasDestinationAttraction(B,G) \wedge$

$hasDestinationAttraction(B,H) \wedge A = QLD \wedge C = Adventure \wedge D = FiveStar \wedge E = Hotel\text{-}Motel \wedge$

$F = Spa \wedge G = Beaches \wedge H = GuidedTours$

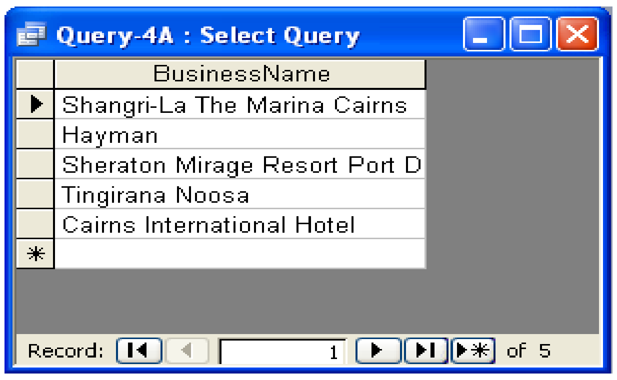


Figure 81: Conjunctive Query-4A results in Access.

<u>Conjunctive Query-4B Represented in Clausal Form</u>

$Q(X) \leftarrow Hotel\_Motel(X) \wedge hasAccommodationDestination(X,A) \wedge hasDestinationClassification(B,C) \wedge$

$hasStarRating\ (X,D) \wedge hasAccommodationFacility(X,F) \wedge hasDestinationAttraction(X,G) \wedge$

$hasDestinationAttraction(X,H) \wedge A = QLD \wedge C = Adventurers \wedge D = FiveStar \wedge F = Spa \wedge G =$

$Beaches \wedge H = GuidedTours$

<u>Conjunctive Query - 4B Represented as an Ontology Concept</u>

$(Hotel\_Motel \sqcap (\exists\ hasAccommodationDestination.\{QLD\})$

$(\exists\ hasStarRating.\{FiveStar\}) \sqcap$

$(\exists\ hasAccommodationFacility.\{Spa\}) \sqcap$

$(\exists\ hasAccommodationFacility.\{ConferenceFacilities\}) \sqcap$

$(\exists\ hasDestinationClassification.\{Adventurers\}) \sqcap$

$(\exists\ hasDestinaiionAttraction.\{GuidedTours\}) \sqcap$

$(\exists\ hasDestinaiionAttraction.\{Beaches\}))$

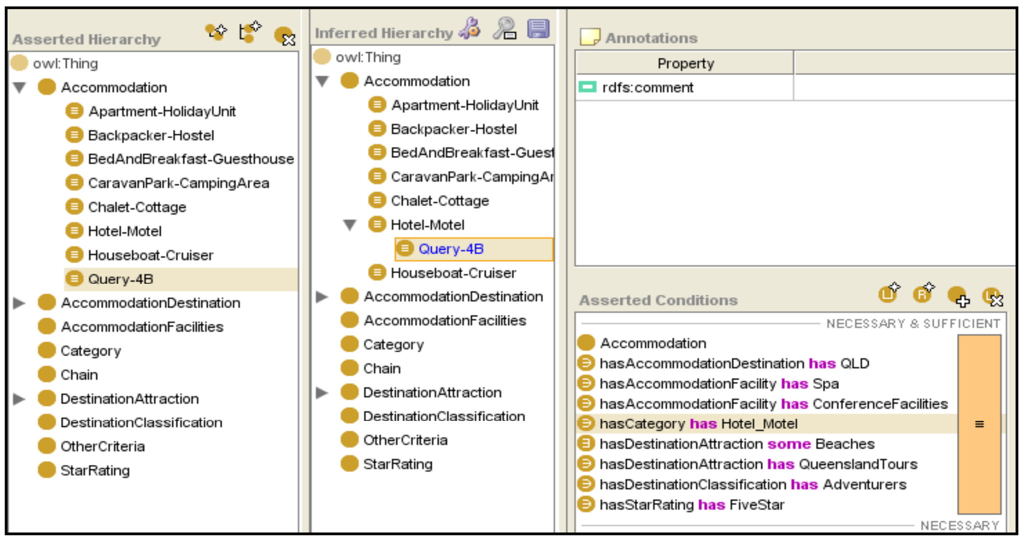


Figure 82: Conjunctive Query-4B as a Protégé ontology concept.

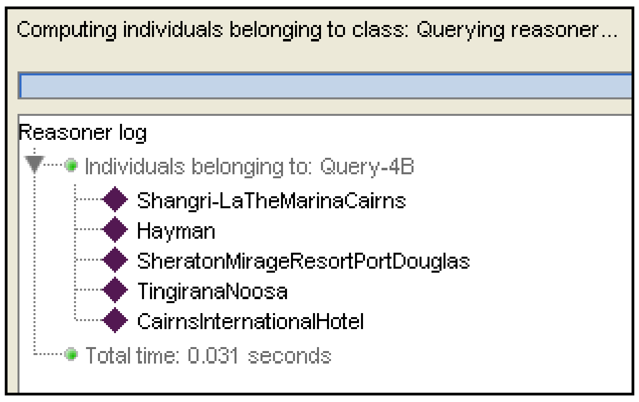


Figure 83: Conjunctive Query-4B results in Racer.

Query Comparison

Conjunctive Query-4A

$Q(X) \leftarrow Accommodation(X) \wedge hasAccommodationDestination(X,A) \wedge hasAccommodationDestination(X,B) \wedge hasDestinationClassification(B,C) \wedge hasStarRating(X,D) \wedge hasCategory(X,E) \wedge hasAccommodationFacility(X,F) \wedge hasDestinationAttraction(B,G) \wedge hasDestinationAttraction(B,H) \wedge A = QLD \wedge C = Adventure \wedge D = FiveStar \wedge E = Hotel\text{-}Motel \wedge F = Spa \wedge G = Beaches \wedge H = GuidedTours$

Conjunctive Query-4B

$Q(X) \leftarrow Hotel\_Motel(X) \wedge hasAccommodationDestination(X,A) \wedge hasDestinationClassification(B,C) \wedge hasStarRating(X,D) \wedge hasAccommodationFacility(X,F) \wedge hasDestinationAttraction(X,G) \wedge hasDestinationAttraction(X,H) \wedge A = QLD \wedge C = Adventurers \wedge D = FiveStar \wedge F = Spa \wedge G = Beaches \wedge H = GuidedTours$

Figure 84: Versions of Query 4 to be compared.

Measure	Results
Conjunctive Query-4A	
X^s	50
B^s (10*10*5)	500
n^s (50*500)	25,000
Query terms	10
Equivalent equi joins	2
Ordinal complexity ranking	1
Conjunctive Query-4B	
X^s	5
B^s	0
n^s	5
Query terms	8
Equivalent equi joins	0
Ordinal complexity ranking	2

Table 9: Query evaluation model applied to Query 4.

<u>Query 4 Evaluation Summary</u>

Once again Query 4 showed that the use of OWL semantics substantially reduced query complexity. *Query-4B* had 8 query terms compared to 10 for *Query-4A*. *Query-4A* also had 500 value to variable assignments compared to 5 for *Query-4B*. The reduction in query complexity was achieved by using the transitive property *hasDestinationAttraction,* which allowed knowledge to be inferred about attractions associated with a particular resort based on the resort's location (in the same way this was inferred for Query 2). The resort locations were automatically reclassified as Adventure destinations (see Figure 85), based on the attractions and accommodation types associated with the location, in the same was that this was done in Query 3 for backpacker destinations. For Query 4, the task of integrating information in the Accommodation domain was made easier by using Semantic Web technologies.

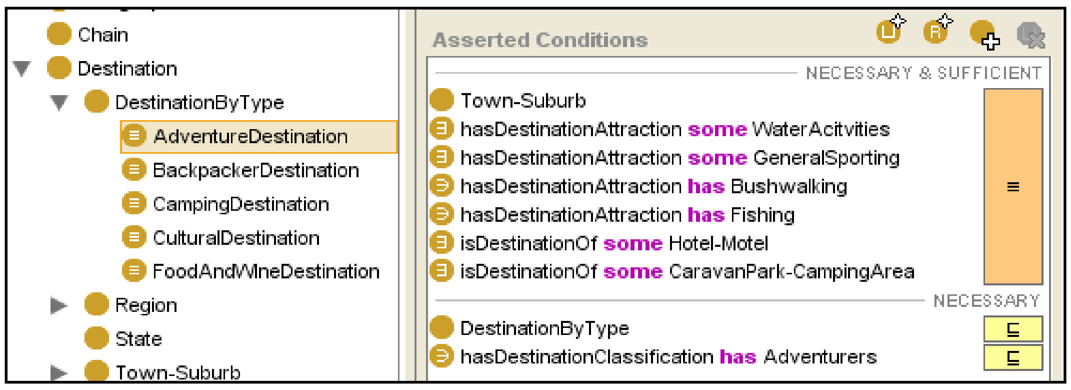


Figure 85: Class restrictions for specifying adventure destinations.

4.4 Chapter 4 Summary

The chapter presented detailed designs for the AcontoWeb semantic portal. The software requirement specification (SRS) provided a high-level view of the system's functional requirements, as well as showing system interfaces, individual screen designs and a usability guide demonstrating typical system processing. More detailed technical specifications are available with the software distribution which is included on the CD accompanying the thesis.

The second part of the chapter detailed the AcontoWeb query experiment that compared the complexity of queries and subsequent ease of information integration for a semantic portal, as opposed conventional portal based on a relational data model. The experiment tested four queries in a hierarchical order starting from a basic query searching only for directly asserted attributes, to increasingly complex queries that made use of OWL semantics in the ontology and complex table joins in the relational model. The experiment showed that there was little difference in complexity when querying directly asserted knowledge about a domain as in Query 1. The main advantages of using the ontology model for the first query were that firstly, the number of query terms was able to be slightly reduced. Secondly, by searching for the underlying concept *Conference Facilities*, results were returned for both the Mantra Erskine and Cumberland resorts - even though the two resorts used a different keyword to describe the concept. This showed that seamless integration was achieved without the need for explicit, runtime data mapping.

At the next level, Query used a transitive property in the ontology model to infer the attractions associated with resorts based on their location. This reduced the query terms, the value to variable assignments, as well as the equivalent equi-joins, thereby improving the integration process. Query 3 was also made less complex when processed in the Semantic Web environment. Ontology class restrictions were used to infer which resorts had a Backpacker classification based on the characteristics of the resorts location. Query 4, when processed using the ontology model, made use of both a transitive property in the same way as Query 2 to infer the attractions associated with resorts, as well as class restrictions in the same way as Query 3, to infer which resorts had an Adventure location. Using the ontology model was therefore shown to have reduced the complexity of Query

4, and eased the integration task in a similar manner to that demonstrated by queries 2 and 3.

In summary, the query experiment contributed to the research data requirements by demonstrating that a portal using a rich domain ontology for indexing purposes as opposed to a flat keyword list, was able to be queried with less complexity, which in turn improved the integration process. Complexity was shown to have been reduced in three of the four queries that were tested and the number of query terms was reduced for all four queries. It is important to note that a comparable number of rules still need to be implemented for a reasoner to process and interpret an ontology (as required in a database environment). The difference with a Semantic Web environment, however, is that part of the information processing occurs in advance of the queries rather than at runtime. For example, the inferred knowledge about locations associated with a particular resort and a resort's location classification still needs to be processed, but is done in advance of any queries over the data model, therefore at runtime, a query can be formulated to search directly for resorts with certain destination attractions and a destination classification, rather than searching for resorts that have a certain destination that has a particular attractions classification.

5 SURVEY OF TOURISM OPERATORS

5.1 Chapter 5 Overview

Chapter 5 reports on an investigation about attitudes towards adoption of new online technology among tourism operators, specifically accommodation service providers. The chapter includes analyses of the tourism operator survey and supporting secondary data interviews obtained from industry stakeholders. As previously noted, the survey was Web based and conducted by the researcher with businesses listed on the Royal Automobile Club of Victoria (RACV) online accommodation portal. Commencing on February 16, 2005, the survey ran for four weeks with 383 valid responses received. The principle reason for conducting the survey was to determine the degree of interest among Australian accommodation enterprises in an advanced, new online technology. Information was also sought that would provide a general overview of the purpose and functionality of accommodation Websites, as well as user preferences for the design of the AcontoWeb annotation tool.

It should be noted that the researcher was responsible for all aspects of the design and conduct of the survey: including its design, development of the Web-based instrument, obtaining access to survey subjects, data collection and data analysis. General guidance only was provided by the researcher's supervisor, Professor G. Michael McGrath, but Professor McGrath did request that some additional questions be included. These related primarily to a compatible, STCRC[100]-funded research project in which the researcher participated as a project team member. Outcomes of this research are detailed in McGrath et al. (2006), McGrath et al. (2005a), McGrath et al. (2005b) and McGrath et al. (2005c).

Outcomes from the secondary data interviews with industry stakeholders were published in McGrath et al. (2005c). Direct quotes from these interviews included in this chapter are all attributed to 'Interviews (2004)'. Finally, most survey results are presented here in bar-chart form. Tables underlying these charts are presented in Appendix G.

[100] The Australian *Sustainable Tourism Cooperative Research Centre (STCRC).*

5.2 General Information Concerning Participant Websites

Geographically the respondents' distribution as shown in Figure 86, was slightly biased towards Victoria, i.e. 24.0% of the sample enterprises were in Victoria compared with an actual figure of 21.4% (ABS 2002). The number of responses from WA, the ACT and NT were very low (7, 8 and 13 respectively).

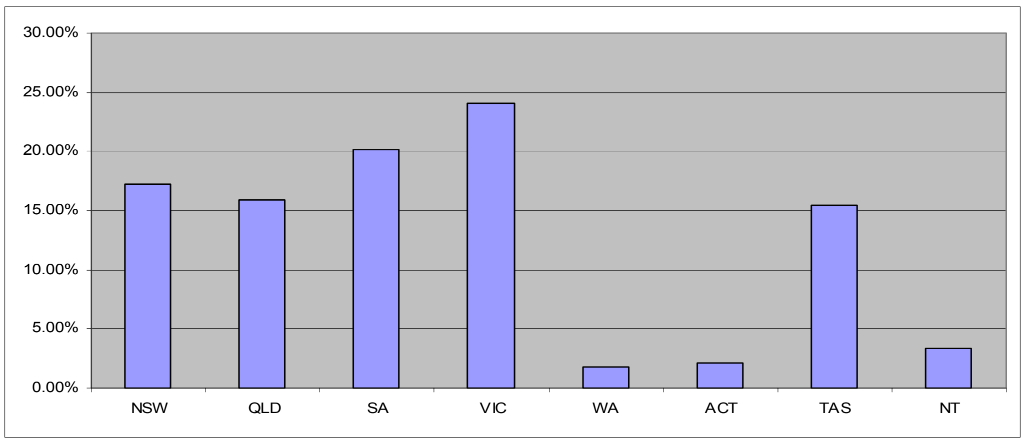


Figure 86: Respondents by state.

Figure 87 shows that the largest category of respondents was hotel/motel operators at 31.6%, followed by B&B/guesthouse operators at 27.2%.

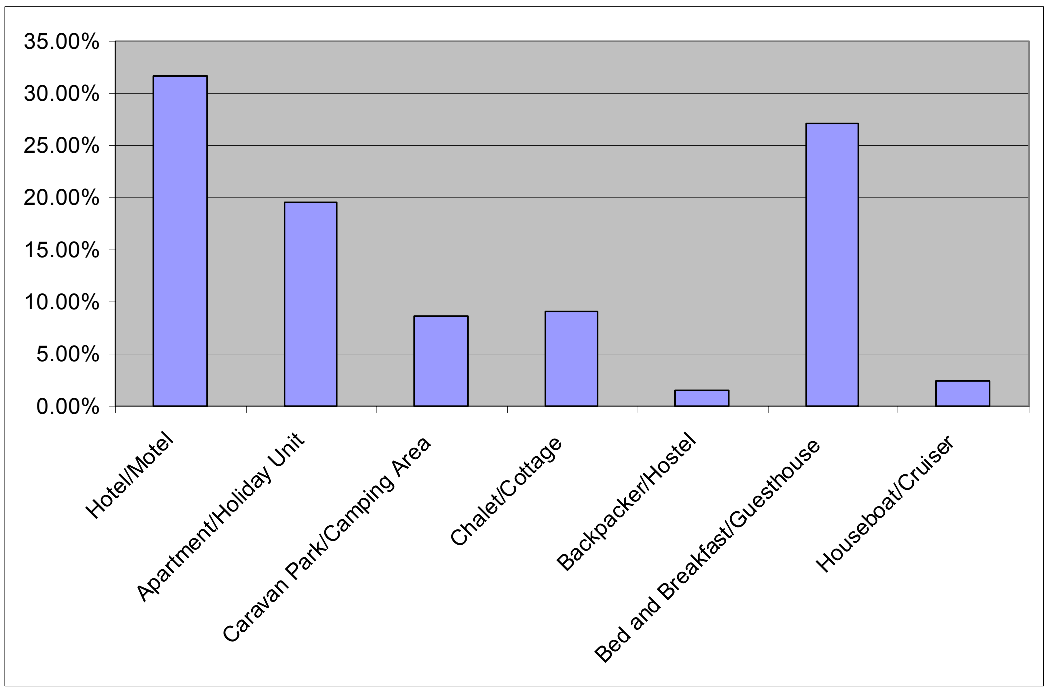


Figure 87: Respondents by business type.

Figure 88 shows that most enterprises (57.7%) were rated at the 4-4.5 Star level, 30.5% were 3-3.5 star operations and only 4.4% were rated 2.5 Star or less which is not representative. The implications for external validity were discussed in section 3.7.

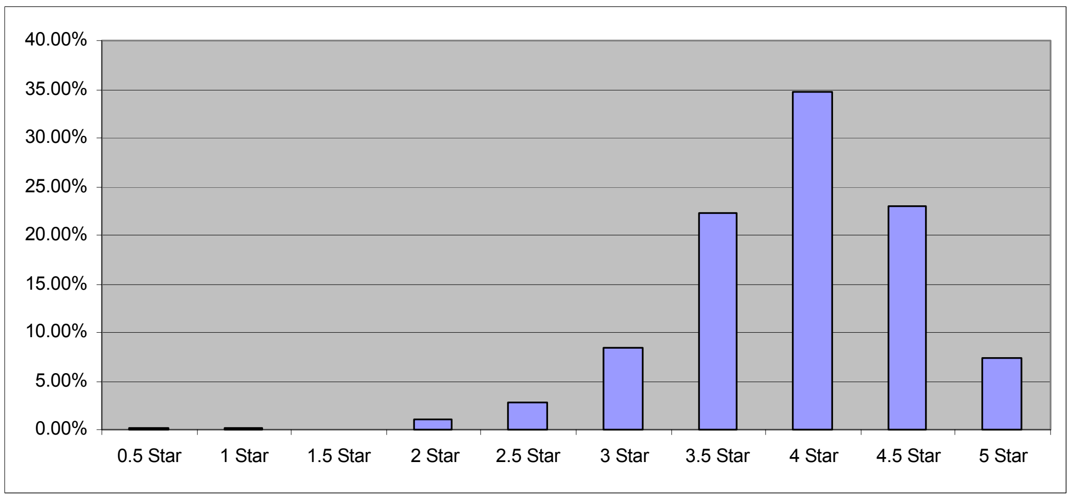


Figure 88: Respondents by star rating.

There are a number of excellent, general-purpose Web development software packages on the market (e.g. *FrontPage®)*. However, many SMTE operators have demonstrated a reluctance to take advantage of these software packages (McGrath et al. 2006, p. 3). Information was therefore sought about who was responsible for developing and maintaining business Websites to indicate likely users of the AcontoWeb annotation tool. The survey showed that in 63.2% of cases an IT professional developed a Website (see Figure 89), and in 53% of cases IT professionals were hired to maintain a Website (see Figure 90). This possibly suggests that the reluctance of business operators to use packages such as FrontPage may not be such a major issue for AcontoWeb, because mostly IT professionals use Web development tools.

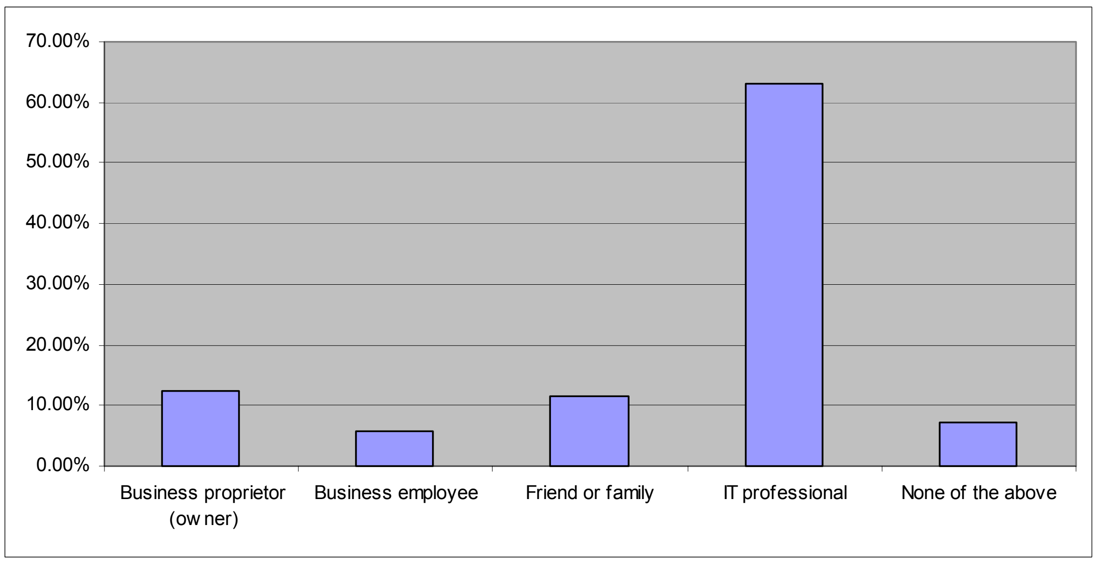


Figure 89: Creator of business Website.

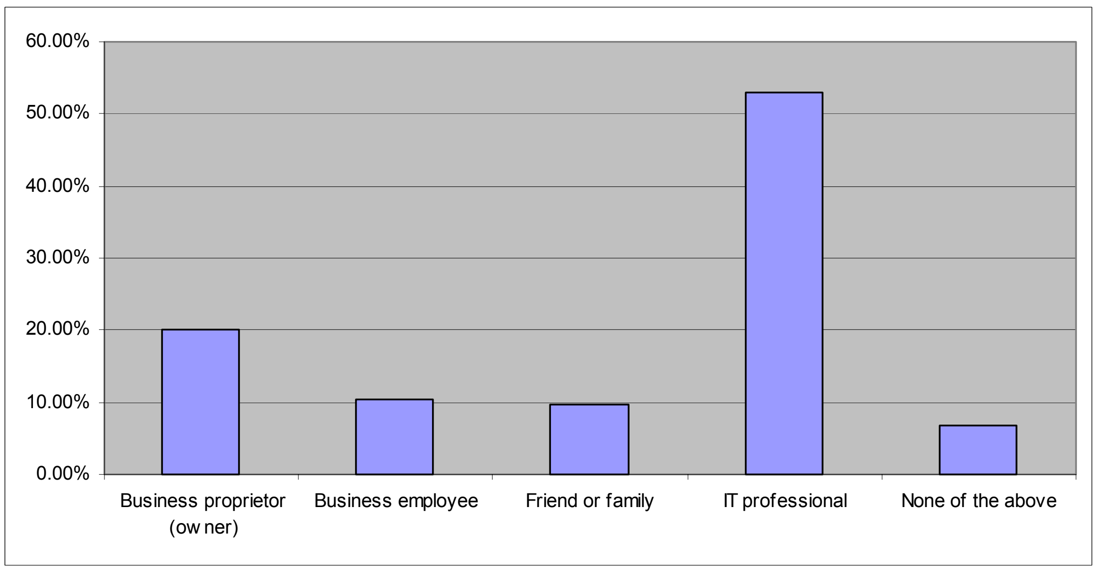


Figure 90: Maintainer of business Website.

