2013

# A Semantic Web Based Search Engine with X3D Visualisation of Queries and Results

Gkoutzis, Konstantinos

http://hdl.handle.net/10026.1/1595

http://dx.doi.org/10.24382/4562
University of Plymouth

# A Semantic Web Based Search Engine with X3D Visualisation of Queries and Results

by

Konstantinos Gkoutzis

A thesis submitted to Plymouth University

in partial fulfilment for the degree of

**Doctor of Philosophy**

School of Computing and Mathematics

Faculty of Science and Technology

Plymouth University

**September 2012**

# Dedication

To my parents Ioannis and Maria,
and my fiancée Evie :)
Thank you.

# Copyright Statement

**A Semantic Web Based Search Engine**

**with X3D Visualisation of Queries and Results**

**by**

**Konstantinos Gkoutzis**

# Abstract


The Semantic Web project has introduced new techniques for managing information. Data can now be organised more efficiently and in such a way that computers can take advantage of the relationships that characterise the given input to present more relevant output. Semantic Web based search engines can quickly educe exactly what is needed to be found and retrieve it while avoiding information overload.

Up until now, search engines have interacted with their users by asking them to look for words and phrases. We propose the creation of a new generation Semantic Web search engine that will offer a visual interface for queries and results. To create such an engine, information input must be viewed not merely as keywords, but as specific concepts and objects which are all part of the same universal system.

To make the manipulation of the interconnected visual objects simpler and more natural, 3D graphics are utilised, based on the X3D Web standard, allowing users to semantically synthesise their queries faster and in a more logical way, both for them and the computer.

# Acknowledgements

I would like to take this opportunity to express my sincere thanks, from the bottom of my heart, to:

My father, Ioannis, and my mother, Maria, for helping me and supporting me, in so many ways, throughout my life. Without them I would not be the man I am today.

My fiancée, Evie, for all her help, patience and support all these years that I have been implementing my thesis. Her presence and impact in my life has given me the strength and courage to go on.

My aunts, Aggeliki and Eleni, and my sisters, Vivian and Stella, for providing all their help and support; each one in their own, unique, way.

My friend, and future brother-in-law, Anthony, for his invaluable help whenever I needed additional support.

My friend, and future best man, Thodori, for his patience and understanding during the long implementation of this thesis.

My second supervisor, Mr Brian Mushens, and my extremely helpful R&I administrators, Carole and Lucy, for always being there for me to answer my questions and help me overcome all problems.

And finally, but more importantly, my exceptional Director of Studies, Professor Vladimir Geroimenko, who has helped me, supported me and made me feel at home even from the first day we met. Thank you for all your valuable comments, suggestions, guidance, patience and understanding over all these years. I hope this thesis will manage to demonstrate all the hard work I have done and make you proud of the results.

# Declaration

- At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

- This study was 100% self-funded.

- Relevant scientific seminars and conferences were attended and work was presented and published.
    - Scientific Publications: 1
    - Relevant Presentations, Conferences and Seminars Attended: 5

- Word count of the thesis: 34,670

**Konstantinos Gkoutzis**

Signed: ____________________

Date: ____________________

# Table of Contents

# Table of Figures

# Chapter 1: Introduction

*"All men by nature desire to* **know***. An indication of this is the delight we take in our senses; for even apart from their usefulness they are loved for themselves; and above all others the sense of* **sight***. For not only with a view to action, but even when we are not going to do anything, we prefer sight to almost everything else. The reason is that this, most of all the senses, makes us know and brings to light many differences between things"*

*– Aristotle (Metaphysics, 980a21)*

## *1.1 Philosophical Overview*

### 1.1.1 From Data to Wisdom

We live in the Information Era, where knowledge is everywhere and everything. The proper management of knowledge, though, presupposes the existence of mechanisms for organising information, in order to match the appropriate pieces together.

When the Internet came along, it offered users the ability to access many different types of electronic data. Unfortunately, those data are still mostly uncategorised and the human mind simply cannot absorb and process the increasingly huge amount of information available. This is the reason why computers are utilised to gather and present information in a more streamlined manner, so that people are able to browse through the findings in a faster and more convenient way.