When asked the main purpose of their exiting Website, Figure 91 shows that the most popular answers (multiple answers were permitted for the question) were 'Advertising and Promotion', 'Means of contact' and 'Means of providing information', which all rated ahead of 'Online bookings'. This appears to suggest that having online booking and payment facility, while strongly desired (as was indicated in later parts of the survey), is not necessarily considered by accommodation providers to be the most important feature of a Website.

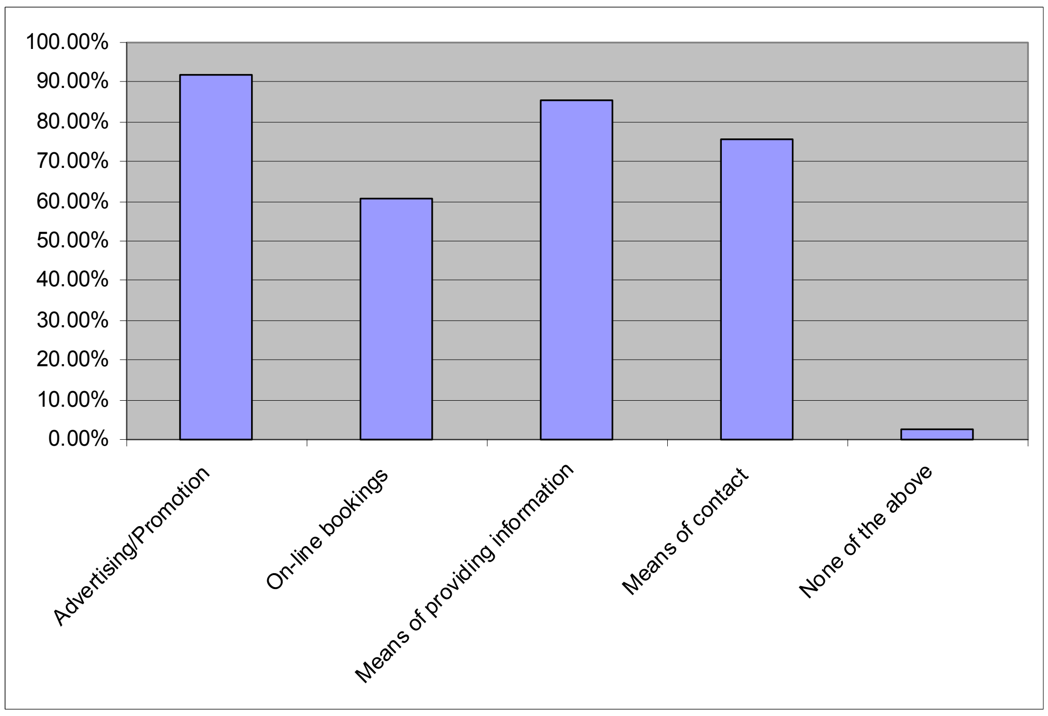


Figure 91: Purpose of business Websites.

The survey found that 60.8% of respondents had an online booking facility and 26.4% had a secure online payment capability (see Table 10). Overall, 73.4% reported that 20% or less of their customers booked their accommodation online. Still, 17.2% reported that between 21 and 50% of their customer base booked online and another 10.4% indicated that more than 50% of their customers generally used their online booking facility. This contrasts with the findings of Weeks & Crouch (1999), who estimated that less than 50% of Australian accommodation enterprises had Websites and, of these, only about one-third had booking facilities. Other Australasian studies conducted around the year 2000 (e.g. Applebee & Richie 2000) report similar, low levels of Net-readiness among tourism and hospitality enterprises and, thus, it is argued that the survey provides some support for the belief that accommodation enterprises (in particular) and their customers have now embraced Internet technology to a significantly greater extent than was the case some six years back.

Table 10 suggests a relationship between the quality level (AAA Star rating) of a property and the percentage of online bookings. Merging the percentage data from Table 10 into three categories *(less than 4 Star, 4 Star* and *more than 4 Star)* and applying a chi-squared test yields a value for that variable of 44.3. With 10 degrees of freedom, that is well above the value of 23.2 which might be expected (at the .01 level). Thus, the data indicates that there is a significant relationship between enterprise quality level and the percentage of customers booking online.

	Customers Booking Online													
	0%		1-5%		6-10%		11-20%		21-50%		51-100%		Totals	
Star Rating	Cnt	Pct	Cnt	Pct	Cnt	Pct	Cnt	Pct	Cnt	Pct	Cnt	Pct	Cnt	Pct
2.5 Star or Less	3	17.6	5	29.4	5	29.4	2	11.8	2	11.8	0	0	17	100
3 Star	4	12.5	13	40.6	5	15.6	4	12.5	6	18.8	0	0	32	100
3.5 Star	10	11.8	31	36.5	16	18.8	13	15.3	9	10.6	6	7.1	85	100
4 Star	9	6.8	24	18.0	26	19.5	37	27.8	27	20.3	10	7.5	133	100
4.5 Star	5	5.7	12	13.6	17	19.3	17	19.3	17	19.3	20	22.7	88	100
5 Star	1	3.6	7	25.0	5	17.9	6	21.4	5	17.9	4	14.3	28	100
Totals	32	8.4	92	24.0	74	19.3	79	20.6	66	17.2	40	10.4	383	100

Table 10: Customers booking online by star rating and within percentage ranges. For example, with properties rated at 2.5 Star or less, 17.6% of hotels reported none online (McGrath, et al., 2006).

Furthermore, as illustrated in Table 10 (and Figure 92) and described by McGrath et al. (op. cit), it would appear that better quality accommodation enterprises seem more likely to have their customers book online. This seems to contrast with the findings of Mistilis et al. (2004) who, in a survey of the use of ICT in a small number of Sydney hotels, reported a significantly higher proportion of Internet bookings in 3 Star hotels than in those belonging to more luxurious categories. The conclusion drawn here, however, does appear to be broadly consistent with the results of a recent study by Fotiodis et al. (2005): specifically, in looking at ICT adoption and use among Greek hotels, they reported a positive correlation between hotel size (and quality) and Internet use.

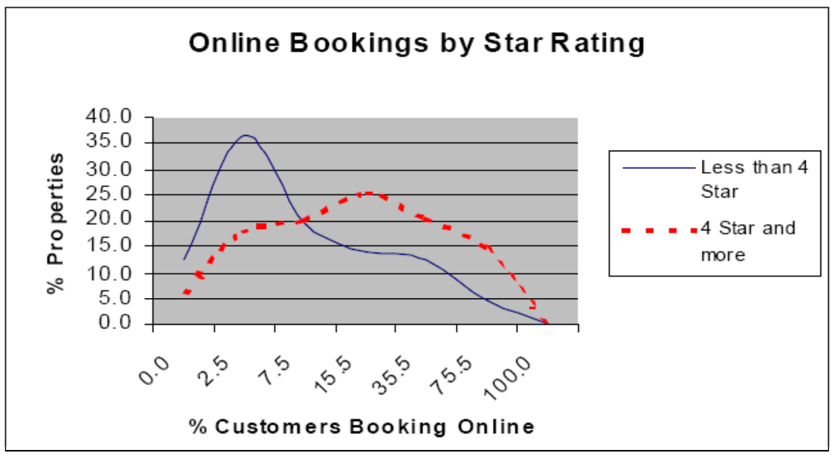


Figure 92: Percentage of customers booking online, broken down into properties rated less than 4 Star and those rated 4 Star and above (note that the X-axis is not to scale) (McGrath, et al., 2006).

Respondents were asked to nominate where they listed online (in addition to their own Websites). The results are illustrated in Figure 93, and clearly show that operators like to promote their enterprises on promotional sites close to home. Also, properties rated 4 Star and above seem to be considerably more likely to list on international sites.

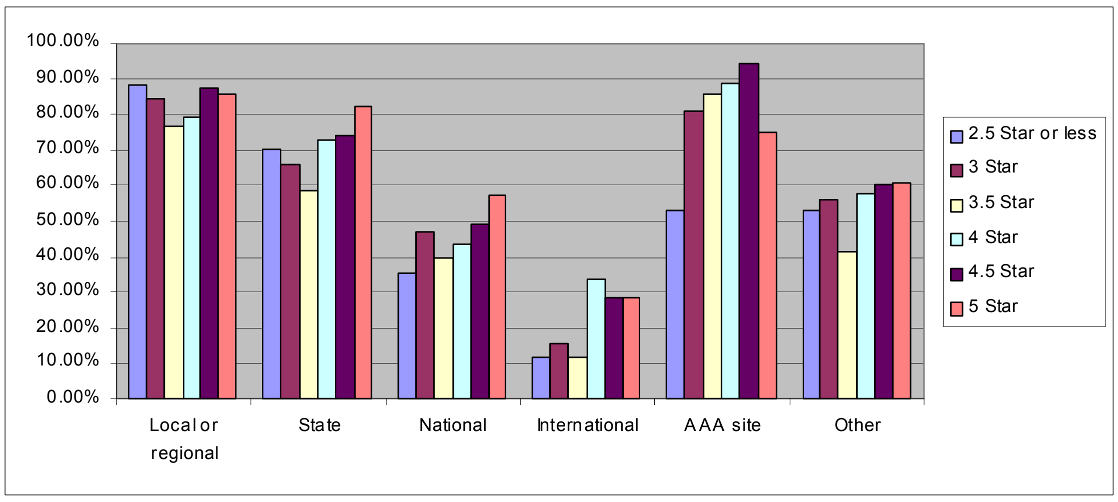


Figure 93: Additional online promotional outlets (McGrath, et al., 2006).

The desire to list closer to home was also apparent in the interviews used as supporting secondary data. For example, McGrath et al. (2005c) reported that several interviewees believed that SMTEs are reluctant to list at the national level – perhaps unreasonably. For example:

SMTEs have a negative attitude towards national sites ----- they don't see that they get any inbound custom. I suspect they do though – particularly from second and third-time visitors, who have done the capital cities and the other major attractions and are now looking to get off the beaten track a bit.
(Interviews, 2004)

5.3 Attitudes Towards Adoption of New Online Technology

Before taking full advantage of the technical benefits of using Semantic Web technology for tourism information integration and utilization, tourism operators need to be willing to adopt the technology. Previous research, however, suggests that the uptake of online ICT among Australian 'Small-to-Medium' Tourism Enterprises', including accommodation resorts, has been poor. McGrath et al. (2006) explain that this hostility was evident in a recent local newspaper article by Mitchell (2003) that focussed on the rapidly-diminishing profit margins of many Australian SMTEs. The article quoted one B&B operator as referring to "that monster the computer". The Victoria (Australia) Government's 'Victoria Tourism Online' (VTO) initiative (Morrison and King, 2002), also portrayed a negative view about attitudes towards online adoption. Here, SMTE's were categorized into *Techno-whizzos, Early adopters, Wait-and-sees* and *Wilderness* operators. The Wilderness group were described as generally aged 45+, with no computer or interest in them, they felt they were too old too learn more and they viewed the Internet as a waste of time. They also had a dislike of officialdom/bureaucracy and were reluctant to participate in RTO activities and networks (McGrath et al. 2006, p. 4). Morrison & King (2002) estimated that 60% of Victorian SMTEs were in the Wilderness category.

The research mentioned above, however, is now somewhat dated. For instance, Morrison and King's data was collected some 4-5 years ago and, even in this short timeframe, it is reasonable to expect that today's entrants to the smaller end of the tourism industry to be less resistant to IT than their earlier counterparts (even if the difference is only marginal)

(McGrath et al. 2006, p. 4). This is supported by the fact that, between December 2000 and June 2003, the percentage of Australians accessing the Internet increased from 44% to 59% and, perhaps even more impressively, the corresponding increase for the 55+ age bracket was from 18% to 29% - clear evidence of diminishing resistance to the use of online technology among older Australians. Locally, recent research suggests substantial growth in travel product purchases over the Internet (Roy Morgan 2003, 2004); and internationally, tourism-related businesses (and accommodation enterprises in particular) are experiencing rapid growth in online sales (PhoCusWright 2003).

Recalling that the primary reason for conducting the survey was to ascertain current attitudes among accommodation enterprise operators to the use of new Web-based technology, survey subjects were asked whether they would consider overhauling or rebuilding their websites in the next 12 to 18 months. Figure 94 shows that 56.7% or more than half of respondents indicated that they would maybe, likely, or definitely overhaul their Website in the next 12 to 18 months. This represents a possible opportunity for adding RDF markup to overhauled Websites.

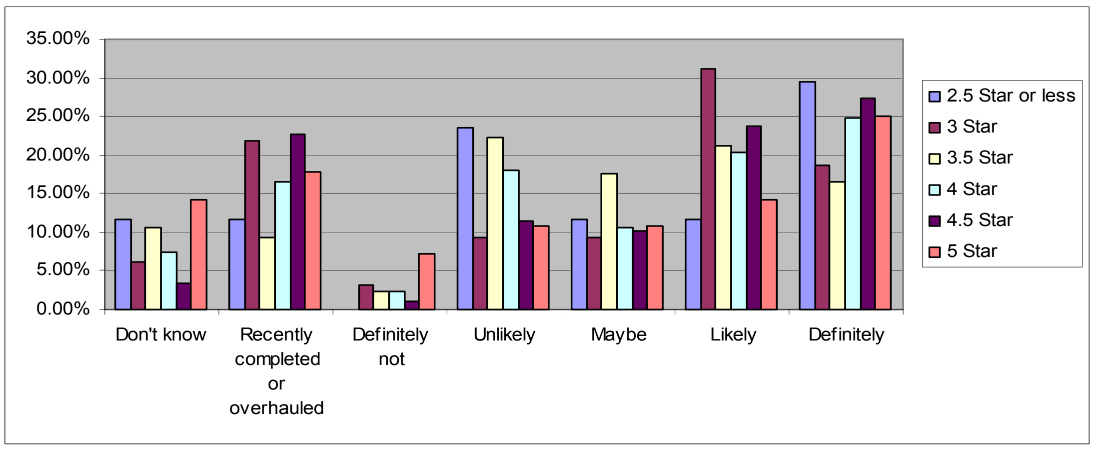


Figure 94: Likelihood of overhauling Website in next year to 18 months.

Perhaps the distribution presented in Figure 95 provides some clues to this positive attitude to improving Websites. Here, survey subjects nominated factors that would influence them in overhauling or rebuilding their websites within the next 12-18 months. Better marketing and promotion, improved efficiency and improved quality of service all rated reasonably highly. However, a desire to improve website layout and usability was the most significant factor nominated. This may indicate a fairly common dissatisfaction with current technology and, judging by the number of hospitality and tourism industry software packages now available, one might reasonably assume that there is real demand for these products. One of the secondary data interviewees endorsed the above view about demand for technology, but expressed doubts about the worth of many current vendor offerings:

Add up all the money being spent on software across the [accommodation] industry and you'd shudder. There are some very good PMS, but they've been purpose-built for larger hotels. It's the same with CRM systems: the really good ones have been built for banks etc. and require major customisation before they can be used in the accommodation sector. The price of this is coming down but it's still expensive for mid-range operators. At the other end of the market, there are lots of cheap packages but they're pretty useless. ---- The other problem here is knowledge. Many of my [operators] complain to me that hardly a day goes by when they aren't approached by 4-5 computer vendors with 'the answer to all their problems'. They just don't have the skills – or the time – to evaluate these products.

(Interviews, 2004)

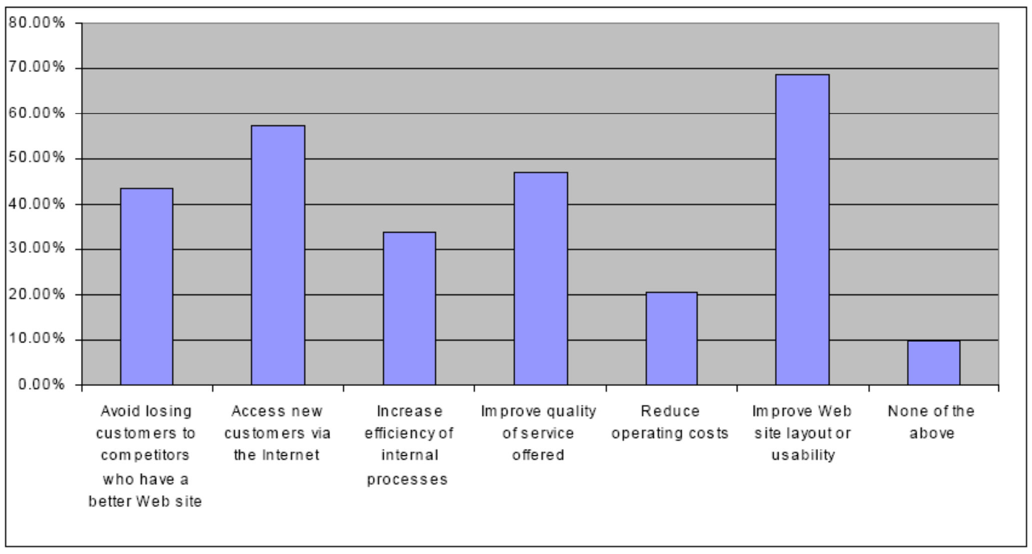


Figure 95: Factors that would encourage businesses to overhaul or rebuild a Website.

Factors that would discourage the overhauling of Websites are shown in Figure 96. The most significant factor here was 'Advantages are outweighed by cost implications with 38.6% followed by 'No significant benefits likely' with 30.8%.

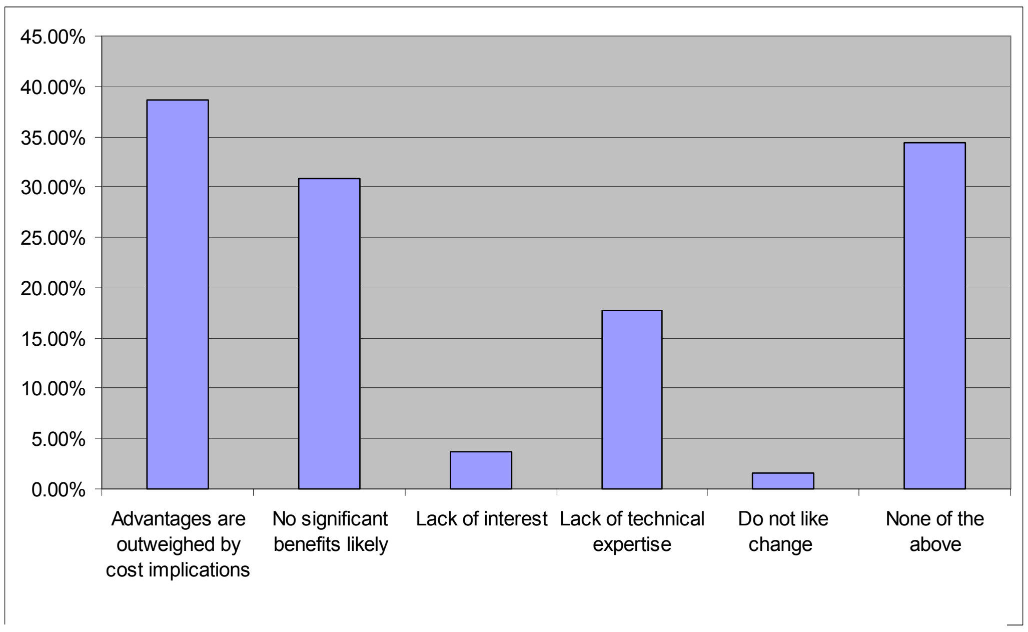


Figure 96: Factors that would discourage businesses to overhaul Website.

As indicated in Figure 97, while many respondents were equivocal about using a new technology (the 'Maybe' group), a great many more respondents were receptive to the idea than were against it (only 13 in the 'Unlikely' category against a total of 200 in the 'Likely'/'Definitely' groupings). McGrath et al. (2005a) reported that moreover, and perhaps somewhat surprisingly, the quality level of an enterprise does not appear to be a significant determinant of its interest in *new* technology. More specifically, merging the data into the same three quality groupings used previously and applying a chi-squared test to the data in Figure 95 yields a value for this variable of 13.3. This is not significant at the .05 level.

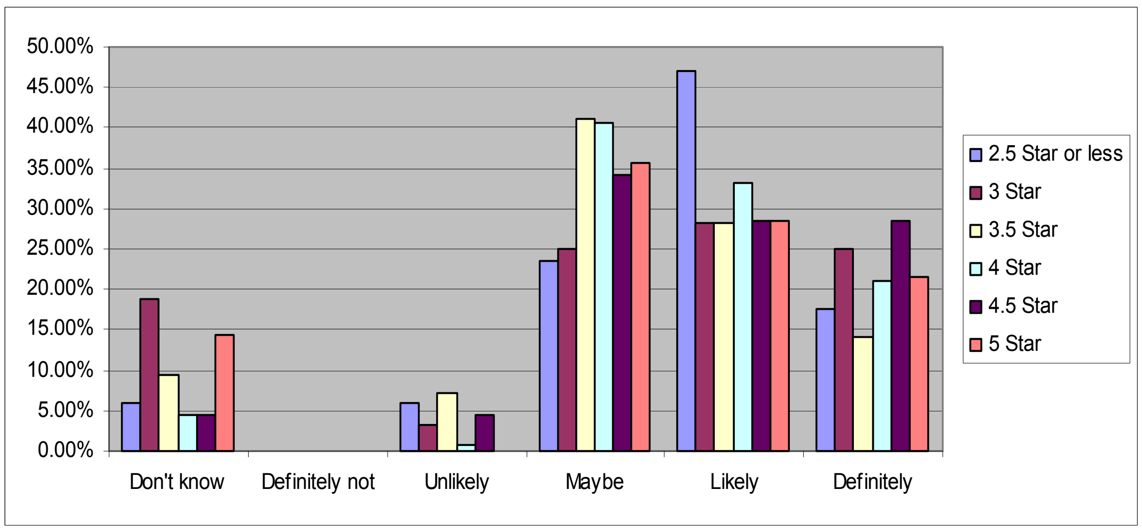


Figure 97: Likelihood of adopting new online technology.

Figure 98 shows that the most significant factor that would encourage business to adopt a new technology was 'If it was proven to increase Web exposure'. This was closely followed by 'It was easy to use', and 'If the cost of implementing it was low'. It can reasonably be assumed from these indictors that Semantic Web technology would have a better chance of being more widely accepted among tourism operators if: 1) annotations could be applied at a very low cost or even free of charge; 2) annotation software was user friendly; and 3) the commercial benefits of using Semantic Web technology for information integration and utilization were well communicated to potential users.

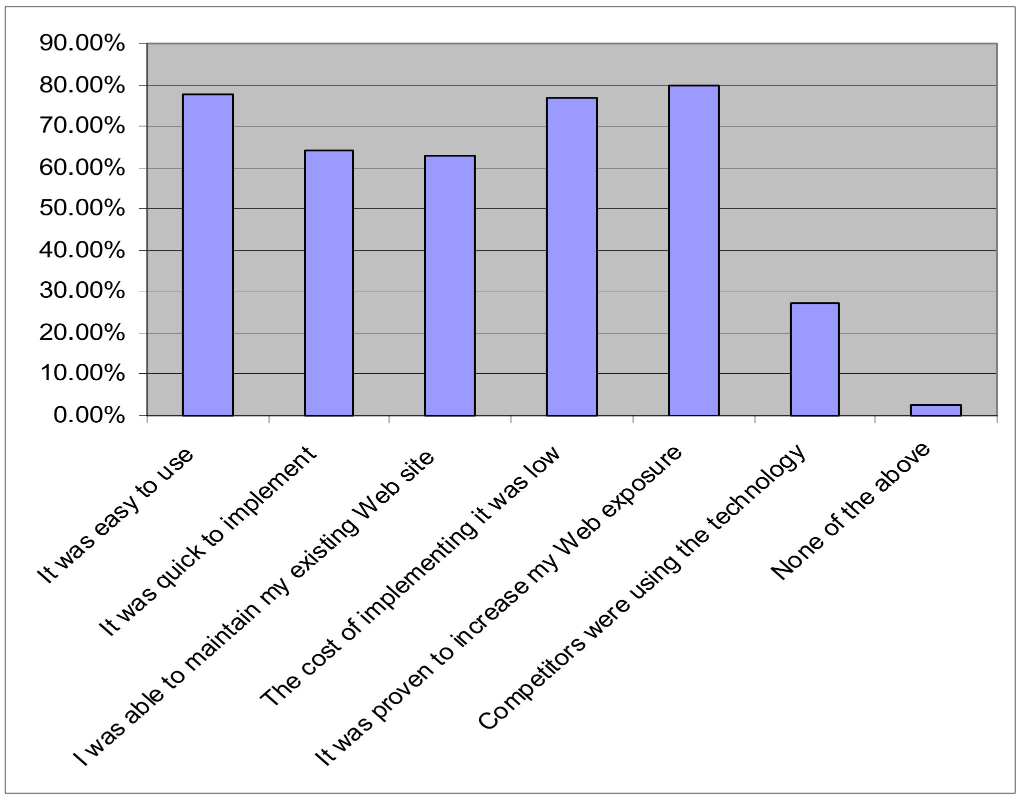


Figure 98: Factors that would encourage business to adopt a new Internet technology.

5.4 Implementation Preferences for New Online Technology

Respondents were asked how they would prefer any new technology be applied to their Website. Results show (see Figure 99) that there was an overwhelming desire that the technology be added to their existing Websites. This information was used in the design of the AcontoWeb annotation tool. Originally the tool generated new Websites, but was changed into a tool that marked up existing Webpages after the survey analysis.

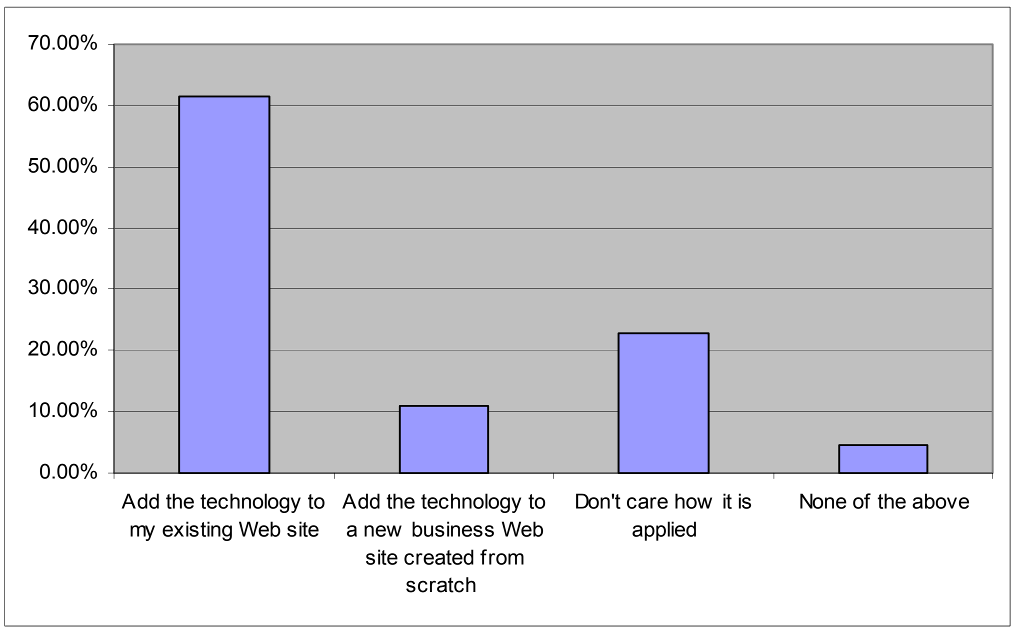


Figure 99: Preference for how a new Internet technology might be applied.

Finally, respondents suggested that they also prefered to include an online payment facility in a new or overhauled Website. Figure 100 shows that 25.1% were equivocal (the ‘Maybe’ category), 26.1% indicated definitely, and 19.6% chose likely.

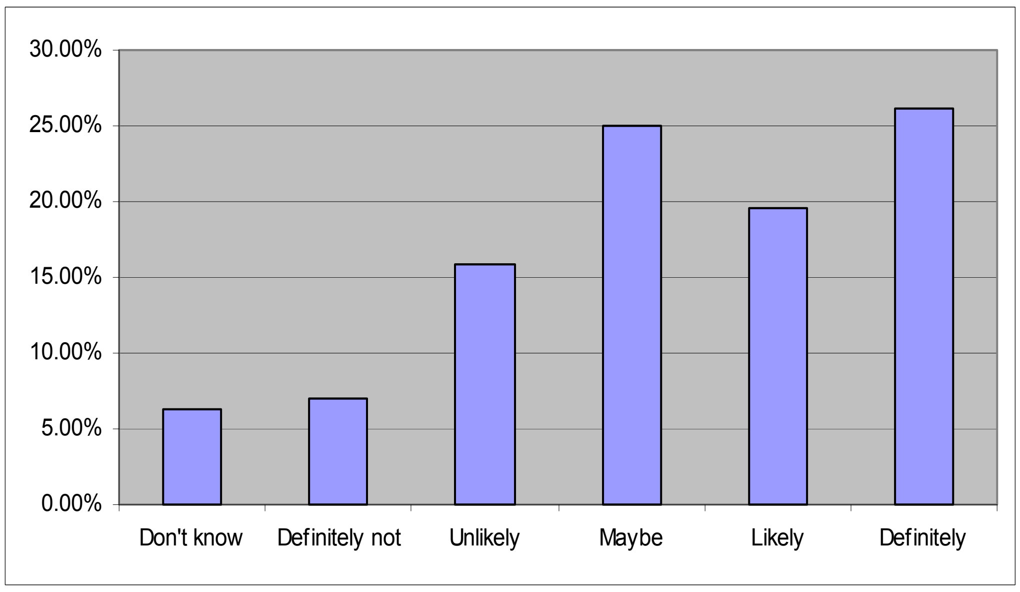


Figure 100: Preference for online payment facility.

5.5 Chapter 5 Summary

The chapter presented the results of an investigation into the attitudes towards adoption of new online technology among tourism operators. This investigation included a survey of accommodation Website owners and secondary data interviews conducted in 2005 and documented in McGrath et al. (2005c). The first part of the chapter focused on providing a general understanding of accommodation Websites. It was shown that Websites were created and maintained mainly by IT professionals. This information suggests that annotation software such as AcontoWeb perhaps should be developed and marketed primarily for use by IT professionals. The survey indicated that operators considered the main purpose of business Websites to be advertising/promotion, followed by information dissemination and providing a means of contact. Online booking and secure payment facilities, although not considered a primary function of a Website, were strongly desired for any future overhauled Website. These facilities were more common in higher quality Star rated hotels, and better quality hotels were also more likely to have additional Website listings outside of their region, including internationally.

Overall, in spite of some bleak results and prognoses a few years ago, and still with some scepticism remaining about technology in the tourism and hospitality industry, the research suggests there now seems a more positive trend and attitude. Moreover, those companies at the leading edge in the diffusion of innovation processes, clearly are engaging with technology in an additional competitive way by not only collaborating with suppliers and customers effectively, but also enhancing collaboration within the broader industry sector, and setting the agenda for technology adoption (McGrath et al 2005a p.10). Businesses were also enthusiastic towards new online technology and showed a willingness to consider adopting such technology if it could be proven to increase Web exposure, and was easy to use and inexpensive.

The major consideration incorporated into the AcontoWeb annotation tool resulting from tourism operator feedback was to add RDF annotations to exiting Websites rather than to a new Website, as originally specified. Consideration was also given to the fact that an overwhelming majority of Website owners indicated a preference for having an online payment facility incorporated into any overhauled Website. This preference, however, was outside the scope of the AcontoWeb prototype (mainly because of a lack of available development resources).

Finally, while the uptake of and use of ICT is currently very uneven across the industry, the pressures on owners and operators were summed up very neatly by one of the interviewees and reported by McGrath et al (2005, a p.10) as follows:

How many operators have a technology strategy? Very few! ----- Try and get [SMTE] operators to keep their websites up-to-date via control and you're beating your head against a brick wall. You will never get operators to come online via control or coercion ----- but, commercial factors will dictate they will have to: i.e. they will either learn from smart operators – or go out of business!
(Interviews, 2004)

Intuitively this seems reasonable, and the assertion could be tested in a more in-depth follow-up study that could include analysis of the Australian situation within an international context.

6 CONCLUSION

6.1 Chapter 6 Overview

Chapter 6 concludes the thesis with a summary of the research findings. The chapter commences with a discussion of findings in relation to each of the minor research questions, followed by a statement answering the major research question. This statement also represents the proposition of a grounded hypothesis that can be tested in further research, about the extent to which the Semantic Web and related technologies can assist with the creation, capture, integration and utilization of accurate, consistent, timely, up-to-date Web based tourism information. The chapter then describes how each of the research aims have been met and what the specific outcomes of the study were. Directions for potential future research in the topic area are also discussed.

6.2 Answers to Minor Research Questions

This section presents the findings in relation to the minor research questions.

6.2.1 Ease of Ontology Development, Availability and Website Annotation

An often quoted concern about the Semantic Web is the ease of ontology development, availability and Website annotation. Ontology library systems such as Protégé[101] or SHOE[102] offer a limited selection of ontologies for download. The ontologies that are available are generally purpose built, meaning there is often a reusability-usability trade-off problem as described by Klinker et al.(1991). The idea of a single consistent ontology for every domain sounds like an ideal solution, but such a wide ranging all-encompassing approach clearly won't scale and can't be enforced. Ontologies usually need to be developed and tailored for individual systems. Development of the AcontoWeb accommodation ontology showed that this can be a relatively complex and time consuming process. Numerous axioms had to be specified in the accommodation ontology to facilitate the types of inference that were required. The time and cost of

[101] http://protege.stanford.edu/plugins/owl/owl-library/index.html

[102] http://www.cs.umd.edu/projects/plus/SHOE/onts/

ontology development and the need for continuing maintenance can therefore be viewed as likely impediments to wide-scale adoption of the Semantic Web.

On the positive side, in certain commercial applications the potential profit and productivity gain from using well structured coordinated vocabulary specifications will outweigh the sunk costs of developing an ontology and the marginal costs of maintenance (Shadbolt, Hall & Berners-Lee 2006, p. 99). If it is assumed as Shadbolt et al. (2006) have done, that ontology building costs are spread across user communities, the number of ontology engineers required increases as the log of the user community's size. The amount of building time then increases as the square of the number of engineers, and so the effort involved per user in building ontologies for large communities gets very small very quickly. In many areas the costs will be easy to re-coup. These are reasonable assumptions for a basic model.

Data annotation also remains problematic from a practical perspective. As yet there are few means to routinely and effortlessly generate Semantic Web annotations. The RDF and OWL formats are for machines so Web authors can no longer embed information in plain English. The information needs to be formatted as RDF triples, which are separate from any natural language representations. These formats have seen extremely low adoption rates, thus, there is a real need for representations to be made easier to translate to and from natural language. The AcontoWeb annotation tool proves that this is quite achievable for an individual domain. AcontoWeb accepts user input from accommodation providers and translates it into RDF instance data consistent with the accommodation ontology. The RDF markup is then imbedded into readily extractable comment tags in an HTML file. AcontoWeb demonstrated that this approach works well in a managed portal environment with well defined functionality and limited Web access. It can be said though (i.e. Hepp, 2006), that embedding RDF markup within HTML code violates the one fact in one place paradigm which has contributed so much to data consistency since Codd (1970) introduced it. This potentially causes problems with data inaccuracy if an annotator fails to update the information when the human readable content changes.

More flexible approaches to content creation are required if wide-scale adoption of the Semantic Web is to occur. Human Language Technology (HLT)[103] and Latent Semantic Indexing[104] are promising alternatives. These techniques can place data into a semantic structure using an algorithmic approach. Hepp (2006) states that this raises the obvious question as to whether physical annotation of data needs to occur at all if techniques such as HLT or LSI can apply at query run time. The annotation of dynamic content also remains a problem. Most annotators work for static pages only. A possible solution is to leave RDF metadata in databases and generate dynamic Webpages from it. This is how query results are displayed in AcontoWeb. Here, the results page is dynamically generated from instance data about accommodation resorts stored in a backend database.

6.2.2 Level of Ontology and Annotation Richness that can be Obtained

Knowledge representation is a technique with mathematical roots in the work of Codd (op. cit) in which the theory is to translate information, which humans represent with natural language, into sets of tables that use well defined schema to define what can be entered in rows and columns (McCool 2005, p. 86). The technique led to the creation of the relational database revolution in the 1980's and also forms the basis of OWL ontologies. The problem with these forms of knowledge representation is that they create a fundamental barrier in terms of richness of representation, as well as creation and maintenance, compared to the written language that people use and HTML incorporates. In the OWL DL AcontoWeb accommodation ontology, cardinality of constraints were unable to be included in the class restrictions for destination classifications without the ontology changing into OWL Full. It was not possible for example, using the OWL DL language, to say that a backpacker location has a minimum of *3 pubs*. The class restriction relating to pubs could only express the fact that a backpacker location has at least *some pubs*.

OWL full is more expressive than OWL DL but still suffers from an inability to represent exceptions to rules and the contexts in which they are valid. Depending on the level of expressiveness required, there can be a need for more powerful languages other than RDF

[103] http://www.mitre.org/work/ird_human_language.html

[104] http://www.cs.utk.edu/~lsi/

and OWL. SWRL[105] is one such language that builds on OWL. The more expressive markup languages like SWRL allow developers to write application-specific declarative knowledge, and can improve the ontology and annotation richness of information on the Semantic Web.

6.2.3 Maturity and Ease of Use of Semantic Web Development Tools

Tool development support for building Semantic Web applications has increased enormously in recent years. The first packages to emerge were ontology development tools which appeared in the mid-1990s. Since then, a range of other tools have been created to assist with developing Semantic Web applications. Some tool suites integrate tools from different groups, while others provide a limited set of isolated functions used for carrying out specific tasks (e.g. ontology merge tools). Three types of tools were used to develop AcontoWeb: 1) an ontology development tool; 2) an ontology based annotation tool; and 3) a semantic middleware application and inference engine.

The protégé[106] tool was used to develop the accommodation ontology. Protégé was user friendly and provided a range of formats for exporting ontologies. It also provided a SPARQL[107] query tab that was used to test the syntax of the experimental queries of Chapter 4. A range of annotators were also tried while conducting the research, including tools such as Ontomat Annotizer[108] and the COHSE[109] annotator. Existing tools were generally found to be awkward and slow to use and required annotations to be manually inserted into Webpages by dropping and dragging class instances from an ontology concept to the Webpage. A new annotation tool was therefore developed as part of AcontoWeb, so that user input of resort details could be automatically transformed into RDF triples and imbedded within the HTML code of accommodation Websites. The tool is considerably easier to use than other existing annotators.

[105] http://www.daml.org/2003/11/swrl/

[106] http://protege.stanford.edu/

[107] http://www.w3.org/TR/rdf-sparql-query/

[108] http://annotation.semanticweb.org/ontomat/index.html

[109] http://www.ecs.soton.ac.uk/~tmb/cohse/annotator/

A number of semantic middleware applications and inference engines were experimented with throughout the development phase of the research. Jena[110] was found to be the best middleware environment because it was completely open source, had a vast array of functional libraries, and was compatible with numerous reasoners via its DIG interface. Jena also offered support for processing SPARQL queries. The pellet reasoner was used as an inference engine because it had sufficient performance capability to handle the types of queries that AcontoWeb needed to run. It was also the easiest Description Logic (DL) reasoner to configure. Jena in conjunction with the pellet reasoner proved capable of processing all the complex axioms specified in the AcontoWeb accommodation ontology, and of returning accurate results based on inferred knowledge of the accommodation domain.

6.2.4 Robustness of Semantic Web Operational Environments

Because the Semantic Web is likely to be built with RDF as a foundation, its success depends on the availability of a scalable and robust infrastructure for storing and accessing RDF data. As noted, a number of semantic middleware applications are available for use as operational environments to support information storage and retrieval. Jena, Sesame[111], RDF Gateway[112], and Kowari[113] to name just a few. Generally, the environments are fairly robust when used to develop applications for a specific domain. Jena provided AcontoWeb with a relatively stable and robust operational environment that supported both RDF and OWL storage, inference, and information retrieval for the accommodation domain.

Despite the inherently distributed nature of the Semantic Web, most current operational environments do not support distributed storage and retrieval of data. In fact, none of the off the shelf products can provide a general technical infrastructure for distributed queries of RDF data. An attempt to build a distributed RDF storage and retrieval system (see

[110] http://jena.sourceforge.net/

[111] http://www.openrdf.org/

[112] http://www.intellidimension.com/

[113] http://www.kowari.org/

Figure 101) is being undertaken by Adamanku and Stuckenschmidt (2005). Their work involves extending the Sesame environment into a multi-server system. Investment on a massive scale is required, however, if the Semantic Web vision as outlined by Tim Berners-Lee (2001) is to become a reality.

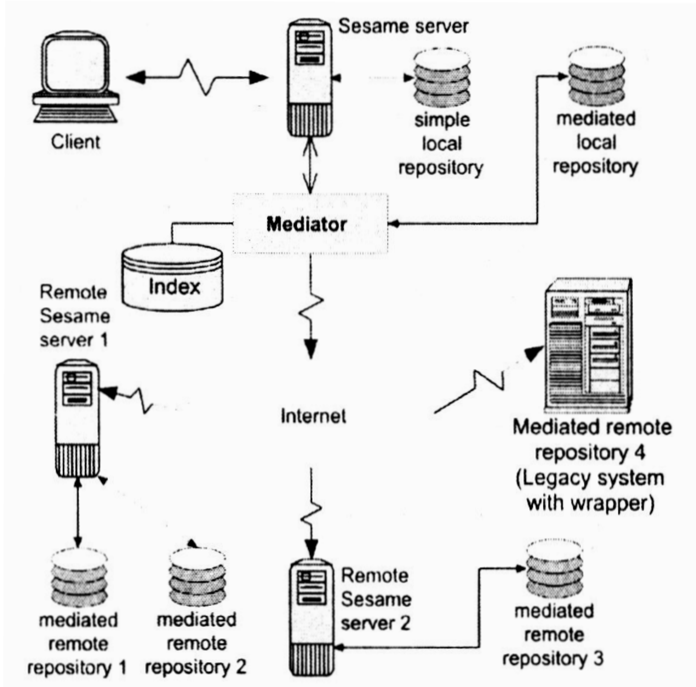


Figure 101: Distributed system architecture (Adamanku & Stuckenschmidt 2005).

6.2.5 How the Semantic Web Can Best be Queried

The original Semantic Web vision, as laid out in 2001 by Berners-Lee (op. cit) and others, was one in which Intelligent agents would be able to crawl a World Wide Web of metadata and exchange information and rules for how to interact with that information, with or without human intervention. In this new Web environment, agents could perform tasks automatically such as schedule appointments, and find information easily without relying on keywords. The agents would be able to search the Web itself providing an interface to the user, rather than using database lookups from a knowledge base to query information in the way that conventional search engines do. Such a vision requires large quantities of machine processable content (e.g. RDF and OWL) available on the Web. At present there is very little. There are a number of shopbots and auction bots on the Web, but these are essentially handcrafted applications and have little ability to interact with heterogenous data and information types.

With the limited Semantic Web infrastructure in place, typical Semantic Web projects of the past five years have demonstrated a distinctive set of characteristics. Shadbolt et al. (2006) explain that typically, they generate new ontologies for the application domain — whether it's information management in breast diseases or computer science research. They either import legacy data or else harvest and redeposit it into a single, large repository. They facilitate semantic integration by using ontologies as mediators. Then they carry out inference on the RDF graphs held within repositories and represent the information using a custom-developed interface. This was the approach taken in designing the AcontoWeb system. It is likely to remain the most practical method to query Semantic Web content until there is substantially more infrastructure (services and RDF data) to support intelligent agent functionality.

6.2.6 Potential Query Results and Accuracy

The research has shown that Semantic Web technologies have the potential to enhance Web search results and accuracy by facilitating more in-depth and precise querying of Web resources. Because conventional Web search engines use keywords for indexing concepts, they are subject to the two well-known linguistic phenomena that strongly degrade a query's precision and recall; 1) Polysemy (one word might have several meanings); and 2) Synonymy (several words or phrases might designate the same concept). These limitations have resulted in significant problems for accessing reliable up-to-date information that urgently need to be solved. By allowing Web authors to explicitly define their words and concepts, the Semantic Web creates an environment in which Web agents are able to analyse the Web on our behalf, making smart inferences that go beyond the simple linguistics performed by today's search engines.

Because intelligent Web agents search the underlying concepts of a Webpage rather than just matching keywords, they are can understand the meaning of information. This allows them to return more relevant and accurate query results to the end user than conventional search engines. The AcontoWeb experiment demonstrated that the principles of semantic search are sound and achievable with presently available technology. The experiment showed how ambiguity of search results can be overcome by searching the underlying concepts of a Webpage. It was also demonstrated that queries could retrieve inferred knowledge about accommodation resorts that had not been explicitly stated on a resort's Website.

6.2.7 How Ontology Based Query Results Compare to those of Conventional Database Systems

The AcontoWeb query experiment compared the complexity of querying the data model of a conventional portal that uses a flat keyword list backed by a relational database for indexing purposes, to a semantic portal that uses a rich domain ontology for indexing. The query results were the same for both data models. Query complexity on the other hand, was reduced in the Semantic Web environment through the use of OWL semantics and inference. A major advantage of using the ontology was that it is dynamic. Classifications specified for each location (e.g. Backpacker destination classification) changed automatically if the attractions or accommodation types associated with a particular location changed. In the relational database environment, destinations classifications need to be updated manually unless additional programming code is implemented at the application level. The use of a transitive property in the accommodation ontology also allowed destination attractions to directly associate with resorts at query run time without the need for equivalent SQL table joins (equi joins).

6.2.8 Usefulness and Limitations of the Semantic Web

The Semantic Web is gaining momentum in both academia and industry. The recent International Semantic Web Conference (ISWC)[114] in Osaka Japan, attracted more than 500 researchers. Major vendors including IBM, Oracle, and Software AG have released or announced Semantic Web based products; and the recent Semantic Technology Conference[115] held in San Jose, California, was an impressive showcase for venture capitalists and executives on the business potential of semantic technologies. Although the increased interest by business and academia is encouraging, so far the Semantic Web vision as laid out by Berners-Lee (op. cit) and others has not eventuated on any real scale. Neither has there been widespread application deployment or the formation of scalable

[114] http://www-static.cc.gatech.edu/gvu/ccg/iswc05/

[115] http://www.semantic-conference.com/

simple systems. The applications that do exist are generally contrived and often consist of examples involving travel, appointments, and sales bookings, and as previously noted, there is little RDF or OWL data available on the Web.