The basis of computing technology is **data** (Ackoff, 1989). Data are represented by symbols that need to be put into context to make sense. For example, the letter 'A' is a symbol that could either represent "*the first letter of the English alphabet*" or "*the sixth note in the musical scale of C major*" (Longman, 2009).

To move from data to **information**, we require a meaning that can help us translate the symbols into something useful and comprehensible. This meaning presents the relational connection between data and the audience they are meant for and provides answers to simple questions such as “Who? When? Where?” and “What?”.

With the help of meaning providing a context, we receive information that is something useful. We can now relate the given information to other pieces of information and learn it, acquiring **knowledge**. With knowledge we manage to answer the “How?” question, which puts the information into action by applying it to a specific domain, for example “A comes before B in the alphabet”.

It has been suggested that an extra layer should be added to this scheme called **understanding** (Bellinger, Castro & Mills, 2004). This is based on the argument that knowledge is information we memorise, but after we understand the true nature of “How?” it works, then we can answer “Why?” it works. According to their theory, memorising “2 + 2 = 4” is knowledge but answering the question “1267 x 300” requires understanding of the mechanisms behind this.

Together, knowledge and understanding lead to **wisdom**. Wisdom is produced after using and applying knowledge and understanding to many different cases and then reflecting on the results. Now, in addition to the “Why?” question, we can also answer the “When?” question. Expecting or estimating a result or an outcome (i.e. the future) needs wisdom, which derives from knowledge. This is something a human mind learns to do as a person grows up and matures, but with computers it is different.

### 1.1.2 From Computing to Creativity

Computers are programmed by people to respond in a predefined way to specific actions. Artificial intelligence scientists have made several attempts to make computers come up with original ideas but, up to this point, it has not been achieved. Given the fact that computer applications can only follow the rules programmed into them (or the rules that those rules produce), they cannot yet cross the line from **computing** to **improvising**.

Since Search Engines are computer applications too, their search results are bound to be calculated and precise, based on mathematical formulae for text searching. This way, results may match the given words or phrases, but the meaning is totally lost. A search for the word “apple” will return many links to webpages and pictures of apple trees, as well as Apple computers, among many other things. Researchers suggest (iProspect, 2004) that in the circumstance of the results being too numerous, users usually go through the first few pages and then either accept what they found as the truth or just give up.

The most practical method to deal with this problem is to help the computer “understand” what each piece of information “means”, by using the relations between interconnected objects. The Semantic Web project is trying to address this issue by grouping information into Ontologies, which are logically organised datasets, a formal and well defined version of Taxonomies (Pidcock & Uschold, 2003). This means a user could specifically search for the fruit called apple, retrieving more relevant results.

A few Semantic Web based Search Engines already exist ('Swoogle,' 2007) ('Hakia,' 2012) but, currently, these search engines lack a way of visualising queries and results, in order to make ontology browsing simpler and more understandable.

## *1.2 Problem Overview*

Computer users have grown accustomed to the habit of text search engines where they must type in keywords which they believe are related to their search. Those keywords are then matched to the index of the engine by using proprietary algorithms and a result list is produced, usually ordered by relevance or link popularity.

Semantic Web based search engines are trying to redefine this procedure. They may still be asking the user to type in text but they match the given keywords to ontology objects, thus making results more logically relevant rather than just plain keyword relevant (Li et al., 2004). This is feasible because ontologies organise items based on logical connections derived from the item meaning.

This thesis deals with adding an extra layer of abstraction on top of the keyword searching Semantic Web search engines currently offer. This is achieved by taking advantage of modern visualisation techniques already widely used by other types of applications. In other words, the thesis will combine the Computer Science sections of **Semantic Web** and **Search Engines** with that of **Information Visualisation**. The final result will provide a graphical way of accessing search engines so that users will be able to create search queries and view results in a more visualised manner.

## *1.3 Objectives and Contributions*

The main topics explored and defended in this thesis are:

1) The introduction of an additional **visual searching layer** on top of Semantic Web ontology queries, offering a graphical representation of the underlying system data as an alternative to plain text search, resembling the sorting and sifting of “tangible” items.
2) The creation and evaluation of a visual searching mechanism entitled “**Semantic Synthesis**”, aiming to provide unified visualisation and manipulation of queries and results, seamlessly bridging the two using interconnected graphical items (concepts and objects).
3) The design of a **customisable visual searching interface**, based on the style and preferences of each user, which learns from and adapts to the searching habits of members, personalising item popularity on a per user basis.
4) The suggestion of a **game-like user ranking scheme** that rewards constructive use of the system by assigning appropriate percentage levels, titles and colours to each member, depending on the usefulness of their syntheses, as backed by the social community of the system.

The four contributions mentioned above are the “cornerstones” of this PhD and are repeated throughout this dissertation. A schematic representation of these contributions can be found in Figure 1.

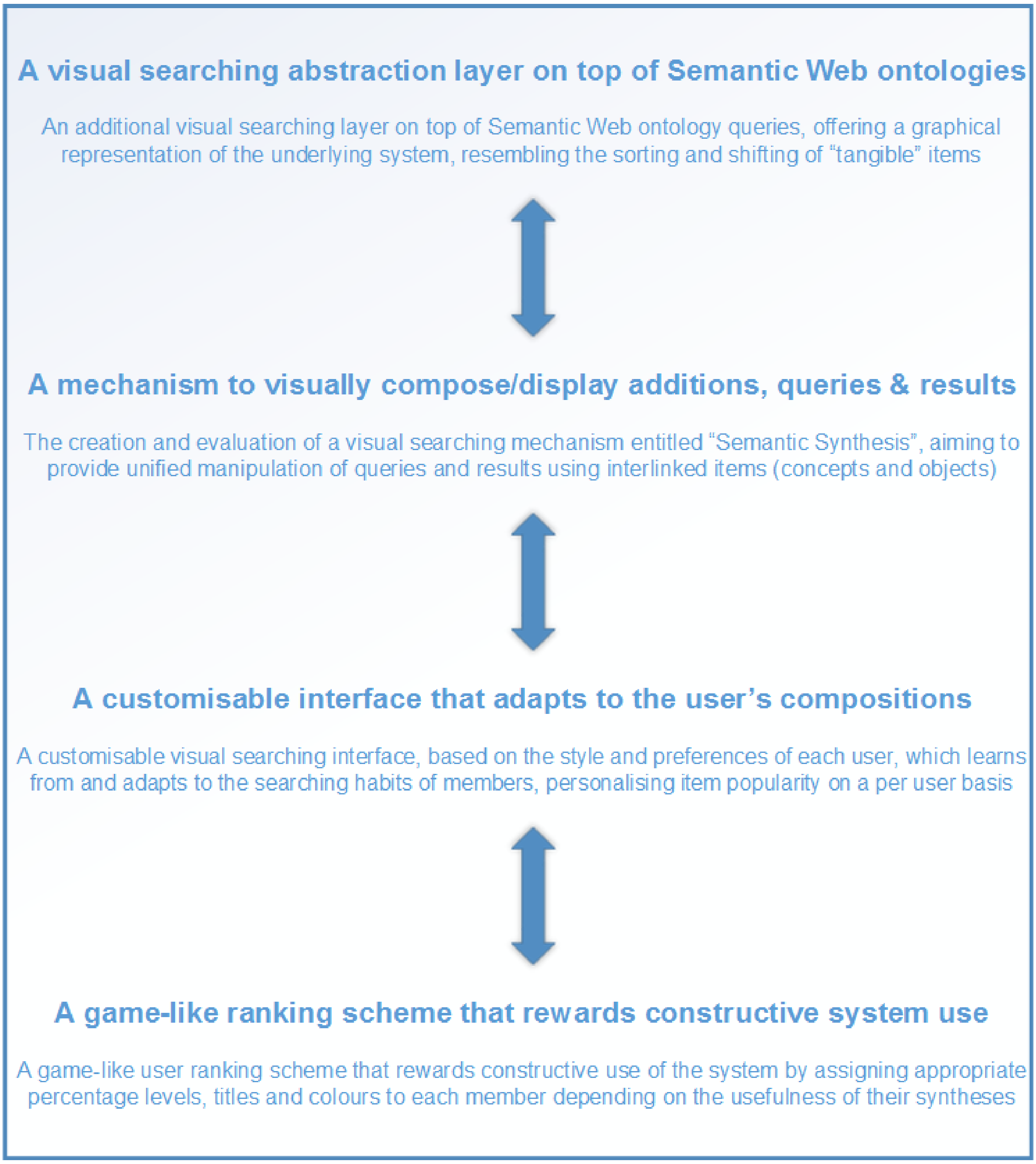



**Figure 1: The Scientific Contributions of the thesis**

The aforementioned visualisation of items aims to help users "synthesise" their compositions in such a way that, each time, the computer will deal with a very specific set of instructions (i.e. query), thus attempting to return more **well-filtered** results. This is going to be further analysed in the following chapters.