The usefulness of Semantic Web technologies is mainly limited at present to purpose-built domain-specific small scale applications such as AcontoWeb. Semantic Web standards provide the necessary, languages and tools to allow Web based systems to integrate data effectively within a small to medium sized organization or domain. Because of the lack of infrastructure and limited availability of RDF and OWL content, it is proposed here that a less data-centric approach is required for the Semantic Web to succeed on a wider scale. There needs to be more emphasis on system functionality through the creation of Semantic Web services. Exposing functionality in the form of Web services is likely to be more attractive to Internet users and participants than trying to annotate all Web documents worldwide with RDF metadata. The creation of Semantic Web services using standards such as OWL-S (Web Ontology Language for Services) has the potential to one day allow intelligent tourism applications, for example, to be directed to sites offering travel information (i.e. flight availability for a specific airline to a certain location on a certain date) enabling them to automate some of the travel planning and booking processes that currently require human intervention.

6.2.9 Managerial Issues Faced in Gaining User Acceptance of the Semantic Web in the Tourism Industry

Resistance to technical change has long been a major problem in the implementation of new information systems. Bernard (1990) discusses numerous cases of an underlying tension between the control of process and the control of workers during implementation of new computer systems in business. For the Semantic Web to grow there needs to be a significant uptake of the available technologies. Chapter 5 highlighted that previous research (Morrison & King 2002) in the tourism industry indicated a reluctance among tourism enterprises to make effective use of important advances in ICT. This reluctance might also apply to the uptake of Semantic Web technologies.

Despite the bleak prognoses of a few years ago, the research conducted here, which included a survey of tourism operators and analysis of secondary data stakeholder interviews, indicated that there now appears to be a more positive attitude towards the adoption of advanced new online technology by tourism operators. It was reported in Chapter 5 that many respondents were equivocal about using a new technology (the 'Maybe' group). A great many more respondents were receptive to the idea than were against it (only 13 in the 'Unlikely' category against a total of 200 in the 'Likely'/'Definitely' groupings). The survey also indicated that the Semantic Web would have a better chance of being widely accepted in the tourism industry if 1) annotations could be applied at a low cost or even free of charge; 2) annotation software was user friendly; and 3) the technical benefits of using Semantic Web technology for information integration and utilization were communicated effectively to potential users. Tourism operators also expressed a clear preference for any new technology to be added to their existing Website, and that online payment facilities be incorporated into overhauled Websites.

6.2.10 How Successfully Tourism Information can be Integrated on the Semantic Web

A major obstacle for tourism ICT applications is the well-known interoperability problem (Dell'Erbra et al. 2005). Different tourism entities have different views of the world which leads to a plethora of different tourism information systems, each with its own data model and structure. Although tourism is just one small application domain, researchers have naturally identified it as an ideal showcase because of its information heterogeneity, market fragmentation, and rather complex discovery and matchmaking tasks, including substitution and composition — all of which are limitations that Semantic Web technologies promise to overcome (Hepp 2006, p. 85).

The Semantic Web provides the universal standards needed to create common conceptualizations of tourism domains that people or organization can choose to adopt if they wish to make their data interoperable via the Web. Harmo-TEN[116], described in detail in sub-section 2.3.6, is a good example of Semantic Web technologies successfully used for the integration of online tourism information. The Harmo-TEN solution allows any actor to map their data model at the conceptual level to a common exchange data

[116] http://www.harmo-ten.info/

model. The Harmonise ontology allows the individual or organization to communicate and interoperate with all other tourism actors who have done the same. At present there are twelve participating tourism bodies involved with Harmo-TEN. The project has demonstrated that tourism information can be effectively integrated using Semantic Web technologies in a real-world setting. The AcontoWeb system, while not yet up and running in a real industry setting, also demonstrated the successful integration of online tourism information. In the AcontoWeb experiment, resort facilities were queried using underlying concepts when different keywords were actually used to describe the facilities on Websites (e.g. Conference Facilities and Convention Centre).

6.3 Answer to the Major Research Question and Proposition of a Grounded Hypothesis

The major research question was defined in Chapter 1 as:

To what extent can the Semantic Web and related technologies assist with the creation, capture, integration and utilization of accurate, consistent, timely, up-to-date Web based tourism information?

The exploratory nature of the research means that the answer to the above question is in itself grounded theory. The grounded theory was established through a comprehensive investigation that included a review of all available literature, a process of system development and experimentation, and a survey of tourism operators supported by secondary data interviews. Based on the findings of this investigation, the answer to the major research question and the proposition of a grounded hypothesis are stated as follows:

The Semantic Web provides the necessary standards, languages, and tool development support for building applications that can integrate and utilize Web based tourism information more effectively than the current Internet allows. More specifically, Semantic Web technologies facilitate ontology based annotations that describe precisely the meaning of certain parts of a Website so that advanced applications such as Web search agents can reason more effectively about this information. The AcontoWeb system demonstrated that the Website of a hotel could be suitably annotated to distinguish *between hotel name, location, category, available facilities, its destination type, and associated attractions etc. This enabled information to be effectively integrated and processed using the AcontoWeb semantic search tool. The survey and secondary data provided an understanding of the managerial issues faced in gaining wider acceptance of the Semantic Web in tourism. This component of the research indicated that there is a positive attitude towards the adoption of new online technologies among tourism operators, provided the technical benefits are well communicated.*

Unfortunately, the limitations of the Semantic Web at present, which primarily relate to the difficulties of RDF knowledge representation, change management issues, and a lack of global infrastructure supporting distributed operational environments, mean that its usefulness for tourism information integration and utilization will be limited in the short to medium term (next 4 or 5 years) to well-managed, strictly-controlled environments such as Harmo-TEN or AcontoWeb. Beyond the short to medium term, however, the need for greater information interoperability on the Web will see Semantic Web standards (in whatever form they may evolve to) being widely used to assist intelligent Web agents in carrying out sophisticated tasks on behalf of Internet users such as, "Arrange a one-week holiday, somewhere near the Great Barrier Reef Queensland (Australia), during September. Services like 'Car Hire' and 'Airline Bookings' are also likely to be automated by such systems via Semantic Web technologies.

6.4 Findings in Relation to Research Aims

The following research objectives have been realized:

1) An understanding has been provided of the issues and problems involved in defining, establishing, capturing, integrating and using the heterogeneous, scattered and diverse supplier source data necessary for the development of Semantic Web based tourism applications. This was achieved by investigating available Semantic Web technologies, standards, and development tools, and using them to develop AcontoWeb, which is an annotation tool and semantic portal system for accommodation Websites.

2) A theoretical and conceptual solution to the data-related problems named above was specified to address the technical limitations of existing Web-based integration approaches by taking into account the critical social dimension. This solution is represented by the design of the AcontoWeb architecture and incorporates tourism operator preferences.

3) The research has succeeded in developing a proof of concept DMS prototype (based on the conceptual model discussed above), restricted to matching tourism customers' accommodation needs to suppliers' offerings. This prototype (titled *AcontoWeb)* is 'ontology-driven', and allows accommodation Website owners to conveniently annotate their Websites with RDF metadata in accordance with the constructs of the domain ontology. The query component also allows the tourism customer to query Websites based on inferred knowledge of the accommodation domain.

4) The effectiveness of the DMS, with regard to usability and value-adding potential for tourism industry customers and service providers, was demonstrated via an experiment that compared the complexity and subsequent ease of information integration of querying the data model of a conventional portal to that of a semantic portal.

5) Insight has been gained into the attitudes towards the adoption of semantic Web technology by SMTEs and their requirements and preferences for the implementation and usability of such systems. This insight was gained through analysis of the tourism operator survey and secondary data interviews obtained from industry stakeholders.

6) A grounded hypothesis has been proposed about the extent to which the Semantic Web and related technologies can assist with the creation, capture integration, and utilization of accurate, consistent, timely, up-to-date Web based information. The grounded theory can now be tested in further research.

6.5 Future Research Directions

A number of potential areas for future research have been identified throughout the thesis. Firstly, Better knowledge representation formalisms are clearly required if the Semantic Web is achieve widespread uptake. Current formalisms create significant barriers to adoption because manual annotation of Web documents with RDF metadata is inefficient and problematic. Automatic generation of metadata by means of semi-automated annotation and text mining promise much. These techniques, however, are not yet mature and are prone to numerous errors (McCool 2005). Improving these processes is vital if the Semantic Web is to succeed.

The continued development of Semantic Web services is also crucial. It is the strong opinion of the researcher that a more service-oriented approach to building the Semantic Web is required, rather than the data-centric approach that has largely been the focus to date. With an emphasis on system and application functionality, Semantic Web services based on the OWL-S standard could be used effectively for automated, discovery, composition and orchestration of information for business functions such as dynamic tourism product packaging. Work in this area is already being undertaken by Cardoso et al. (2005). Much more work is required, though, before Semantic Web services are widely available. At present there are few.

It was noted in the Methodology (Chapter 3) that AcontoWeb is somewhat unique because its modularized architecture allows any OWL DL ontology to be plugged into the Jena supported backend, reasoned over, and then queried using the SPARQL query language. Work is continuing on this project to evolve the AcontoWeb backend components into Web services that could be utilized by other remote systems. The idea is to provide generic Web based semantic middleware capable of performing reasoning and query functions for any remote application that may wish to tap into the service. Such an initiative, if successfully implemented, would represent significant research in the area of the Semantic Web. It could provide substantial benefits to industries such as tourism by facilitating greater information Interoperability by way of easy access to reasoning and query services.

Further development and refinement of the AcontoWeb prototype is continuing as part of the Phoenix[117] research program. It is planned that this extended research will include a prototype evaluation study incorporating tourism operators. This study will emphasize and strengthen the link between the two research components documented in this thesis (i.e. prototype development and survey).

Substantial other challenges remain for the Semantic Web and its application to tourism ICT systems. These challenges present a myriad of opportunities for further valuable research in the topic area. For instance: how can huge numbers of decentralized information repositories of varying scales be queried? Or, how can a semantic browser be developed that can effectively navigate and visualise large RDF graphs? And, how can the problems associated with ontology versioning be adequately dealt with? These are just some of the issues that need further investigation before the Semantic Web reaches it full potential of revolutionizing areas such as tourism e-commerce.

6.6 Chapter 6 Summary

The chapter presented the research findings and proposed a grounded hypothesis about the usefulness of the Semantic Web for online tourism information integration and utilization. At the beginning of the thesis, the research problem was categorized into three distinct parts that should be viewed as follows: 1) there are a number of limitations associated with the current Internet; that 2) create significant challenges for information systems integration; which 3) have negative consequences for tourism ICT applications. Traditional approaches to data integration were essentially ‘top-down’, in that they are driven by senior management, or even governments or industry bodies. It was emphasized that while seeming to make sense theoretically, the evidence strongly suggests that these approaches do not work in practice (Markus & Tanis 2000). The AI

[117] http://www.staff.vu.edu.au/PHOENIX/phoenix/index1.htm

literature suggested that a better solution may lie with a bottom-up Semantic Web approach. It was the benefits and limitations of such an approach that the thesis set out to investigate.

The research was conducted by following a systems development methodology (Burstein, 2002) to generate grounded theory about the extent to which the Semantic Web and related technologies assist with the creation, capture, integration and utilization of accurate, consistent, timely, up-to-date Web based tourism information. The systems development process was supplemented with a survey of tourism operators designed to provide an understanding of the attitudes towards the adoption of advanced new online technologies within the industry. It was concluded from the investigation that Semantic Web technologies provide the necessary standards, markup languages and development tool support for building applications that can integrate and process online tourism information more effectively than the current Internet allows. It was also concluded, however, that the usefulness of the Semantic Web for tourism ICT applications is likely to be limited in the short to medium term (next 4 or 5 years) to well managed strictly-controlled environments.

On the positive side, the theory was proposed that beyond the short to medium term, the need for greater information interoperability on the Web will see Semantic Web standards (in whatever form they may have evolved to) being widely used to assist intelligent Web agents in carrying out sophisticated tasks on behalf of Internet users. The creation of widely available Semantic Web services, and easier forms of knowledge representation such as automatic creation of metadata by means of text mining and semi-automated annotation, were identified as potential areas for future research.

In closing, it is emphasized that the Semantic Web should not be viewed as a separate Web, but rather as Berners-Lee et al. (op cit.) described it, as an extension of the current one, in which information and services are given well-defined meaning, thereby better enabling computers and people to work in cooperation. Dealing with heterogeneity has continued to be a key challenge since it was made possible to exchange and share data between computers and applications over the Internet. The tourism industry has been particularly affected by this heterogeneity because of its market fragmentation, and rather complex discovery and matchmaking tasks, including substitution and composition.

Semantic Web standards offer the means to define information on the Web so that it can be used by computers not only for display purposes, but also for interoperability and integration between systems and applications, thus resolving heterogeneity problems. Various languages, development tools and applications were presented in this thesis that are capable of facilitating semantic integration of tourism information sources. These technologies form the technological foundations of the Semantic Web, and are additions to the current Web that are freely available for individuals or organisations who may wish to use them to their advantage.

REFERENCES

Abrahams, B. & Dai, W. 2005a, 'Architecture for automated annotation and ontology based querying of Semantic Web resources', paper presented to IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne, France, September 19-22, 2005.

---- 2005b, 'Meeting Semantic Web challenges with automated annotation and multi-agent querying of Web resources'', paper presented to Victoria University Business Research Conference, Melbourne, Australia.

ABS 2002, *Accommodation Industry Australia, Report No. 8695.0*, Australia Bureau of Statistics:, Canberra.

Adamanku, G. & Stuckenschmidt, H. 2005, 'Implementation and evaluation of a distributed RDF storage and retrieval system', paper presented to IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne, France, September 19-22, 2005.

Alesso, P. 2004, 'Semantic search technology', *SIGSEMIS Bulletin*, vol. 1, no. 3, pp. 86-98.

Alesso, P. & Smith, C. 2004a, 'The Semantic Web', in *Developing Semantic Web Services*, AK Peters, Wellesey, MA, USA, pp. 165-84.

---- 2004b, 'Challenges and opportunities', in *Developing Semantic Web Services*, AK Peters, Ltd., Wellesey, MA, USA, pp. 409-17.

---- 2004c, 'Web services', in *Developing Semantic Web Services*, AK Peters Ltd., Wellesey, MA, pp. 121-62.

---- 2004d, 'Semantic search', in *Developing Semantic Web Services*, A K Peters Ltd., Wellesey, MA, pp. 387-406.

---- 2004, 'Markup Languages', in *Developing Semantic Web Services*, AK Peters, Wellesey, MA, USA, pp. 31-43.

---- 2004e, 'The World Wide Web', in *Developing Semantic Web Services*, A K Peters Ltd, Wellesey, MA, pp. 3-29.

Antoniou, G., Skylogiannis, T., Bikakis, A. & Bassiliades, N. 2005, 'A semantic brokering system for the tourism domain', *Information Technology & Tourism*, vol. 7, no. 3/4, pp. 183-200.

Applebee, A. & Richie, B.W. 2000, *The ACT tourism Internet study: attitudes, perceptions and adoption*, University of Canberra.

ATDW 2001, 'Taking Australian tourism to the world', paper presented to Presentation to the TRAVELtech Expo, Sydney, Australia, August 28.

Benjamins, V. & Gomez-Perez, A. 1999, *Overview of knowledge sharing and reuse components: ontologies and problem solving methods*, Amsterdam, the Netherlands.

Benjamins, V., Contreras, J., Corcho, O. & Gómez-Pérez, A. 2004, 'Six challenges for the Semantic Web', *SIGSEMIS Bulletin*, vol. 1, no. 1, pp. 24-5.

Bergamaschi, S., Deomenico, B., G, F. & Maurizio, V. 2005, 'Building a tourism information provider with the Momis system', *Information Technology & Tourism*, vol. 7, no. 3/4, pp. 221-38.

Berger, S., Bry, F., Furche, T. & Linse, B. 2005, 'The Web and Semantic Web query language Xcerpt', *Semantic Web Fact Book 2005*, pp. 99-102.

Bernard, E. 1990, *Management resistance to change: case study results in the introduction of computer information systems*, Harvard University, viewed 26/05/2006 <http://www.law.harvard.edu/programs/lwp/eb/minn.pdf>.

Berners-Lee, T., Hendler, J. & Lassila, O. 2001, 'The Semantic Web', *Scientific American*, May, pp. 34-43.

Bernstein, A., Kaufmann, E., Kaiser, C. & Kiefer, C. 2006, 'Ginseng: a guided input natural language search engine for querying ontologies', paper presented to First Jena User Conference, Bristol, UK.

Blake, S. 1978, *Managing for responsive research and development.*, W.H. Freeman and Company, San Francisco.

Bloodsworth, P. & Greenwood, C. 2005, 'Ontology-centric multi-agent systems in 2005', *Semantic Web Fact Book 2005*, pp. 90-4.

Borst, W. 1997, *Construction of engineering ontologies*, Centre for Telematica and Information Technology, Enschede, Netherlands.

Bray, T., Paoli, J., SPerberg-McQueen, C. & Maler, E. 2004, *Extensible markup language (XML) 1.0 W3C recommendation*, viewed 20/04/2006 <http://www.w3.org/TR/REC-xml/>.

Brickley, D. & Guha, R. 2003, *RDF vocabulary description language 1.0: RDF schema. W3C working draft*, viewed 12/01/2006 <http://www.w3.org/TR/PR-rdf-schema>.

Burstein, F. 2002, 'Systems development in information systems', in *Research Methods for Students, Academics and Professionals*, 2nd edn, Centre for Information Studies, Charles Sturt University, Wagga Wagga, NSW, pp. 147-58.

Burstein, M., Bussler, C., Finin, T., Huhns, M., Paolucci, M., Sheth, A., Williams, S. & Zaremba, M. 2005, 'A Semantic Web services architecture', *IEEE Internet Computing*, vol. 9, no. 5, pp. 72-81.

Bylander, P., Bryant, B., Ide, N., Pareja-Lora, A. & Wilcock, G. 2003, *The roles of natural language and XML*, Language and Linguistics, Academia Sinica, Taiwan.

Calvanese, D., Giacomo, G., Lembo, D., Lenzerini, M. & Rosati, Y. 2006, 'Data complexity of query answering in description logics', paper presented to Proc. of the 10th Int. Conf. on the Principles of Knowledge Representation and Reasoning, UK, June 2-5.

Cardoso, J., Jorge, D. & Fernandes, D. 2005, 'SEED, semantic e-tourism dynamic packaging', *Semantic Web Fact Book 2005*, pp. 58-60.

Castellanos, D. & Fernández, T. 2004, *Using semantic technologies to support evaluation processes in e-Learning*, viewed 3 1, <http://www.sigsemis.org/>.

Cerez-Kecmanovic, D. 1994, 'Engineering type information systems research', paper presented to In Proceedings of the 5th Australian Information Systems Conference, Department of Information Systems, Monash University: Caulfield Vic.

Chebotko, Lu, S. & Fotouhi 2004, 'Challenges for information systems towards the Semantic Web', *SIGSEMIS Bulletin*, vol. 1, no. 1, pp. 26-8.

Codd, E. 1970, 'A relational model of data for large shared data banks', *Communications of the ACM*, vol. 13, no. 6, pp. 377-87.

Cristani, M. & Roberta, C. 2005, 'A survey on ontology creation methodologies', *International Journal on Semantic Web and Information Systems*, vol. 1, no. 2, pp. 49-69.

Dai, W. & Abrahams, B. 2005, 'A multiagent architecture for Semantic Web resources', paper presented to IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Compiegne, France.

Damjanoviæ, V., Devedžiæ, V., Djuriæ, D. & Gaševiæ, D. 2004, 'Framework for analysing ontology development tools', *SIGSEMIS Bulletin*, vol. 1, no. 3, pp. 43-7.

Daniele, R., Misitilis, N. & Ward, L. 2000, 'Partnership Australia's national tourism data warehouse: preliminary assessment of a destination marketing system', in *Information and Communication Technologies in Tourism 2000*, Springer, Vienna, Austria.

Decker, S., Mitra, P. & Melnik, S. 2000, 'Framework for the semantic Web: an RDF tutorial', *IEEE Internet Computing*, vol. 4, no. 6, pp. 687-73.

Dell'Erbra, M., Fodor, O., Hopken, W. & Werthner, H. 2005, 'Exploiting semantic Web technologies for harmonizing e-markets', *Information Technology & Tourism*, vol. 7, no. 3/4, pp. 201-19.

Denny, M. 2002, *Ontology building: a survey of editing tools*, 20/04/2006, <http://www.xml.com/pub/a/2002/11/06/ontologies.html?page=1>.

Ding, L., Finin, T., Joshi, A., Peng, Y., Pan, R. & Reddivari, P. 2005, 'Search on the Semantic Web'.

Donaldson Dewitz, S. 1996, 'The process of systems development two paradigms', in *Systems Analysis and Design and the Transition to Objects*, McGraw-Hill, Singapore, pp. 92-117.

El Sawy, O. 2001, *Redesigning enterprise processes for e-business*, McGraw-Hill, Boston, MA.

Eysenback, G. 2003, 'The Semantic Web and healthcare consumers: a new challenge and opportunity on the horizon?' *International Journal on Healthcare Technology and Management*, vol. 5, no. 3/4/5, pp. 194-212.

Fernandez-Lopez, M., Gomes-Perez, A. & Juritso, N. 1997, 'Methontology: from ontology art towards ontology engineering', paper presented to Spring Symposium on Ontological Engineering of AAAI, Stanford University, California, USA.

Ford, P. 2004, *Semantic Web roundup*, viewed 10/03/2006 <http://www.xml.com/pub/a/2004/05/26/www2004.html>.

Fotiodis, T., Vassiliadis, C., Hatzithomas, L. & Gkotzamanis, E. 2005, 'An IT approach brand positioning confusion on hospitality enterprises: the case of the Greek Islands.' in *Information and Communication Technologies in Tourism*, Springer, Vienna, Austria, pp. 371-82.

Furche, T. 2004, *Survey over existing query and transformation Languages*, viewed 11/01/2006 <http://rewerse.net/deliverables/i4-d1.pdf>.

Galliers, R. 1991, 'Choosing appropriate information systems research approaches: a revised taxonomy', in *Information Systems Research: Contemporary Approaches and Emerging Traditions*, B.V, Noth Holland, pp. 327-45.

García, E. & Sicilia, M. 2003, 'User interface tactics in ontology-based information seeking', *PsychNology*, vol. 1, no. 3.

Geroimenko, V. & Chen, C. 2006, *Visualizing the Semantic Web*, Springer, London, UK.

Glaser, A. 1967, *The discovery of grounded theory: strategies for qualitative research*, Aldine, Chicago, USA.

Gomes-Perez, A., Fernandez-Lopez, M. & Corcho, O. 2004a, 'Methodologies and methods for building ontologies', in *Ontological Engineering*, Springer, London, UK, pp. 107-97.

---- 2004b, 'The most outstanding ontologies', in *Ontological Engineering*, Springer, London, UK, pp. 47-106.

---- 2004c, 'Theoretical Foundations of Ontologies', in *Ontological Engineering*, Springer, London, UK, pp. 3-45.

---- 2004d, 'Ontology tools', in *Ontological Engineering*, Springer, London, UK, pp. 293-362.

Gruber, T. 1993a, 'A translation approach to portable ontology specifications', *Knowledge Acquisition*, vol. 6, no. 2, pp. 199-221.

---- 1993b, 'A translation approach to portable ontology specification', *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220.

Gruininger, M. & Fox, M. 1995, 'Methodology for the design and evaluation of ontologies', paper presented to Skuce D (ed) IJCAI95 Workshop on Basic Ontological issues in Knowledge Management Sharing.

Guarino 1998, 'Formal ontology in information systems', paper presented to 1st International Conference on Formal Ontology in Information Systems (FOIS'98), Trento, Italy.

Handschuh, S., Staab, S. & Maedche, A. 2001, 'Cream - creating relational metadata with a component-based, ontology driven annotation framework', paper presented to First International Conference on Knowledge Capture, Victoria, Canada.

Handschuh, S., Staab, S. & Volz, R. 2003, 'On deep annotation', paper presented to 12th International World Wide Web Conference, Budapest, Hungary, 20-24 May.

Hepp, M. 2006, 'Semantic Web and Semantic Web services', *IEEE Internet Computing*, vol. 10, no. 2, pp. 85-8.

Hitch, C. & McKean, R. 1960, *The economics of defence in the nuclear age*, Harvard University Press, Cambridge Ma.

Hopken, W. 2002, 'Analysis of tourism standards', paper presented to 2nd Harmonise Workshop, 22 January.

Horridge, M. 2004, *A Practical guide to building OWL ontologies using the Protege-OWL plugin and CO-ODE tools edition 1.0*, The University Of Manchester, viewed 5/10/2005 <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>.

Horrocks, I. 2000, *A denotional semantics for OIL-Lite and standard OIL. Technical report*, viewed 16//09/2006 <http://www.cs.man.ac.uk/~horrocks/OIL/Semantics/>.

Horrocks, I. & Tessaris, S. 2000, 'A conjunctive query language for description logic ABoxes', paper presented to National Conference on Artificial Intelligence AAAAI/IAAI 2000.

Horrocks, I. & Harmelen, F. 2001, 'Reference description of the DAML + OIL (March 2001) ontology markup language. Technical report.'

Howe, W. 2005, *A brief history of the internet*, viewed 17/02/2006 <http://www.walthowe.com/navnet/history.html>.

Hsu, C. 1996, 'Enterprise integration and modelling: the meta database approach, kluwer', Norwell, MA.

Hyv¨onen, E., Salminen, M. & Kettula, S. 2004, 'A content creation process for the Semantic Web', paper presented to OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments, Lisbon, Portugal, May 29.

Jakoniene, V. 2003, *Ontology integration*, <http://www.ida.liu.se/labs/iislab/courses/LW/slides/ontologyIntegration.pdf>.

Jansen, B. 2000, *The effect of query complexity on Web searching results*, viewed 04/06/2006 <http://informationr.net/ir/6-1/paper87.html>.

Jones, S. 1987, 'Choosing action research: a rationale', in *Organisation Analysis and Development*, Wiley, Sussex, UK.

Kactus 1996, *The Kactus booklet version 1.0. Esprit Project 8145 Kactus*, viewed 31/01/06 <http://www.swi.psy.uva.nl/projects/NewKactus/Reports.html>.

Klein, M. & Fensel, D. 2001, *In Proceedings of the international Semantic Web working symposium (SWWS)*, Stanford University, California, USA.

Klinker, G., Bhola, C., Dallemagne, G., Marques, D. & McDermott, J. 1991, 'Usable and reusable programming constructs.' *Knowledge Acquisition*, vol. 3, pp. 117-36.

Klischewski, R. & Jeenicke, M. 2004, 'Semantic Web technologies for information management within e-governement', paper presented to Proceeding of the 37th Hawaii International Conference on System Sciences., Hawaii.

Kolaitis, P. & Vardi, M. 1998, 'Conjunctive-query containment and constraint satisfaction', *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pp. 205-13.

Kotte, J. & Schlesinger, L. 1979, *Six Change approaches*, viewed 25/04/2006 <http://www.valuebasedmanagement.net/methods_kotter_change_approaches.html>.

Kuhn, T. 1970, *The structure of scientific revolutions*, 2nd edn, University of Chicago Press, Chicaga, USA.

Lara, R., Han, S., Lausen, H., Stollberg, M., Ding, Y. & Fensel, D. 2004, 'An evaluation of Semantic Web portals', paper presented to IADIS Applied Computing Conference, Lisbon, Portugal, 23-26 March.

Lassila, O. & McGuinness, D. 2001, *The role of frame based representation on the Semantic Web*, Stanford University, Stanford, California.

Lausen, H., Stollberg, M., Hernández, R., Ding, Y., Han, S. & Fensel, D. 2003, *Semantic Web portals – state of the art survey*, IFI – Institute for Computer Science, University of Innsbruck, Innsbruck, Austria.

Lederer, A.L. & Sethi, V. 1992, 'Meeting the challenges of information systems planning', *Long Range Planning*, vol. 25, no. 2, pp. 69-80.

Leiner, B., Vinton, G. & Clark, D. 2003, *A brief history of the internet*, viewed 18-04-2006 <http://www.isoc.org/internet/history/brief.shtml>.

Lenat, D.B. & Guha, R.V. 1990, *Representation and inference in the Cyc project*, Addison-Wesley, Bostone, Massachusetts, USA.

Makela, E., Hyv¨onen, E., Saarela, S., & Viljanen, K. (2004). ONTOVIEWS - A tool for creating semantic web portals. University of Helsinki. Retrieved January 8, 2007, from http://whitepapers.zdnet.co.uk/0,39025945,60117603p-39000589q,00.htm

Maedche, A. & Staab, S. 2002, 'Applying Semantic Web technologies for tourism information systems', paper presented to 9th International Conference for Information and Communication Technologies in Tourism (ENTER 2002), Innabruck, Austria, January 23-25.

Manola, F. & Miller, E. 2004, *RDF Primer, W3C Recommendation*, viewed 07/12/2005 <http://www.w3.org/TR/rdf-primer/>.

Markus, M.L. & Tanis, C. 2000, 'The enterprise system experience - from adoption to success', in *Framing the Domains of IT research: Glimpsing the Future Through the Past*, Pinnaflex Educational Resources Inc., Cincinnati, OH.

McCarthy, P. 2005, *Search RDF data with SPARQL*, viewed 24/04/2006 <http://www-128.ibm.com/developerworks/xml/library/j-sparql/>.

McCool, R. 2005, 'Rethinking the semantic Web, part 1', *IEEE Internet Computing*.

McGrath, M. & Moore, E. 2003, 'Information integration within the Australian tourism industry: a proposed approach', paper presented to in Proceedings of the 17th Annual Conference of the Australian & New Zealand Academy of Management., Edith Cowan University, Perth, December 2-5.

McGrath, M. & Abrahams, B. 2006a, 'AcOntoWeb: a Semantic portal for the tourism and hospitality industry', paper presented to Hospitality Information Technology Association (HITA'06), Minneapolis, USA, June 18 - 19.

---- 2006b, 'Ontology-based website generation and utilization for tourism services', *Information Technology in Hospitality*, vol. 4.

McGrath, M., Moore, E. & Abrahams, B. 2005a, 'Attitudes towards online technology among Australian accommodation enterprise operators: a preliminary study', paper presented to Tourism Enterprise Strategies Conference (TES2005), Melbourne Australia, 11-12 July.

McGrath, M., Abrahams, B. & Moore, E. 2005b, 'Online technology use and adoption among Australian accommodation enterprise operators', paper presented to Proceedings of the 19th Annual ANZAM Conference, Canberra, Australia, 7-10 December.

McGrath, M., Abrahams, B. & More, E. 2006, 'Potential use of advanced online technologies among Australian accommodation sector operators', paper presented to Proceedings of the 13th International Conference on Information Technology in Travel and Tourism (ENTER2006), Lausanne, Switzerland, 18–20 January.

McGrath, M., Carson, D., Debenham, J., King, B., Meijerink, H., More, E. & Sandy, G. 2005, *A high-level architecture for the Australian tourism industry*, Sustainable Tourism Cooperative Research Centre, Brisbane, Australia.

McGuinness, D. & Harmelen, F. 2004, *OWL Web ontology language semantics and abstract syntax, W3C recommendation 10 February 2004*, viewed 07/12/2005 <http://www.w3.org/TR/owl-absyn/>.

McGuinness, D. & Van Harmelen, F. 2004, *Web ontology language (OWL)*, viewed 20/0402006 <http://www.w3.org/2004/OWL/>.

Mendes, O. 2003, *État de l'art sur les méthodologies d'ingénierie ontologique*, Centre de recherche LICEF Montréal, Québec, Québec, Canada.

Miller, L. 2001, *RDF Squish query language and Java implementation*, viewed 04/04/2006 <http://ilrt.org/discovery/2001/02/squish/>.

Mills, J. & Morrison, A. 2003, *Measuring customer satisfaction with online travel*, In Information and communication technologies in tourism 2003, Springer, Vienna, Austria.

Missikoff, M., Werthner, H., Hopken, W., Dell'Orba, M., Fodor, O., Formica, A. & Taglino, F. 2003, 'Harmonise: towards interoperability in the tourism domain', *Information and Communication Technologies in Tourism*, pp. 58-66.

Mistilis, N., Presbury, R. & Agnes, P. 2004, 'The strategic use of information technology in marketing and distribution - a preliminary investigation of Sydney hotels', *Forthcoming in The Journal of Hospitality and Tourism Management*.

Mitchell, L. 2003, 'Bed, brecky and red tape', *The Age*, September 16, pp. 4-5.

Mizoguchi, R., Vanwelkenhuysen, J. & Ikeda, M. 1995, 'Task ontology for reuse of problem solving knowledge', in *Mars N (ed) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS'95).* IOS Press, Amsterdam, The Netherlands.

Morgan, R. 2003, *Online travel takes off, finding No. 3676*.

---- 2004, *Online travel takes off, finding No. 3811*.

Morrison, A. & King, M. 2002, 'Small tourism businesses and e-commerce: Victorian tourism online', *Tourism and Hospitality Research*, vol. 4, no. 2, pp. 104-15.

Neches, R., Fikes, R., Finin, T., Gruber, T., Senator, T. & Swartout, W. 1991, 'Enabling technology for knowledge sharing.' *AI Magazine*, vol. 12, no. 3, pp. 36-56.

Neuman, W. 1994, *Social research methods: qualitative and quantitative approaches*, 2nd edn edn, Allyn and Bacon, Needham Heights, Ma.

Nolan, R.L., Puryear, C.R. & Elron, D.H. 1989, 'The hidden barriers to the Bell operating companies and their regional holding companies' competitive strategies', in *Future Competition in Telecommunications*, M.M Parker ed. Harvard School of Business Press, Boston, MA, pp. 301-27.

Noy, N.F. & Musen, M.A. 2002, 'Evaluating ontology mapping tools: requirements and experience', paper presented to Workshop on Evaluation of Ontology-based Tools (EON2002), Siguenza, Spain.

Nunamaker, J. & Chen, C. 1990, 'Systems development in information systems research', paper presented to In Proceedings of the 23rd Hawaii international Conference on Systems Science, Los Alomitos, Ca pp. 631-639.

Nunamaker, J., Chen, M. & Purden, T. 1990-1991, 'Systems development in information systems research', *Journal of Management Information Systems Research*, vol. 7, no. 3, pp. 89-106.

Oberle, D., Staab, S., Struder, R. & Volz, R. 2005, 'Supporting application development in the Semantic Web', *ACM Transactions on Internet Technology*, vol. 5, no. 2.

Ogbuji, C. 2005, *Versa: path-based RDF query language*, viewed 04/04/2006. <http://www.xml.com/pub/a/2005/07/20/versa.html>

Palmer, S. 2002, *RDF in HTML: approaches*, viewed 05/04/2006 <http://infomesh.net/2002/rdfinhtml/>.

Park, J. 1998, 'Mappings for reuse in knowledge-based systems', paper presented to In: Gaines BR, Musen MA (eds) 11th International Workshop on Knowledge Acquisition, Modelling and Management (KAW'98). Banff, Canada.

Parker, C., Wafula, E., Swatman, P. & Swatman, P. 1994, 'Information systems research methods', paper presented to In Proceedings of the 5th Australian Conference on Information Systems, Monash University, Department of Information Systems, Victoria, Australia.

Parker, D. 2003, *Surfing a big online-up of fast breaks*.

Petrie, C. 2006, 'Semantic Web and Semantic Web services: father and son or indivisible Twins?' *IEEE Internet Computing*, vol. 10, no. 2, pp. 85-8.

PhoCusWright 2003, *Hotel and lodging commerce 2002-2005: distribution strategies and market forecasts*.

Poirer, C. & Bauer, M. 2001, *E-supply chain: using the internet to revolutionize your business.*, Berrett-Koehler Publishers, San Francisco, USA.

Pollard , D. 2004, *Knowledge integration leading to personal knowledge management*, Knowledge Management Blog - The Feryman, viewed 31/10/2006 <barryhardy.blogs.com/theferryman/2004/06/knowledge_integ.html>.

Priebe, T., Kiss, C., & Kolter, J. (2005). Semiautomatische Annotation von Textdokumenten mit Semantischen Metadaten. In *Paper Presented to Sixth Internationale Tagung Wirtshaftsinformatik (WI 2005)*, Bamber, Germany.

Prud'hommeaux, E. & Seaborne, A. 2005, *SPARQL query language for RDF*, viewed 02/02/2005 <http://www.w3.org/TR/rdf-sparql-query/>.

Qin, L. 2005, 'Change detection and management for the Semantic Web', *Semantic Web Fact Book 2005*, p. 96.

Reynolds, D. 2001, *SWAD-Europe deliverable 12.1.5: semantic portals - requirements Specification*, viewed 18/01/2006 <http://www.w3.org/2001/sw/Europe/reports/requirements_demo_2/>.

Reynolds, D., Shabajee, P. & Cayzer, S. 2004, *Semantic information portals*, Hewlet-Packard, <http://whitepapers.zdnet.co.uk/0,39025945,60104230p-39000536q,00.htm>.

Ricci, F. 2002, 'Travel recommender systems', *Computer.org/intelligent*, NOVEMBER/DECEMBER 2002.

Rumbaugh, J., Jacobson, I. & Booch, G. 1998, *The unified modelling language regerence manual*, Addison-Wesley, Boston, Massachusetts.

Sandy, G. & Burgess, S. 2003, ' A decision chart for small business Web site content', *Logistics Information Management*, vol. 16, no. 1, pp. 36-47.

Schaffert, S. & Bry, F. 2004, 'Querying the Web reconsidered: a practical introduction to Xcerpt', *Proceedings of the Extreme Markup Languages*.

Seaborne, A. 2004, *RDQL - A query language for RDF*, viewed 01/02/2005 <http://www.w3.org/Submission/RDQL/>.

Shadbolt, N., Hall, W. & Berners-Lee, T. 2006, 'The semantic Web revisited', *IEEE Internet Computing*, pp. 96-101.

Sharma, P., Carson, D. & DeLacy, T. 2000, 'Developing a business information data warehouse for the Australian tourism industry - a strategic response', in *information and Communication Technologies in Tourism 2000*, Springer, Vienna, Austria, pp. 147-56.

Sheth, A., Ramakrishnan, C. & Thomas, C. 2005, 'Semantics for the Semantic Web', *International Journal on Semantic Web and Information Systems*, vol. 1, no. 1, pp. 1-35.

Sheth, A., York, W., Kochut, K. & Miller, J. 2005, 'Bioinformatics for glycan expression: integrated technology resource for biomedical glycomics', *Semantic Web Fact Book*, pp. 73-4.

Singh, R. & Iyer, L. 2003, 'Web Service for knowledge management in e-marketplaces', *eService Journal*, vol. 3, no. 1.

Singh, R. & Murshed, A. 2005, *Evaluation and ranking of ontology construction tools*, University of Trento.

Singh, R., Lakshmi, I. & Salam, A. 2005, 'Semantic eBusiness', *International Journal on Semantic Web and Information Systems*, vol. 1, no. 1, pp. 19-35.

Song, H., Giri, S. & Ma, F. 2004, 'Data extraction and annotation for dynamic Web pages', paper presented to IEEE Int'l Conf, e-Technology, e-Commerce, and e-Service.

Staab, S. 2005, 'Introduction to the special theme', *Information Technology & Tourism*, vol. 7, no. 3/4, pp. 181-238.

STCRC 1999, *Meeting the challenge*, viewed 25/08/2004 <http://www.crctourism.com.au/>.

Stojanovic, L., Stojanovic, N. & Voltz, R. 2002, 'Migrating data-intensive Web sites into the Semantic Web', *Proc. ACM Symp. Applied Computing (SAC 02)*, pp. 1100-7.

Stojanovic, N., Maedche, A., Staab, S., Studer, R. & Sure, Y. 2001, 'SEAL: A framework for developing semantic portals', paper presented to First International Conference on Knowledge Capture, Victoria, British Columbia, Canada.

Strauss, A.L. 1987, *Qualitative analysis for social sciences*, Cambridge University Press, Cambridge, UK.

Strauss, A.L. & Corbin, J. 1994, 'Grounded theory methodology: an overview', in *Handbook of Qualitative Research*, Thousand Oaks, CA, USA, pp. 273-85.

Struder, R., Benjamins, V. & Fensel, D. 1998, 'Knowledge engineering: principles and methods.' *IEEE Transactions on Data and Knowledge Engineering*, vol. 25, no. 1-2, pp. 161-97.

Stuckenschmidt, H. & Harmelen, F. 2005a, 'Ontology languages for the Semantic Web', in *Information Sharing on the Semantic Web*, Springer, Berlin, Germany.

---- 2005b, 'Semantic integration', in *Information Sharing on the Semantic Web*, Springer, Berlin, Germany, pp. 3-22.

Sullivan, D. 2005, *Nielsen net ratings search engine ratings*, Neilson Net Ratings, viewed 19-06-2006 <http://searchenginewatch.com/reports/article.php/2156451>.

Tanner, K. 2002, 'Survey research', in *Research Methods for Students, Academics and Professionals*, 2nd edn, Centre for Information Studies, Charles Sturt University, Wagga Wagga, NSW.

Teswanich, W., Anutariya, C. & V, W. 2002, 'Unified representation for e-government knowledge management', paper presented to Proceeding of the 3rd International Workshop on Knowledge Management in E-Government.

Ticehurst, G. & Veal, A. 2000a, 'Questionnaire surveys', in *Business Research Methods - A Managerial Approach*, Pearson Education Pty Ltd., Frenchs Forest, NSW, Australia, pp. 135-58.

---- 2000b, 'Qualitative methods', in *Business Research Methods*, Pearson Education Limited, NSW, Australia, pp. 93-112.

Tourism, C. 2002, *CRC for sustainable tourism rebid proposal (executive summary), CRC for sustainable tourism*, Brisbane.

Uschold, M. & King, M. 1995, 'Towards a methodology for building ontologies', paper presented to Skuce D (eds) IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada.

Van Harmelen, F., Horrocks, I., Hendler, J. & McGuinness, D. 2000, 'The Semantic Web and its languages', *IEEE Intelligent Systems*, pp. 67-73.

van Heijst, G., Schreiber, A. & Wielinga, B. 1997, 'Using explicit ontologies in KBS', *International Journal of Human-Computer Studies*, vol. 45, pp. 183-292.

Vardi, M. 1982, 'The complexity of relational queries', paper presented to ACM SIGACT Symp. on Theory of Computing, Stockholm, Sweden.

Venturini, A. & Ricci, F. 2006, 'Applying Trip@dvice Recommendation Technology to www.visiteurope.com', paper presented to 17th European Conference on Artificial Intelligence, Riva del Garda, Italy, Aug 28th - Sept 1st.

Wache, H. 2003, 'Semantic mediation for heterogenous information sources', University of Bremem.

Weber, N., Schegg, R. & Murphy, H. 2005, 'An Investigation of satisfaction and loyalty in the virtual hospitality environment', in *Information and Communication Technologies in Tourism 2005,*, Springer, Vienna, Austria.

Weeks, N. & Crouch, I. 1999, 'Sites for sore eyes: an analysis of Australian tourism and hospitality Web sites', *Journal of Information Technology and Tourism*, vol. 2, no. 3-4, pp. 153-72.

Werthner, H. 2003, 'Intelligent systems in travel and tourism', paper presented to 18th International Joint Conference on Artificial Intelligence (IJCAI2003), Acapulco, Mexico, August 9-15.

APPENDIX A – Methontology Framework

Name of the Phase	Input	Description	Output
Planning	Nothing: first step	Plan the main tasks to be done, the way in which they will be arranged, the time and resources that are necessary to perform these tasks	A project plan
Specification	A series of questions such as: "Why is this ontology being built and what are its intended uses and end-users?"`	Identify ontology goals	Ontology requirement specification document written in natural language, using a set of intermediate representations or using competency questions, respectively. The document has to provide at least the following information: the purpose of the ontology (including its intended users, scenarios of use, end users etc.); the level of formality used to codify terms and meanings (highly informal, semi-informal, semi-formal, rigorously formal ontologies; the scope; its characteristics and granularity. Properties of this document are: concision, partial completeness, coverage of terms, the stopover problem and level of granularity of ache and every term, and consistency of all terms and their meanings.
Conceptualization	A good specification document	Conceptualize in a model that describes the problem and its solution. To identify and gather all the useful and potential usable domain knowledge and its meanings	A complete glossary of terms (including concepts, instances, verbs, and properties). Then a set of intermediate representations such as concepts, classification trees, verb diagram, table of formulas, and table of rules. The aim is to allow the final user to ascertain whether or not an ontology is useful and to compare the scope and completeness of several ontologies, their reusability, and share-ability.
Formalization	Conceptual model	Transform conceptual model into a formal or semi-compatible model, using frame-oriented or description logic representation systems	Formal conceptualization
Integration	Existing ontologies and the formal model	Processes of inclusion, polymorphic refinement, circular dependencies, and restriction. For example, select meta ontologies that better fit the conceptualization	
Implementation	Formal model	Select target language	Create a computable ontology
Maintenance		Including, modifying definition in the ontology	Guidelines for maintaining ontologies
Acquisition		Searching and listing knowledge sources through non-structured interviews with experts to have detailed information on concepts, terms, meanings, and so on.	A list of the sources of knowledge and a rough description of how the process will be carried out and what techniques will be used.
Evaluation	Computable ontology	Technical judgment with respect to a frame of reference	A formal and correct ontology
Documentation			Specification document must have the property of concision

Table 11: The Methontology framework.

APPENDIX B – Tourism Market Segment Characteristics

Adventure Tourist Characteristics

Adventure Activities	Victoria	Australia
Bushwalking or rainforest walks	48%	41%
Go fishing	28%	34%
Other outdoor activities	25%	23%
Scuba diving	*	1%
Snorkelling	*	4%
Surfing	8%	9%
Water activities / sports	14%	16%

*denotes less than 1%

Table 12: Adventure activities.

Accommodation	Victoria	Australia
Hotel, resort, motel	12%	18%
Guest house / B&B	2%	1%
Rented house, apartment, flat or unit	10%	12%
Caravan park / camping ground	20%	19%
Friends or relatives property	26%	30%
Own property (e.g. holiday house)	13%	6%
Caravan or camping (non-commercial)	14%	9%
Other	3%	3%

Table 13: Adventure accommodation.

Backpacker Tourist Characteristics

Activities	Victoria	Australia
Go shopping for pleasure	79%	81%
Go to the beach	76%	80%
Go to markets	73%	68%
Pubs, clubs, discos etc	73%	75%
Visit national parks / State parks	70%	65%
Visit wildlife parks / zoos / aquariums	65%	59%
Visit botanical or other public gardens	63%	60%
Visit museums or art galleries	61%	51%
Charter boat / cruise / ferry	58%	55%
Bushwalking / rainforest walks	58%	52%

Table 14: Backpacker activities.

* The complete list of accommodation preferences was not available for the backpacker market segment.

Caravan and Camping Tourist Characteristics

Activities	Victoria	Australia
General sight seeing	31%	35%
Bushwalking or rainforest walks	29%	26%
Eat out at restaurants	27%	34%
Go to the beach (including swimming)	26%	36%
Pubs clubs discos etc	22%	24%
Visit national parks or State parks	20%	19%
Go fishing	18%	24%
Go shopping for pleasure	15%	19%
Picnics or BBQs	14%	16%
Other outdoor activities	13%	13%

Table 15: Caravan and camping activities.

* The complete list of accommodation preferences was not available for the caravan and camping market segment

Cultural Tourist Characteristics

Cultural Activities	Victoria	Australia
Attend theatre concerts or other performing arts	18%	21%
Visit museums or art galleries	44%	41%
Visit art or craft workshops or studios	6%	9%
Attend festivals or fairs or cultural events	16%	17%
Experience aboriginal art or craft and cultural displays	2%	3%
Visit an aboriginal site or community	1%	2%
Visit history, heritage buildings sites or monuments	36%	37%

Table 16: cultural activities.