## *1.4 Thesis Structure*

This dissertation is divided into the following chapters:

- Chapter 1 presents an introduction to the thesis, stating the objectives and the contribution to the scientific community.
- Chapter 2 presents background information of scientific material and work related to this dissertation, as well as relevant technologies and techniques.
- Chapter 3 introduces the proposed system via a problem statement, which identifies the issues addressed in this project, and the proposed solution, which describes the steps taken to address the problem.
- Chapter 4 presents the requirements analysis, based on the previously defined problem statement, as well as a variety of factors closely related to the system.
- Chapter 5 contains the original 2D designs of the system and its elements, along with an evaluation of the searching mechanism via a user study.
- Chapter 6 presents the suggested development process that the defined system should follow in order to be implemented and commercially launched in the future, including 3D designs of its suggested interfaces.
- Chapter 7 sums up the entire dissertation, with a discussion on conclusions and scientific contribution, an evaluation of the outcome and suggestions for future work and expansions.

## *1.5 Summary*

In this chapter, an introduction to the thesis was presented, providing general information about the philosophy behind the subject, while stating the objectives and the contribution of this dissertation to the scientific community.

# Chapter 2: Background & Related Work

## *2.1 Historical Breakthroughs*

### 2.1.1 Moving Away from Paper

In the past, before the digitalisation of written work, documents were compiled using pen and paper, which a print-shop would then typeset into some other medium. Books and newspapers were synthesised letter by letter using special glyphs and afterwards they were printed out using letterpress machines. This was a long and tiresome process for a compositor, who would work on it manually for hours.

In order for writers to explain to the compositor how they wanted each section to look like, they cooperated with a proofreader to “**mark up**” the manuscript with notes. To achieve this, special printing instructions were added alongside the text, which specified what the result should look like. Font style, size, text indentation as well as various other attributes, were defined based on these notes.

Many years later, the word “**markup**” would take a totally new meaning. Document markup languages started to be created during the computer era so that the print format of digital documents – both on the screen and the printer – would change, according to the writer’s choices. After many different attempts, the result was the creation of several markup languages that provided extra information about the data as well as their presentation. This transition is briefly presented below.

### 2.1.2 The Creation of "The Web"

When computers were used for text editing for the first time, electronic documents used to contain control codes or macro commands that defined the desired format of the entire file. In the late 1960s, a new concept was proposed called **Generic Coding**. The idea behind it was that descriptive tags could be used inside the text, which would differentiate the segments from one another (for example: "heading", "footnote" etc.).

This proposal is credited to William Tunnicliffe, who mentioned it during a meeting with the topic "The separation of the information content of documents from their format" at the Canadian Government Printing Office in September 1967. At the same time, Stanley Rice, a New York book designer, made the suggestion of "a universal catalogue of parameterised, editorial structure, tags" (Goldfarb, 1996). Hence, the demand for a more organised presentation of data was becoming obvious.

William Tunnicliffe was at the time chairman of the Composition Committee of the Graphic Computer Communications Association (GCCA), which was then renamed to Graphic Communications Association (GCA) and is now called IDEAlliance. Norman Scharpf, who was then the active director of the GCA, created a Generic Coding project (called "The System X Project") which was under the control of Composition Committee and Tunnicliffe began working on this new idea.

The result of the project was **GenCode**, a language that defined a set of standards for the markup of digital documents. Shortly after its inauguration, another markup language called **SGML** came along, so the GCA aborted GenCode and started supporting the standardisation of this new method.