Accommodation	Victoria	Australia
Hotel, resort, motel	34%	30%
Guest house / B&B	2%	1%
Rented house, apartment, flat or unit	5%	8%
Caravan park / camping ground	13%	14%
Friends or relatives property	36%	36%
Own property (e.g. holiday house)	5%	2%
Caravan or camping (non-commercial)	3%	5%
Other	2%	3%

Table 17: Cultural accommodation.

Food and Wine Tourist Characteristics

Food & Wine Activities	Victoria	Australia
Visit wineries	6%	7%
Eat out at restaurants	98%	98%

Table 18: Food and wine activities.

Accommodation	Victoria	Australia
Hotel, resort, motel	33%	32%
Guest house / B&B	2%	1%
Rented house, apartment, flat or unit	7%	11%
Caravan park / camping ground	8%	9%
Friends or relatives property	41%	39%
Own property (e.g. holiday house)	6%	3%
Caravan or camping (non-commercial)	2%	3%
Other	2%	2%

Table 19: Food and wine accommodation.

APPENDIX C - Logic Notation

Description	Symbol
Disjunction	∨
Material implication	⊃
Material equivalence	≡
Negation of material equivalence	≢
Negation of equality	≠
Therefore	∴
Semantic consequence	⊨
Syntactic consequence	⊢
Existential quantifier	∃
Universal quantifier	∀
Set membership	∈
Denial of set membership	∉
Set intersection	∩
Set union	∪
Subset	⊆
Proper subset	⊂
One-to-one correspondence	≃
Aleph	ℵ
Gamma	Γ
Delta	Δ
Necessity	□
Possibility	◇

Table 20: Logic notation.

APPENDIX D – Accommodation ER Diagram

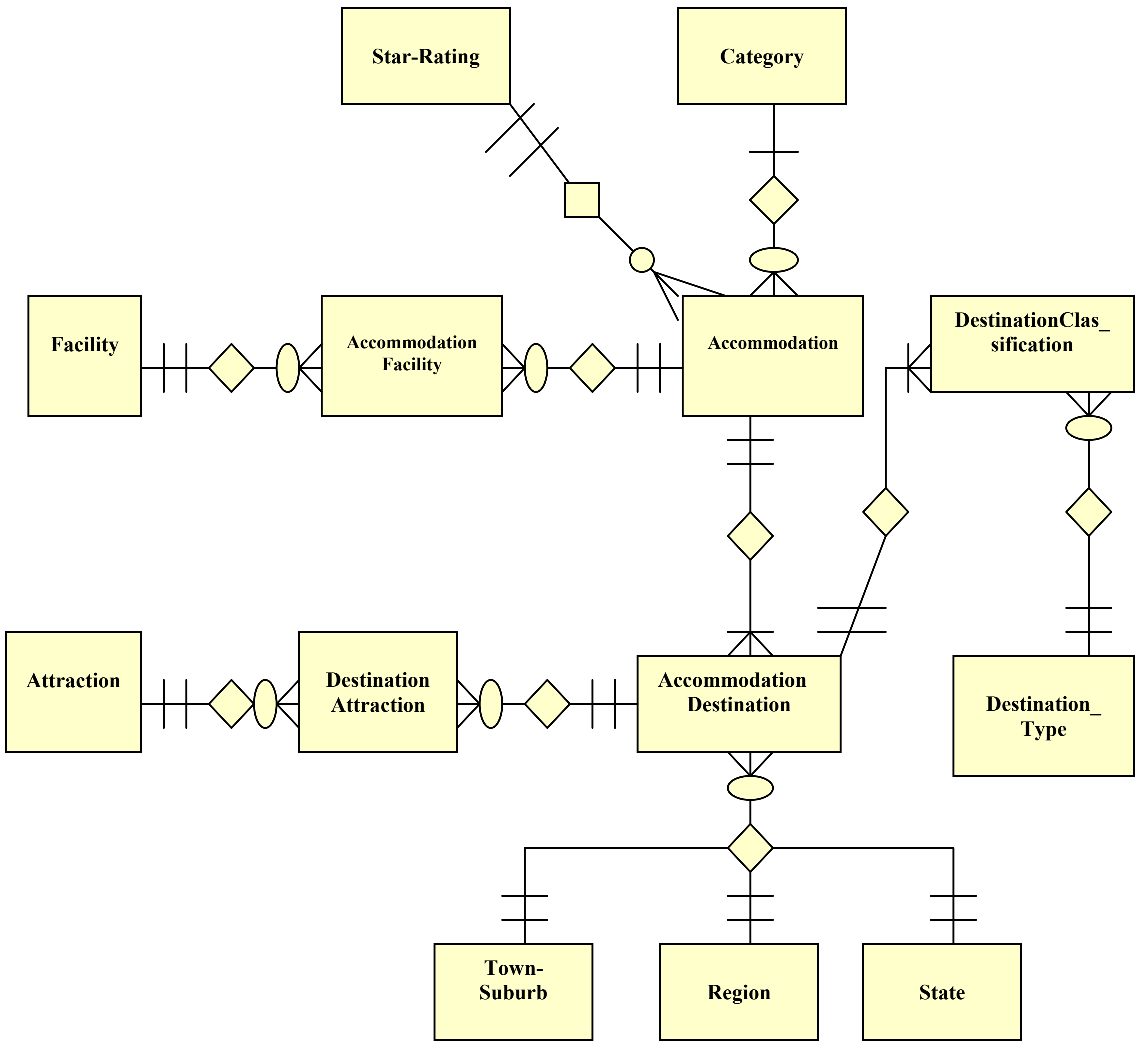


Figure 102: Accommodation ER diagram.

APPENDIX E – Accommodation Ontology

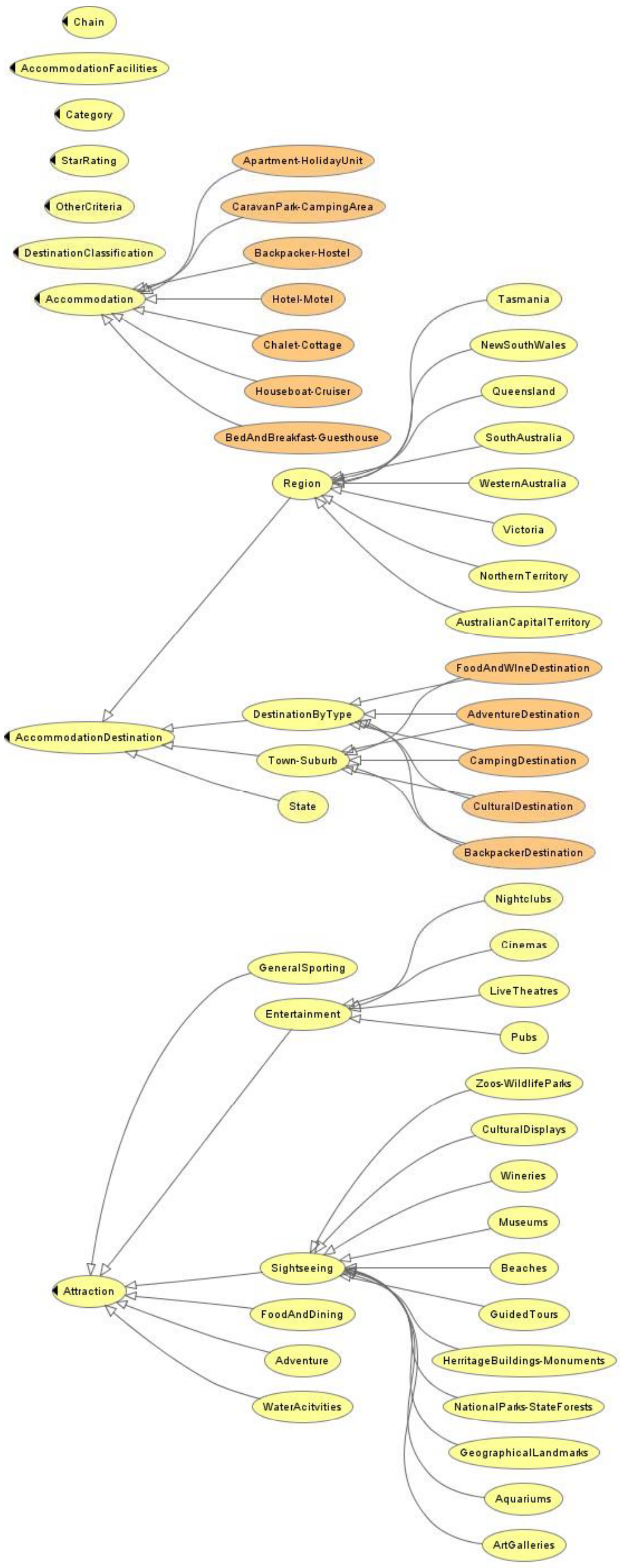


Figure 103: Accommodation ontology.

APPENDIX F – Accommodation Web Survey

Email to Survey Participants

Dear Accommodation Provider,

I am a PHD student at Victoria University in Melbourne. I am presently working on the development of an improved internet technology called the Semantic Web.

The aim of my research is to implement the Semantic Web in the tourism industry in order to provide greater Web exposure for tourism operators. Part of the research involves conducting a short on-line survey designed to gain an understanding of the requirements that tourism operators have for their Web sites.

If possible could you please assist with my research by participating in a pilot for this survey? The survey is very easy to complete and will take no more than five minutes of your time. The information obtained will be used purely for academic purposes and has no commercial use. The survey is available on-line at:

http://www.users.bigpond.com/brookeabrahams/AccommodationWebSurvey.htm

Feel free to email me with any suggestions on how the survey may be improved or made easier to complete for other participants. Your feedback would be greatly appreciated.

Kind Regards,

Brooke Abrahams
Victoria University

Copy of Survey

1. Please enter the name of your accommodation business:
   Business Name: ____________________

2. In which state is or territory is your business located?
   - ❍ NSW
   - ❍ QLD
   - ❍ SA
   - ❍ VIC

- ❍ WA
- ❍ ACT
- ❍ TAS
- ❍ NT

3. Please specify the location (town or suburb) of your business:
   Business Location: ____________________

4. What type of resort is your business?
   - ❍ Hotel/Motel
   - ❍ Apartment/Holiday Unit
   - ❍ Caravan Park/Camping Area
   - ❍ Chalet/Cottage
   - ❍ Backpacker/Hostel
   - ❍ Bed and Breakfast/Guesthouse
   - ❍ Houseboat/Cruiser

5. What is the Star Rating of your business?
   - ❍ 0.5 Star
   - ❍ 1 Star
   - ❍ 1.5 Star
   - ❍ 2 Star
   - ❍ 2.5 Star
   - ❍ 3 Star
   - ❍ 3.5 Star
   - ❍ 4 Star
   - ❍ 4.5 Star
   - ❍ 5 Star

6. What is the purpose of your business Web site (multiple answers permitted)?
   - ❑ Advertising/Promotion
   - ❑ On-line bookings
   - ❑ Means of providing information
   - ❑ Means of contact
   - ❑ None of the above

7. In addition to your own Web site, what additional online listings do you have (multiple answers permitted)?
   - ❑ With a local or regional authority, agency or business
   - ❑ With a State authority or agency (e.g. visitvictoria.com)
   - ❑ With a national authority or agency (e.g. the Australian Tourism Data Warehouse)
   - ❑ With an international authority, agency or business
   - ❑ With a AAA site (NRMA, RACV etc.)
   - ❑ Other online content provider

8. What proportion of your customers book their accommodation on-line (estimation only)?
   - ❍ 0%
   - ❍ 1-5%

- ❍ 6-10%
- ❍ 11-20%
- ❍ 21-50%
- ❍ 51-100%

9. Does your business have an on-line payment facility?
- ❍ Yes (Go to question 10)
- ❍ No (Go to question 11)

10. What proportion of your customers pay for their accommodation on-line (estimation only)?
- ❍ 0%
- ❍ 1-5%
- ❍ 6-10%
- ❍ 11-20%
- ❍ 21-50%
- ❍ 51-100%

11. Who created your business Web site?
- ❍ Business proprietor (owner)
- ❍ Business employee
- ❍ Friend or family
- ❍ IT professional
- ❍ None of the above

12. Who maintains or modifies your business Web site when the need arises?
- ❍ Business proprietor (owner)
- ❍ Business employee
- ❍ Friend or family
- ❍ IT professional
- ❍ None of the above

13. How likely are you to overhaul or rebuild your business Web site in the next 12 to 18 months?
- ❍ Don't know
- ❍ Recently completed or overhauled
- ❍ Definitely not
- ❍ Unlikely
- ❍ Maybe
- ❍ Likely
- ❍ Definitely

14. What factors would influence you to rebuild or overhaul your business web site in the next12 to 18 months (multiple answers permitted)?
- ❑ Avoid losing customers to competitors who have a better Web site
- ❑ Access new customers via the Internet
- ❑ Increase efficiency of internal processes
- ❑ Improve quality of service offered
- ❑ Reduce operating costs

- ❑ Improve Web site layout or usability
- ❑ None of the above

15. What factors would discourage you from rebuilding or overhauling your business Web in the next 12 to 18 months (multiple answers permitted)?

- ❑ Advantages are outweighed by cost implications
- ❑ No significant benefits likely
- ❑ Lack of interest
- ❑ Lack of technical expertise
- ❑ Do not like change
- ❑ None of the above

16. If a new Internet technology was available that could substantially increase your Web exposure, would you consider overhauling or rebuilding your Web site in order to use the technology?

- ❍ Don't know
- ❍ Definitely not
- ❍ Unlikely
- ❍ Maybe
- ❍ Likely
- ❍ Definitely

17. What factors may influence you to use a new Internet technology (multiple answers permitted)?

- ❑ It was easy to use
- ❑ It was quick to implement
- ❑ I was able to maintain my existing Web site
- ❑ The cost of implementing it was low
- ❑ It was proven to increase my Web exposure
- ❑ Competitors were using the technology
- ❑ None of the above

18. How would you prefer a new internet technology to be applied to your business?

- ❍ Add the technology to my existing Web site
- ❍ Add the technology to a new business Web site created from scratch
- ❍ Don't care how it is applied
- ❍ None of the above

19. If you were to overhaul your existing Web site in the next 12 to 18 months, would you want your overhauled web site to include an on-line payment facility?

- ❍ Don't know
- ❍ Definitely not
- ❍ Unlikely
- ❍ Maybe
- ❍ Likely
- ❍ Definitely

APPENDIX G – Survey Results

Question 1 Answers

383 business names received.

Question 2 Answers

Choice	Count	Percentage of Sample Answering
NSW	66	17.2%
QLD	61	15.9%
SA	77	20.1%
VIC	92	24.0%
WA	7	1.8%
ACT	8	2.1%
TAS	59	15.4%
NT	13	3.4%

Table 21: Businesses by state.

Question 3 Answers

Item	Frequency	Percent
Swansea	2	0.5%
Adelaide	4	1.0%
New Norfolk	5	1.3%
Cairns	5	1.3%
Auburn, Clare Valley	2	0.5%
Avoca Beach	2	0.5%
Mclaren Vale	2	0.5%
Darwin	5	1.3%
Burleigh Heads	2	0.5%
Townsville	2	0.5%
McLaren Vale	3	0.8%
Mount Gambier	4	1.0%
Merimbula	2	0.5%
Sawtell	2	0.5%
Wynyard	2	0.5%
Yarra Glen	2	0.5%
Barossa Valley	2	0.5%
Stanley	2	0.5%
Launceston	2	0.5%
Naracoorte	3	0.8%
Box Hill	2	0.5%
Port Augusta	3	0.8%
Airlie Beach	3	0.8%
Jervis Bay	2	0.5%
Alexandra	2	0.5%
Town	3	0.8%
Tamworth	2	0.5%
Beauty Point	2	0.5%
Bright	2	0.5%
Port Lincoln	3	0.8%
Hobart	4	1.0%
Albury	2	0.5%
Strahan	2	0.5%
Warrnambool	3	0.8%

Table 22: Business locations.

Ballarat	3	0.8%
Warragul	2	0.5%
Marysville	2	0.5%
Sydney	2	0.5%
Aireys Inlet	2	0.5%
Alexandra Headland	2	0.5%
Coober Pedy	2	0.5%
Bundaberg	2	0.5%
Eaglehawk Neck	2	0.5%
GOONDIWINDI	2	0.5%
Bicheno	3	0.8%
Devonport	2	0.5%
Kapunda	2	0.5%
Victor Harbor	2	0.5%
Noosa Heads	2	0.5%
Burnie	2	0.5%
(Unique responses)	259	67.6%
(Total)	383	100.0%

Table 23: Business locations continued.

Question 4 Answers

Choice	Count	Percentage of Sample Answering
Hotel/Motel	121	31.6%
Apartment/Holiday Unit	75	19.6%
Caravan Park/Camping Area	33	8.6%
Chalet/Cottage	35	9.1%
Backpacker/Hostel	6	1.6%
Bed and Breakfast/Guesthouse	104	27.2%
Houseboat/Cruiser	9	2.3%

Table 24: Businesses by category.

Question 5 Answers

Choice	Count	Percentage of Sample Answering
0.5 Star	1	0.3%
1 Star	1	0.3%
1.5 Star	0	0.0%
2 Star	4	1.0%
2.5 Star	11	2.9%
3 Star	32	8.4%
3.5 Star	85	22.2%
4 Star	133	34.7%
4.5 Star	88	23.0%
5 Star	28	7.3%

Table 25: Businesses by star rating.

Question 6 Answers

Choice	Count	Percent of Sample Asked
Advertising/Promotion	352	91.9%
On-line bookings	233	60.8%
Means of providing information	327	85.4%
Means of contact	290	75.7%
None of the above	10	2.6%

Table 26: Purpose of business Website.

Question 7 Answers

Choice	Count	Percent of Sample Asked
With a local or regional authority, agency or business	313	81.7%
With a State authority or agency (e.g. visitvictoria.com)	268	70.0%
With a national authority or agency (e.g. the Australian Tourism Data Warehouse)	172	44.9%
With an international authority, agency or business	95	24.8%
With a AAA site (NRMA, RACV etc.)	330	86.2%
Other online content provider	209	54.6%

Table 27: Additional online listings.

Question 8 - What proportion of your customers book their accommodation on-line (estimation only)?

Question 8 Answers

Choice	Count	Percentage of Sample Answering
0%	32	8.4%
1-5%	92	24.0%
6-10%	74	19.3%
11-20%	79	20.6%
21-50%	66	17.2%
51-100%	40	10.4%

Table 28: Online bookings.

Question 9 Answers

Choice	Count	Percentage of Sample Answering
Yes (Go to question 10)	70	18.3%
No (Go to question 11)	313	81.7%

Table 29: Businesses with online payment facility.

Question 10 Answers

Choice	Count	Percentage of Sample Answering
0%	102	51.3%
1-5%	38	19.1%
6-10%	23	11.6%
11-20%	12	6.0%
21-50%	15	7.5%
51-100%	9	4.5%

Table 30: Online payments.

Question 11 Answers

Choice	Count	Percentage of Sample Answering
Business proprietor (owner)	47	12.3%
Business employee	22	5.7%
Friend or family	44	11.5%
IT professional	242	63.2%
None of the above	28	7.3%

Table 31: Creator of business Website.

Choice	Count	Percentage of Sample Answering
Business proprietor (owner)	77	20.1%
Business employee	40	10.4%
Friend or family	37	9.7%
IT professional	203	53.0%
None of the above	26	6.8%

Table 32: Maintainer of business Website.

Question 13 Answers

Choice	Count	Percentage of Sample Answering
Don't know	30	7.8%
Recently completed or overhauled	64	16.7%
Definitely not	9	2.3%
Unlikely	63	16.4%
Maybe	46	12.0%
Likely	82	21.4%
Definitely	89	23.2%

Table 33: Likelihood of overhauling Website.

Question 14 Answers

Choice	Count	Percent of Sample Asked
Avoid losing customers to competitors who have a better Web site	166	43.3%
Access new customers via the Internet	219	57.2%
Increase efficiency of internal processes	129	33.7%
Improve quality of service offered	180	47.0%
Reduce operating costs	79	20.6%
Improve Web site layout or usability	262	68.4%
None of the above	37	9.7%

Table 34: Factors influencing the overhaul of Website.

Question 15 Answers

Choice	Count	Percent of Sample Asked
Advantages are outweighed by cost implications	148	38.6%
No significant benefits likely	118	30.8%
Lack of interest	14	3.7%
Lack of technical expertise	68	17.8%
Do not like change	6	1.6%
None of the above	132	34.5%

Table 35: Factors discouraging the overhaul of Website.

Question 16 Answers

Choice	Count	Percentage of Sample Answering
Don't know	29	7.6%
Definitely not	0	0.0%
Unlikely	13	3.4%
Maybe	141	36.8%
Likely	118	30.8%
Definitely	82	21.4%

Table 36: Willingness to use a new Internet technology.

Question 17 Answers

Choice	Count	Percent of Sample Asked
It was easy to use	298	77.8%
It was quick to implement	245	64.0%
I was able to maintain my existing Web site	240	62.7%
The cost of implementing it was low	294	76.8%
It was proven to increase my Web exposure	305	79.6%
Competitors were using the technology	104	27.2%
None of the above	9	2.3%

Table 37: Factors influencing uptake of technology.

Question 18 Answers

Choice	Count	Percentage of Sample Answering
Add the technology to my existing Web site	235	61.4%
Add the technology to a new business Web site created from scratch	42	11.0%
Don't care how it is applied	88	23.0%
None of the above	18	4.7%

Table 38: Preference for how technology is applied.

Question 19 Answers

Choice	Count	Percentage of Sample Answering
Don't know	24	6.3%
Definitely not	27	7.0%
Unlikely	61	15.9%
Maybe	96	25.1%
Likely	75	19.6%
Definitely	100	26.1%

Table 39: Preference for online payment facility.

APPENDIX H - AcontoWeb Queries

Query 1

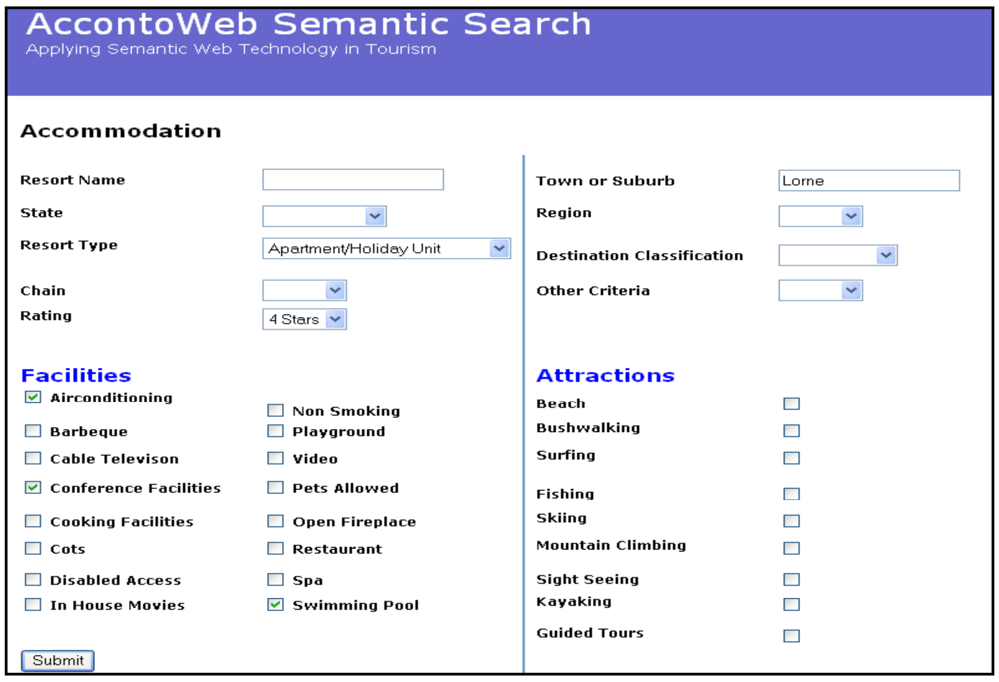


Figure 104: Query 1 in AcontoWeb.

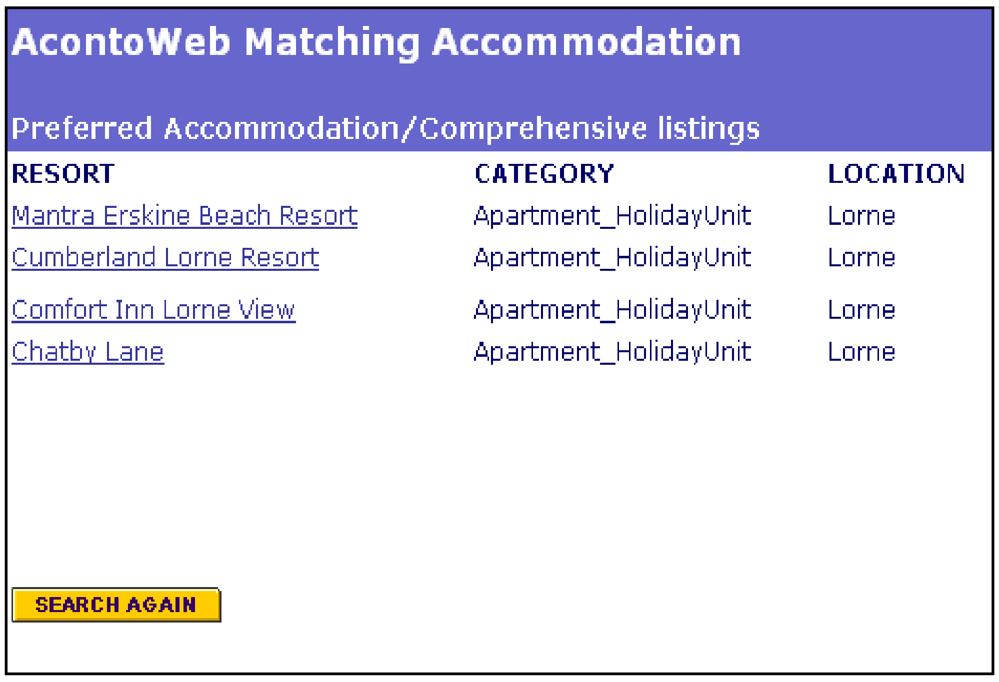


Figure 105: Query 1 results in AcontoWeb.

Query 2

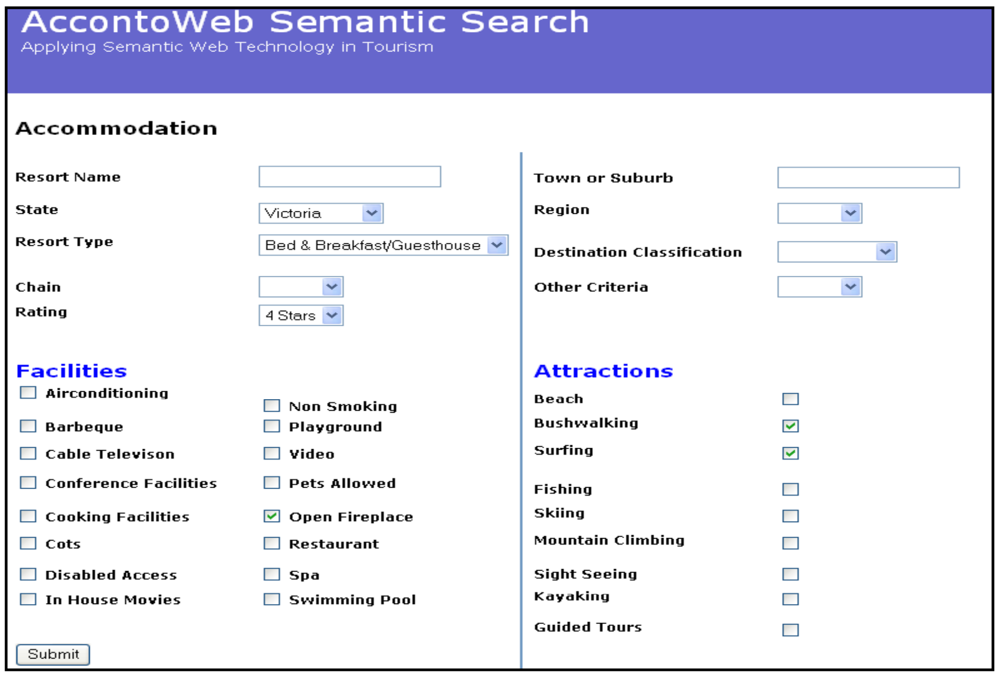


Figure 106: Query 2 in AcontoWeb.

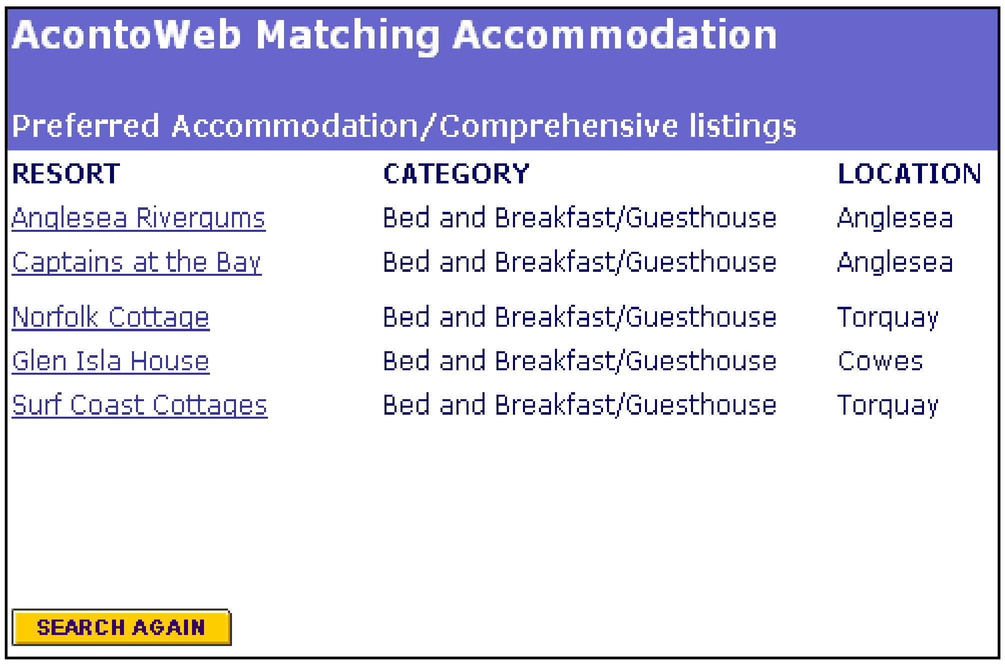


Figure 107: Query 2 results in AcontoWeb.

Query 3

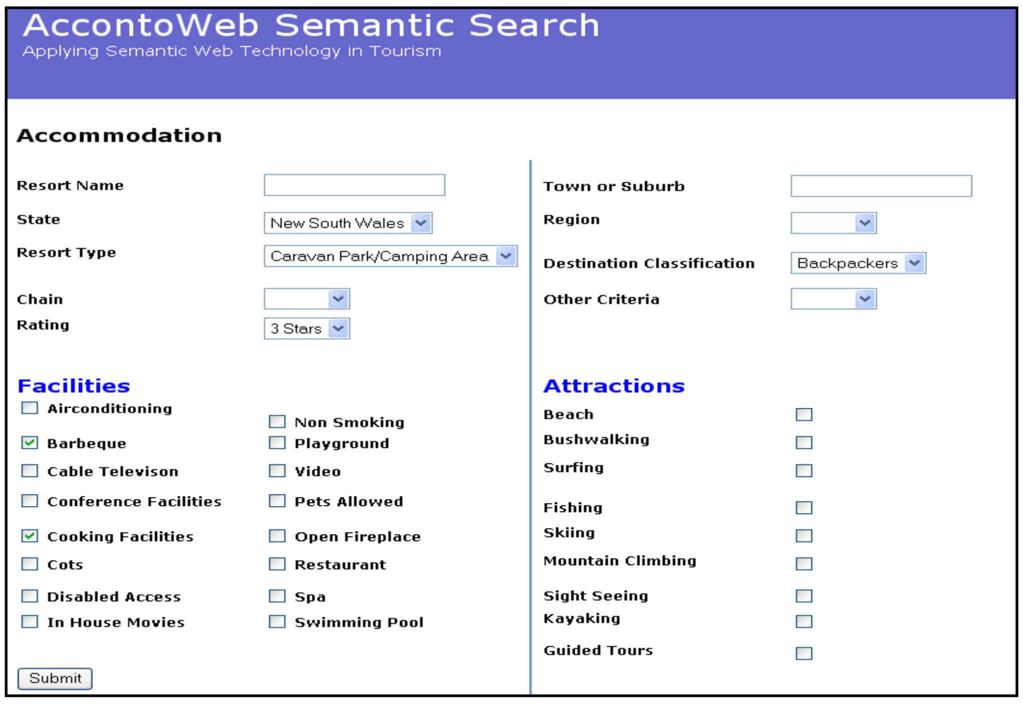


Figure 108: Query 3 in AcontoWeb.

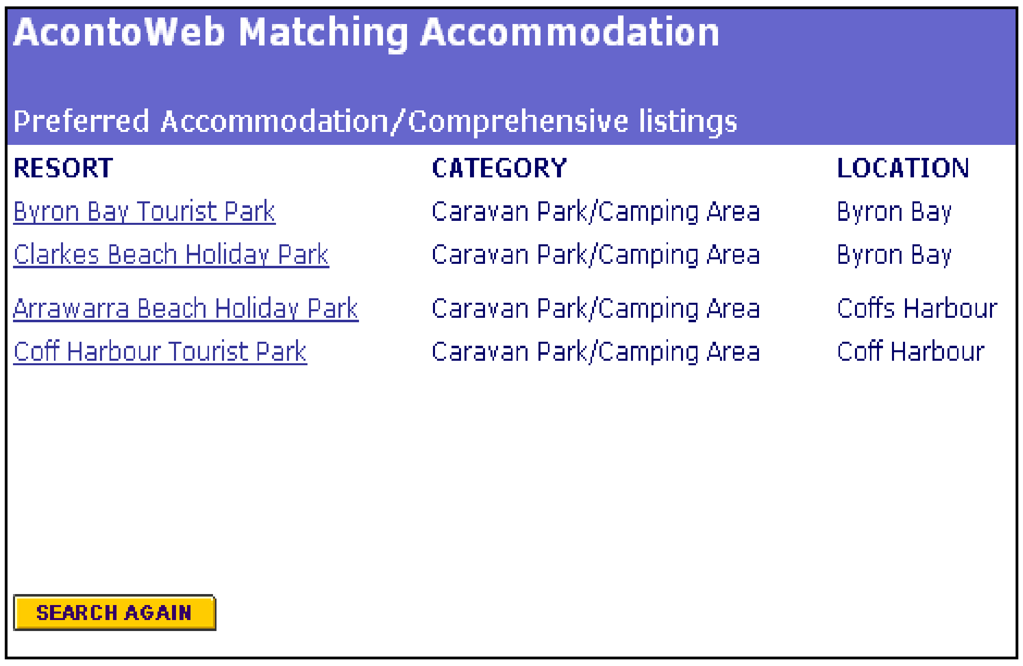


Figure 109: Query 3 results in AcontoWeb.

Query 4

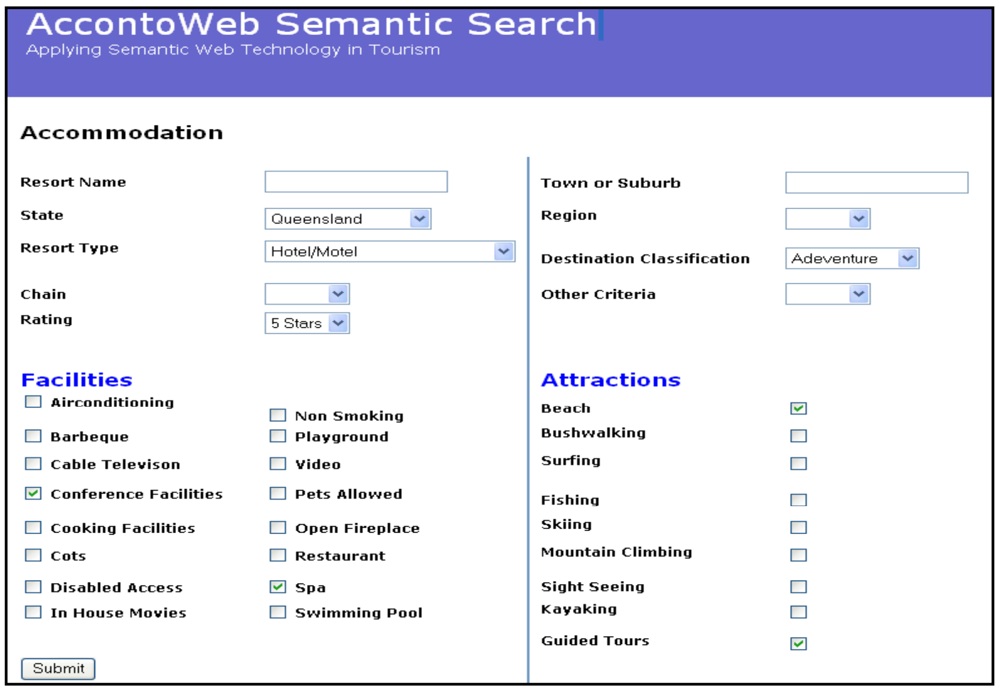


Figure 110: Query 4 in AcontoWeb.

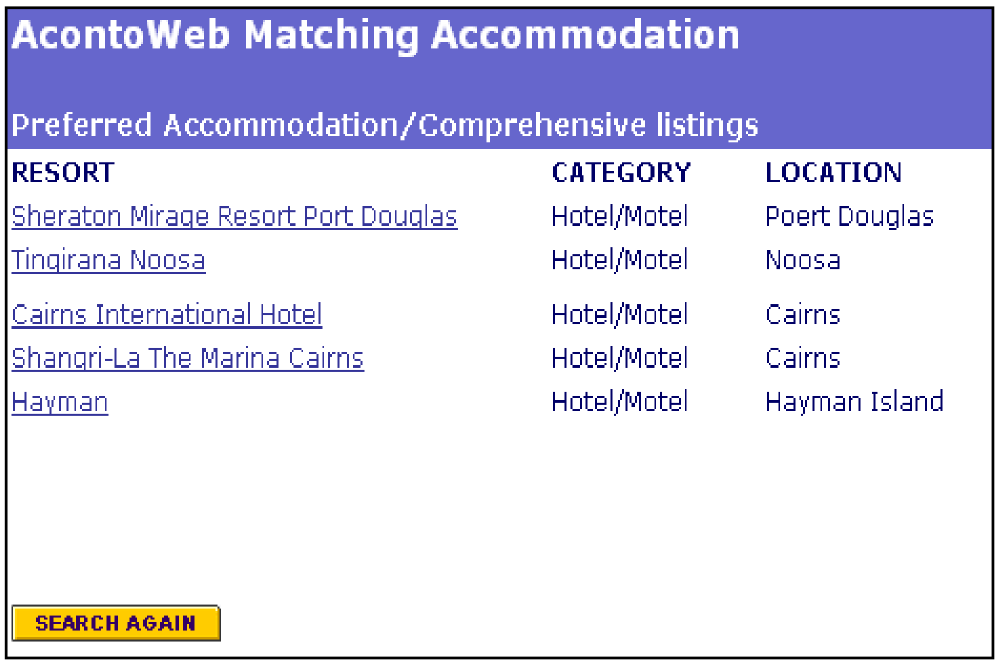


Figure 111: Query 4 results in AcontoWeb.

APPENDIX I – Experimental Queries SQL Syntax

Query 1

```
SELECT Accommodation.BusinessName
FROM AccommodationFacility AS AccommodationFacility_1, AccommodationFacility
AS AccommodationFacility_2, (Accommodation INNER JOIN AccommodationFacility
ON Accommodation.AccommodationID = AccommodationFacility.AccommodationID)
INNER JOIN AccommodationDestination ON Accommodation.AccommodationID =
AccommodationDestination.AccommodationID
WHERE (((Accommodation.Category)="Apartment_HolidayUnit") AND
((Accommodation.StarRating)="FourStar") AND
((AccommodationDestination.DestinationName)="Lorne") AND
((AccommodationFacility.FacilityName)="SwimmingPool") AND
((AccommodationFacility_1.FacilityName)="Airconditioning") AND
((AccommodationFacility_2.FacilityName)="ConferenceFacilities"));
```

Query 2

```
SELECT Accommodation.BusinessName
FROM DestinationAttraction AS DestinationAttraction_1, ((Accommodation INNER
JOIN AccommodationFacility ON Accommodation.AccommodationID =
AccommodationFacility.AccommodationID) INNER JOIN AccommodationDestination
ON Accommodation.AccommodationID =
AccommodationDestination.AccommodationID) INNER JOIN DestinationAttraction ON
AccommodationDestination.DestinationID = DestinationAttraction.DestinationID
WHERE (((Accommodation.Category)="BedAndBeakfast_Guesthouse") AND
((Accommodation.StarRating)="FourStar") AND
((AccommodationFacility.FacilityName)="OpenFireplace") AND
((AccommodationDestination.DestinationName)="Vic") AND
((DestinationAttraction.AttractionName)="Surfing") AND
((DestinationAttraction_1.AttractionName)="Bushwalking"));
```

Query 3

```
SELECT Accommodation.BusinessName
FROM AccommodationFacility AS AccommodationFacility_1, ((Accommodation
INNER JOIN AccommodationDestination ON
Accommodation.AccommodationID=AccommodationDestination.AccommodationID)
INNER JOIN DestinationClassification ON
AccommodationDestination.DestinationID=DestinationClassification.DestinationID)
INNER JOIN AccommodationFacility ON
Accommodation.AccommodationID=AccommodationFacility.AccommodationID
WHERE (((Accommodation.Category)="CaravanPark_CampingArea") AND
((Accommodation.StarRating)="ThreeStar") AND
((AccommodationDestination.DestinationName)="NSW") AND
((DestinationClassification.Classification)="Backpackers") AND
((AccommodationFacility.FacilityName)="CookingFacilities") AND
((AccommodationFacility_1.FacilityName)="Barbeque"));
```

Query 4

```
SELECT Accommodation.BusinessName
FROM DestinationAttraction AS DestinationAttraction_1, (((Accommodation INNER
JOIN AccommodationDestination ON Accommodation.AccommodationID =
AccommodationDestination.AccommodationID) INNER JOIN DestinationClassification
ON AccommodationDestination.DestinationID =
DestinationClassification.DestinationID) INNER JOIN DestinationAttraction ON
AccommodationDestination.DestinationID = DestinationAttraction.DestinationID)
INNER JOIN AccommodationFacility ON Accommodation.AccommodationID =
AccommodationFacility.AccommodationID
WHERE (((Accommodation.Category)="Hotel_Motel") AND
((Accommodation.StarRating)="FiveStar") AND
((AccommodationFacility.FacilityName)="Spa") AND
((DestinationAttraction.AttractionName)="Beaches") AND
((DestinationAttraction_1.AttractionName)="GuidedTours") AND
((DestinationClassification.Classification)="Adventure"));
```

APPENDIX J – Experimental Queries SPARQL Syntax

Query 1

```
PREFIX Q: <http://www.owl-ontologies.com/Accommodation.owl#>
SELECT ?BusinessName ?URL
WHERE  {?Accommodation  Q:hasCategory Q:Apartment_HolidayUnit .
        ?Accommodation  Q:hasStarRating Q:FourStar .
        ?Accommodation Q:hasAccommodationDestination Q:Lorne .
        ?Accommodation Q:hasAccommodationFacility Q:SwimmingPool .
        ?Accommodation Q:hasAccommodationFacility Q:Airconditioning .
         ?Accommodation Q:hasAccommodationFacility Q:ConferenceFacilities .
        ?Accommodation :hasBusinessName ?BusinessName .
        ?Accommodation :hasURL ?URL}
```

Query 2

```
PREFIX Q: <http://www.owl-ontologies.com/Accommodation.owl#>

SELECT ?BusinessName ?URL

WHERE  {?Accommodation  Q:hasCategory Q:BedAndBreakfast_Guesthouse .
        ?Accommodation  Q:hasStarRating Q:FourStar .
        ?Accommodation Q:hasAccommodationDestination Q:VIC .
        ?Accommodation Q:hasAccommodationFacility Q:OpenFireplace .
        ?Accommodation Q:hasDestinationAttraction Q:Surfing .
        ?Accommodation Q:hasDestinationAttraction Q:Bushwalking .
        ?Accommodation :hasBusinessName ?BusinessName .
        ?Accommodation :hasURL ?URL}
```

Query 3

```
PREFIX Q: <http://www.owl-ontologies.com/Accommodation.owl#>
SELECT ?BusinessName ?URL
WHERE  {?Accommodation  Q:hasCategory Q:CaravanPark_CampingArea .
          ?Accommodation  Q:hasStarRating Q:ThreeStar .
          ?Accommodation Q:hasAccommodationDestination Q:NSW .
          ?Accommodation Q:hasDestinationClassification Q:Backpackers .
          ?Accommodation Q: hasAccommodationFacility Q:Barbeque .
          ?Accommodation Q: hasAccommodationFacility Q:CookingFacilities .
          ?Accommodation :hasBusinessName ?BusinessName .
          ?Accommodation :hasURL ?URL}
```

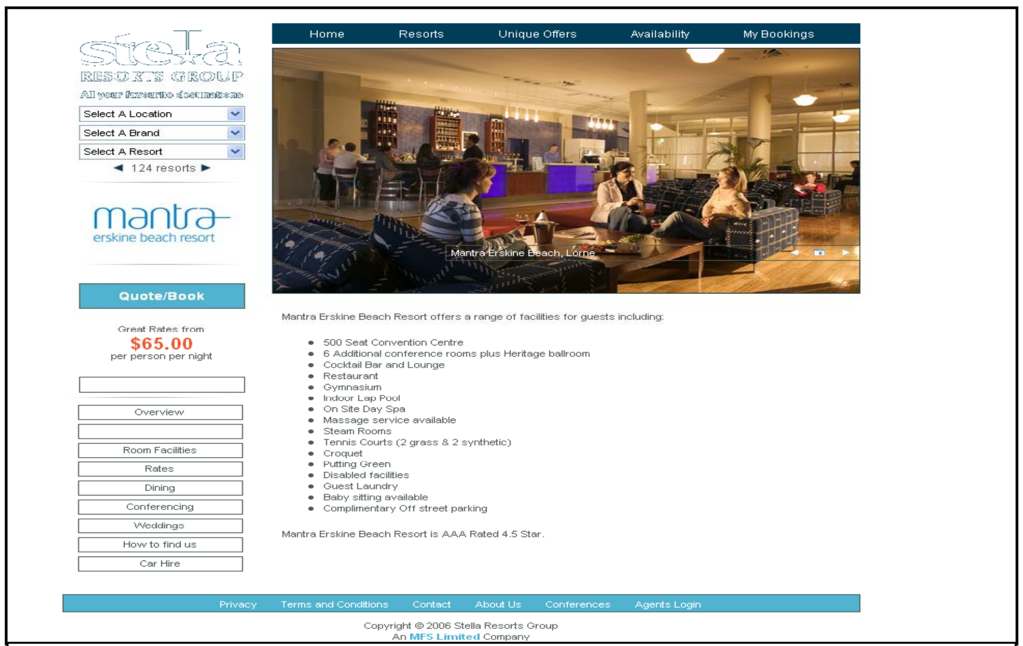
Query 4

```
PREFIX Q: <http://www.owl-ontologies.com/Accommodation.owl#>
SELECT ?BusinessName ?URL
WHERE  {?A Q:hasDestinationAttraction ?C .
          ?C ?B Q:Beaches .
          ?A Q:hasDestinationAttraction ?D .
          ?D ?B Q:GuidedTours .
          ?A Q:hasStarRating Q:FiveStar .
          ?A Q:hasCategory Q:Hotel_Motel .
          ?A Q:hasAccommodationFacility Q:ConferenceFacilities .
          ?A Q:hasAccommodationFacility Q:Spa .
          ?A Q:hasDestinationClassification Q:Adventurers .
           ?A Q:hasAccommodationDestination Q:QLD .
          ?A :hasBusinessName ?BusinessName .
          ?A :hasURL ?URL}
```

APPENDIX K – Annotated Webpages

Mantra Erskine Resort

Web page



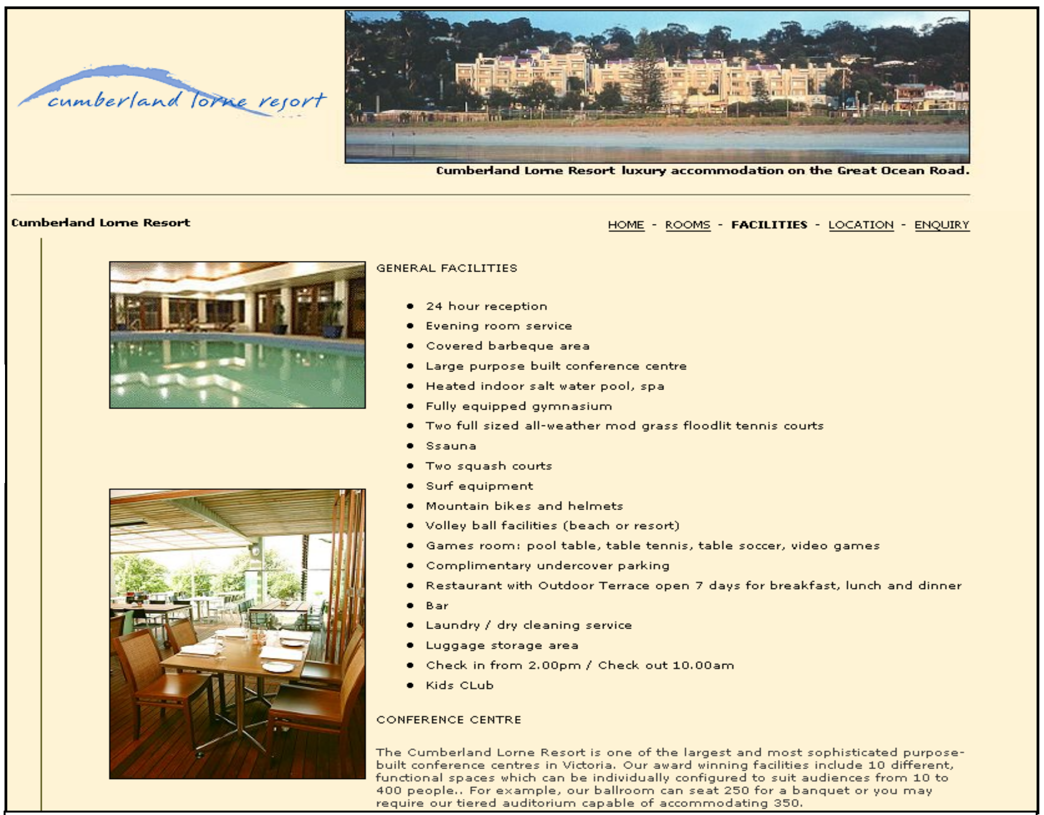
RDF Markup


```
<!--AcontoWeb Annotation<?xml version="1.0"?>
<rdf:RDF
   xmlns:p3="http://www.accommodation.owl#"
   xmlns:p2="http://www.owl-ontologies.com/Accomodation.owl#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:p1="http://www.owl-ontologies.com/assert.owl#"
   xmlns="http://www.owl-ontologies.com/Accommodation.owl#"
   xmlns:p4="http://www.owl-ontologies.com/"
   xmlns:p5="http://www.owl-ontologies.com/Accommodation.ow#"
   xml:base="http://www.owl-ontologies.com/Accommodation.owl">
 <Accommodation rdf:ID="MantraErskineBeachResort">
   <hasOtherCriteria>
     <OtherCriteria rdf:ID="SYCS"/>
   </hasOtherCriteria>
   <hasDestinationClassification rdf:resource="#Lorne"/>
   <hasEmail rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >erskine.res@mantraresorts.com.au &lt;erskine.res@mantraresorts.com.au</hasEmail>
   <hasAccommodationFacility rdf:resource="#ConferenceFacilities"/>
   <hasBusinessName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >Mantra Erskine Beach Resort</hasBusinessName>
   <hasFacility rdf:resource="#SwimmingPool"/>
   <hasFax rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >03 5289 1209</hasFax>
   <hasStarRating rdf:resource="#FourStar"/>
   <hasAccommodationFacility rdf:resource="#CookingFacilities"/>
   <hasAccommodationFacility>
     <AccommodationFacilities rdf:ID="Video"/>
   </hasAccommodationFacility>
   <hasAccommodationFacility>
     <Facilities rdf:ID="Restaurant"/>
   </hasAccommodationFacility>
   <hasAddress rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >Mountjoy Pde</hasAddress>
   <hasAccommodationFacility rdf:resource="#Airconditioning"/>
   <hasCategory rdf:resource="#Apartment_HolidayUnit"/>
   <hasTelephone rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >03 5289 1185</hasTelephone>
   <hasURL rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
   >http://www.lornevictoria.com.au/3.asp?id=81</hasURL>
   <hasAccommodationFacility rdf:resource="#Spa"/>
 </Accommodation>
</rdf:RDF> -->
```


Figure 112: Annotated Webpage 1.

Mantra Erskine Resort

Web page



RDF Markup


```
<!--AcontoWeb Annotation<?xml version="1.0"?>
<rdf:RDF
    xmlns:p3="http://www.accommodation.owl#"
    xmlns:p2="http://www.owl-ontologies.com/Accomodation.owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:p1="http://www.owl-ontologies.com/assert.owl#"
    xmlns="http://www.owl-ontologies.com/Accommodation.owl#"
    xmlns:p4="http://www.owl-ontologies.com/"
    xmlns:p5="http://www.owl-ontologies.com/Accommodation.ow#"
    xml:base="http://www.owl-ontologies.com/Accommodation.owl">
<Accommodation rdf:ID="CumberlandLorneResort">
    <hasTelephone rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >03 5289 2400</hasTelephone>
    <hasEmail rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >chef@cumberland.com.au</hasEmail>
    <hasAccommodationFacility rdf:resource="#Airconditioning"/>
    <hasFax rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >03 5289 2256</hasFax>
    <hasAccommodationFacility rdf:resource="#ConferenceFacilities"/>
    <hasBusinessName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >CumberlandLorneResort</hasBusinessName>
    <hasDestinationClassification rdf:resource="#Lorne"/>
    <hasCategory rdf:resource="#Apartment_HolidayUnit"/>
    <hasStarRating rdf:resource="#FourStar"/>
    <hasAccommodationFacility rdf:resource="#SwimmingPool"/>
    <hasURL rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >http://www.cumberland.com.au/</hasURL>
  </Accommodation>
</rdf:RDF> -->
```


Figure 113: Annotated Webpage 2.

APPENDIX L - Publications Attributable to Thesis

At the time of writing, research outcomes attributable to the thesis had resulted in thirteen refereed DEST publications. The publications, which are listed below, include two journal articles, nine conference papers and two book chapters:

McGrath, G.M., Abrahams, B. 2007 'A Semantic Portal for the Tourism and Hospitality Industry: Its Design, Use and Acceptance ', International Journal of *Internet and Enterprise Management*, Vol. 5 No. 2, (forthcoming).

Abrahams, B. 2007, 'Developing Semantic Portals'. Book Chapter. Encyclopaedia of Portal Technology and Applications. Idea Group Publication, (forthcoming).

Abrahams, B. and Dai, W. 2007, 'Semantic Portals: An Introduction and Overview'. Book Chapter. Encyclopaedia of Portal Technology and Applications. Idea Group Publication, (forthcoming).

McGrath, G.M., Abrahams, B. 2006, 'Ontology-based website generation and utilization for tourism services', Journal of *Information Technology in Hospitality*, vol. 4.

McGrath, M. & Abrahams, B. 2006a, 'AcOntoWeb: A Semantic Portal for the Tourism and Hospitality Industry', paper presented to Hospitality Information Technology Association (HITA'06), Minneapolis, USA, June 18 - 19.

McGrath, G.M., Abrahams, B. and More, E. 2006. 'Potential Use of Advanced Online Technologies Among Australian Accommodation Sector Operators', (M. Hitz, M. Sigala and J. Murphy eds.), Proceedings of the 13th International Conference on Information Technology in Travel and Tourism (ENTER2006), (ISBN 3-211-30987-X), Springer-Verlag: Lausanne, January Switzerland, 18–20, pp.183-195.

McGrath, G.M., Abrahams, B. and More, E. 2005. 'Online Technology Use and Adoption Among Australian Accommodation Enterprise Operators', Proceedings of the 19th Annual ANZAM Conference, (ISBN 1 74088 245 8), Canberra, ACT, 7-10 December 2005, pp. 1-12.

Abrahams, B. & Dai, W. 2005. 'Meeting Semantic Web Challenges with Automated Annotation and Multi-Agent Querying of Web Resources', paper presented at Victoria University Business Research Conference, Melbourne, Australia November 29, 2005.

Abrahams, B. & Dai, W. 2005. 'Architecture for Automated Annotation and Ontology Based Querying of Semantic Web Resources', paper presented to IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne, France, September 19-22, 2005.

McGrath, G.M., Abrahams, B. and More, E. 2005. 'Attitudes Towards Online Technology Among Australian Accommodation Enterprise Operators: A Preliminary Study' paper presented to Tourism Enterprise Strategies

Conference (TES2005), Victoria University, Melbourne, Australia, 11-12 July 2005.

Dai, W. & Abrahams, B. 2005. 'A Multi-agent Architecture for Semantic Web Resourses', paper presented to IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Compiegne, France September 19-22, 2005.

Abrahams, B., Dai, W., and McGrath, M. 2004. 'A Multi Agent Approach for Dynamic Ontology Loading to Support Semantic Web Applications. In Proceedings of 2004 IEEE International Conference on Information Reuse and Integration'. Ed(s). Atif Memon. IEEE, Piscataway, New Jersy, USA. 570-575.

McGrath, B., and Abrahams, B. 2004. 'Ontology-Based Website Generation and Utilization for Tourism Services'. In Proceedings of the Hospitality Information Technology Association Conference: HITA 04. Ed(s). Peter O\'Connor and Andrew J. Frew. HITA, Cergy Pontoise, France. 138-161.

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Browser - A Web client that allows a human to read information on the Web. Microsoft Internet Explorer and Netscape Navigator are two leading browsers.

Class – A set of things; a one parameter predicate; a unary relation.

Client – Any program that uses the services of another program. On the Web a Web client is a program such as a browser, editor or search robot that reads or writes information on the Web.

CSS (Cascading Style Sheets) - A W3C Standard that uses a rule-based declarative syntax that assigns formatting properties to the element either HTML or XML element content.

DAML (DARPA Agent Markup Language) - The DAML language is being developed as an extension to XML and the Resource Description Framework (RDF). The latest release of the language (DAML+OIL) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable. http://www.dami.org/.

DAML+01L Web Ontology Language- A semantic markup language for Web resources. It builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modeling primitives. DAML+OIL provides modelling primitives commonly found in frame-based languages. DAML+OIL (March 2001) extends DAML+OIL (December 2000) with values from XML Schema datatypes.

Data model - A data model is what is formally-defined in a DTD (Document Type Definition) or XML Schema. A document's "data model" consists of the allowable element and attribute names and optional structural and occurrence constraints for a "type" or "class" of documents.

DAML (DARPAAgent Markup Language) - The DAML language is being developed as an extension to XML and the Resource Description Framework (RDF). The latest release of the language (DAML+OIL) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable. http://www.daml.org/.

DAML+01L Web Ontology Language - A semantic markup language for Web resources. It builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modelling primitives. DAML+OIL provides modelling primitives commonly found in frame-based languages. DAML+OIL (March 2001) extends DAML+OIL (December 2000) with values from XML Schema datatypes.

Data model - A data model is what is formally-defined in a DTD (Document Type Definition) or XML Schema. A document's "data model" consists of the allowable element and attribute names and optional structural and occurrence constraints for a "type" or "class" of documents.

Data typing - Data is said to be "typed" when it takes on additional abstract meaning than what its characters usually represent. Integers, dates, booleans, and strings are all examples of typed data (data types). A data value that is typed takes on additional meaning, due to the semantic properties known to be associated with specific named data types.

DTD (Document Type Definition) - A formal definition of the data model (the elements and attributes allowed and their allowable content and nesting structure) for a class of documents. XML DTDs are written using SGML DTD syntax.

E-Business - the term 'ebusiness' refers to using the Internet for doing business. Every time a business uses the Internet to conduct business, it is doing ebusiness.

E-Commerce - Electric commerce: the conducting of business communication and transactions over networks and through computers. Specifically, ecommerce is the buying and selling of goods and services, and the transfer of funds, through digital communications.

Graph - Informally, a graph is a finite set of dots called vertices (or nodes) connected by links called edges (or arcs). More formally a simple graph is a (usually finite) set of vertices V and set of unordered pairs of distinct elements of V called edges.

HTML (Hypertext Markup Language) - A computer language for representing the contents of a page of hypertext; the language that most Web pages are written in.

HyperLink - A medium that includes links and includes media as well as text and is sometimes called hypermedia.

HTTP (HyperText Transfer Protocol) - This is the protocol by which web clients (browsers) and web servers communicate. It is stateless; meaning that it does not maintain a conversation between a given client and server, but it can be manipulated using scripting to appear as if state is being maintained. Donot confuse HTML (Markup language for our browser-based front ends), with HTTP (protocol used by clients and servers to send and receive messages over the Web).

ICT - The use of computer-based information systems and communications systems to process, transmit and store data and information.

Internet - A global network of networks through which computers communicate by sending information in packets. Each network consists of computers connected by cables or wireless links.

Intranet - A part of the Internet or part of the Web used internally within a company or organization. **Invocation Execution** of an identified Web Service by an agent or other service.

IP (Internet Protocol) - The protocol that governs how computers send packets across the Internet. Designed by Vint Cerf and Bob Khan.

Java - A programming language developed (originally as "Oak") by James Gosling of Sun Microsystems. Designed for portability and usability embedded in small devices, Java took off as a language for small applications ("applets") that ran within a Web browser.

GUI (Graphical User Interface) - An end-user sees and interacts with when operating (interacting with) a software application. Sometimes referred to as the "front-end" of an application. HTML is the GUI standard for Web based applications.

Link - A link (or hyperlink) is a relationship between two resources. HTML links usually connect HTML documents together in this fashion (called a "hyperlink), but links can link to any type of resource (documents, pictures, sound and video files) capable of residing at a Web address.

Markup - Comprised of several "special characters" that are used to structure a document's character data into logical components that can then be labeled (named) so that they can be manipulated more easily by a software application.

Markup Language - A language used to structure a document's character data into logical components, and "name" them in a manner that is useful. These labels (element names) provide either formatting information about how the character data should be visually presented (for a word processor or a Web browser, for instance) or they can provide "semantic" (meaningful) information about what kind of data the component represents. Markup languages provide a simple format for exchanging text-based character data that can be understood by both humans and machines.

Meta - A prefix to indicate something applied to itself, for example, a metameeting is a meeting about meetings.

Metadata - Data about data on the Web, including but not limited to authorship, classification, endorsement, policy, distribution terms, IPR, and so on. A significant use for the Semantic Web.

Meta-markup language - A language used to define markup languages. SGML and XML are both metamarkup languages. HTML is a markup language that was defined using the SGML meta-markup language.

Object - Of the three parts of a statement, the object is one of the two things related by the predicate. Often, it is the value of some property, such as the color of a car. See also: subject, predicate.

OIL (Ontology Inference Layer) - A proposal for a web-based representation and inference layer for ontologies, which combines the widely used modeling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. It is compatible with RDF Schema (RDFS), and includes a precise semantics for describing term meanings (and thus also for describing implied information). http://www.ontoknowledge.org/oil/index.shtml.

Ontology - From an IT industry perspective, the word ontology was first used by artificial intelligence researchers and then the Web community to describe the linguistic specifications needed to help computers effectively share information and knowledge. In both cases, ontologies are used to define "the things and rules that exist" within a respective domain. In this sense, an ontology is like a rigorous taxonomy that also understands the relationships between the various classified items.

OWL - Web Ontology Language. Markup language used to specify ontologies for the Internet.

Path - A path is a sequence of consecutive edges in a graph and the length of the path is the number of edges traversed.

P2P or Peer-to-peer - A blanket term used to describe: (1) a peer-centric distributed software architecture, (2) a flavor of software that encourages collaboration and file sharing between peers, and (3) a cultural progression in the way humans and applications interact with each other that emphasizes two way interactive "conversations" in place of the Web's initial television-like communication model (where information only flows in one direction).

Predicate - Of the three parts of a statement, the predicate or verb, is the resource, specifically the Property, which defines what the statement means. See also: subject, object.

Property - A sort of relationship between two things; a binary relation. A Property can be used as the predicate in a statement.

Protocol - A language and a set of rules that allow computers to interact in a well-defined way. Examples are FTP, HTTP, and NNTP.

Range - For a Property, its range is a class which any object of that Property must be in.

RDF (Resource Description Framework) - A framework for constructing logical languages that can work together in the Semantic Web. A way of using XML for data rather than just documents.

RDF Schema-or RDF Vocabulary Description Language 1.0: - The Resource Description Framework (RDF) is a general purpose language for representing information in the Web. This describes how to use RDF to describe RDF vocabularies. This is a basic vocabulary for this purpose, as well as conventions that can be used by Semantic Web applications to support more sophisticated RDF vocabulary description. See http://www.w3.org/TR/rdf-schema/

Reachability - An important characteristic of a directed logic graph which find all paths from every node ni, to any node nj within the graph.

Reasoner – A program that can find new facts from existing data (aka. reasoning).

Resource - That identified by a Universal Resource Identifier (without a "#"). If the URI starts "http:", then the resource is some form of generic document.

Rule - A loose term for a statement that an engine has been programmed to process. Different engines have different sets of rules.

Semantic portal – A Web portal where information resources are indexed in accordance with the constructs of a rich domain ontology.

Semantic Web - The Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it. 'the principle that one should separately represent the essence of a document and the style presented.

Semantic Web Services - Web Services developed using semantic markup language ontologies.

Server - A program that provides a service (typically information) to another program, called the client. A Web server holds Web pages and allows client programs to read and write them.

SGML (Standard Generalized Markup Language) - An international standard in markup languages a basis for HTML and a precursor to XML.

SHOE Simple HTML Ontology Extension - A small extension to HTML which allows web page authors to annotate their web documents with machine readable knowledge. SHOE claims to make real intelligent agent software on the web possible. See http://www.cs.umd.edu/projects/plus/SHOE/

SQL (Structured Query Language) - An ISO and ANSI standard language for database access. SQL is sometimes implemented as an interactive, command line application and is sometimes used within database applications. Typical commands include select, insert, and update.

SGML (Standard Generalized Markup Language) -Since 1986, SGML has been the international ISO standard used to define standards-based markup languages. HTML is a markup language that is defined using SGML. The HTML DTD the specifies HTML is written in SGML syntax. XML is not a markup language written in SGML. There is no pre-defined DTD for "XML Markup." XML is a sub-set of the SGML standard itself **Statement -** A subject, predicate and object which assert meaning defined by the particular predicate used.

Stylesheets - A term extended from print publishing to online media. A stylesheet can contain either formatting information (as is the case with CSS-Cascading Style Sheets, or XSL FOs-XSL Formatting Objects), or it can contain information about how to manipulate the structure of a document, so it can be 11 transformed" into another type of structure (as is the case with XSLT Transformation "style sheets").

Subject - Of the three parts of a statement, the subject is one of the two things related by the predicate. Often, it indicates the thing being described, such as a car whose color and length are being given. See also: object, predicate.

Taxonomy - This term traditionally refers to the study of the general principles of classification. It is widely used to describe computer-based systems that use hierarchies of topics to help users sift through information. Many companies have developed their own taxonomies, although there is also an increasing number of industry standard offerings. Additionally, a number of suppliers, including Applied Semantics, Autonomy, Verity and Semio, provide taxonomy-building software.

TCP (Transmission Control Protocol)-A computer protocol that allows one computer to send the other a continuous stream of information by breaking it into packets and reassembling it at the other end, resending any packets that get lost in the Internet. TCP uses IP to send the packets, and the two together are referred to as TCP/IP.

Transformation - In XSLT, a transformation is the process of a software application applying a style sheet containing template "rules" to a source document containing structured XML markup to create a new document containing a completely altered data structure. UML (Unified Modelling Language)-Derived from three separate modelling languages.

Travel Recommender System (TRS) - Applications that e-commerce sites exploit to suggest travel products and provide consumers with information to facilitate their decision-making processes.

URI (Universal Resource Identifier) - The string (often starting with http:) that is used to identify anything on the Web.

URL (Uniform Resource Locator)-The address of a file or resource on the Internet.

Valid - An XML document is "valid" if it is both well-formed and it conforms to an explicitly-defined data model that has been expressed using SGML:s DTD (Document Type Definition) syntax.

W3C (World Wide Web Consortium) - A neutral meeting of those to whom the Web is important, with the mission of leading the Web to its full potential.

WSDL (Web Service Description Language) - provides a communication level description of the messages and protocols used by a Web Service.

Weblogs - Weblogs (Blogs) are personal publishing Web sites that syndicate their content for inclusion in other sites using XML-based file formats known as RSS. Weblogs frequently include links to content syndicated from other Weblogs and organizations use RSS to circulate news about themselves and their business. RSS version 1.0 supports richly expressive metadata in the form of RDE.

Web portal - A Web site or service that offers a broad array of resources and services, such as e-mail, forums, search engines, and on-line shopping malls. The first Web portals were online services, such as America Online (AOL), that provided access to the Web, but by now most of the traditional search engines have transformed themselves into Web portals to attract and keep a larger audience. 12

Web Services - Web-accessible programs and devices.

Web server - A Web server is a program that, using the client/server model and the World Wide Web's Hypertext Transfer Protocol (HTTP), serves the files that form Web pages to Web users (whose computers contain HTTP clients that forward their requests).

Well formed - A document is "well-formed" if all of its start tags have end tags and are nested properly, with any empty tags properly terminated, and any attribute values properly quoted. An XML document must be well-formed by definition.

XML Schema - A formal definition of a "class" or "type" of documents that is expressed using XML syntax instead of SGML DTD syntax.

XSL (Extensible Srylesheet Language) - XSL has two parts to it: a transformation vocabulary (XSL Transformations-XSLT) and a formatting vocabulary (XSL Formatting Objects (XSL FOs).

XSL FOs (XSL Formatting Objects) - Ihe formatting vocabulary part of XSL that applies style properties to the result of an XSLT transformation.

XSLT (XSL Transformations) - The transformation vocabulary part of XSL. An XSLT "stylesheet" contains template rules that are applied to selected portions of a source document's "source tree" to produce a "result tree" that can then be rendered for viewing, processed by another application, or further transformed into another data structure.