SGML history can be traced back to the late 1960s, when Charles F. Goldfarb was an attorney who IBM had hired to install accounting systems for businesses. After some time he moved to the company's Cambridge Scientific Centre, which was researching the possibility of applying computers to legal practice. Since there were a lot of documents involved, he and his colleagues (Ed Mosher and Ted Peterson) decided to organise them by implementing document markup languages for the various existing operating systems. After some attempts, Steve Furth, a senior IBM industry marketing manager, referred Goldfarb to Stanley Rice's work which inspired him to create a new, unified markup language.

Charles Goldfarb, Ed Mosher and Ray Lorie invented the Text Description Language (**TDL**) as part of IBM's Integrated Text Processing project, and the first prototype used the name Integrated Textual Information Management System (**InTIME**). After Norman Scharpf, who was a former IBM marketing manager, saw a demo of InTIME, he invited Goldfarb to a meeting of the "System X" committee. There, Tunnicliffe and other GCA representatives discussed the future of markup as well as the project that would eventually produce GenCode, and decided to follow a different path from IBM for a while.

A couple of years before the release of InTIME, Goldfarb renamed it **GML** (Generalized Markup Language), a name which he made up from the surnames of the creators (Goldfarb, Mosher and Lorie). In 1973 it was released along with IBM's Advanced Text Management System (ATMS), a text processing system that was the first to make use of the advantages of GML.

By then, Ed Mosher had already developed the "first production quality" (Goldfarb, 1996) Document Type Definition (**DTD**) for the manuals of IBM's

Telecommunications Access Method (TCAM). Goldfarb created a validation parser that read a DTD for GML and proposed the creation of a global standard. Thus, in 1978, SGML (S for Standard) was born.

It took SGML eight years to become an ISO standard (ISO-8879) but this set the basis for all markup languages up to date. After a while, the GCA also joined the efforts and it is still one of the most important supporters of this standard. Still, SGML did not become very popular until someone else decided to use it for a network documentation project called World Wide Web.

The World Wide Web (**WWW**) is a project that aims to provide global access to information and was created by Sir Tim Berners-Lee while he was working at the CERN institute in Switzerland ('World Wide Web CERN Proposal,'). The WWW utilises the HyperText Transfer Protocol (**HTTP**) which defines specific resources using Uniform Resource Identifiers (**URIs**) to grant access to objects from multiple interconnected networks. In order for heterogeneous computer systems to have a common language to read and share information, Tim Berners-Lee invented the HyperText Markup Language (**HTML**) in the early 1990s, based on the SGML standard.

This new markup language instantly became amazingly popular and more advanced Web Servers were developed in order for businesses and organisations to share documents on-line and promote themselves through the WWW. More and more websites were popping up every day and the face of the Web gradually changed from a large document exchange network into an interconnected global “village”. Additionally, the World Wide Web Consortium (**W3C**) was created by Tim Berners-Lee in MIT, to manage and update the Web protocols.

HTML made SGML popular and managed to get many people involved in the movement for sharing information from all around the planet. However, it was quite limited by design (Alesso & Smith, 2006), so computer scientists started designing new markup languages for the Web to use.

### 2.1.3 From "The Web" to "The Semantic Web"

Although HTML was the killer-app development of SGML, which defined the initial form of the World Wide Web, it only offered specific facilities for customisation of the presentation. Apart from this issue, Web designers gradually realised the need for including extra data in the webpages in order to define the information contained inside. For example, the language in which a page was written had to be declared somehow so that each Web browser could display the local characters correctly. Also, there was the need to add keywords and descriptions about the contents of each page which would define the contained objects without mixing up with them.

This led to the idea of using hidden tags that would characterise each document or Web page and would only be visible upon request. They are the data that "follow" (come after) the actual data, carrying definitions and extra information. They are now universally known as **metatags** which provide the webpage **metadata**.

The term metadata derives from the Hellenic word "meta", which means "after", and the Latin word "data", which has come to mean "unprocessed information". A terse definition for metadata states that they are data about data, meaning that this type of data is meant to describe other data (even other metadata). To put it simply, any piece of data that is meant to describe other data, for example by providing further information about them or presenting their connection to something else, is considered to be metadata. Everything else is just data.

The term itself has an interesting background story. In the early 1970s Jack Myers founded the company Metadata Information Partners (now called The Metadata Company) and requested a trademark for the word “Metadata” which he received in 1986. As early as 1984, the term metadata was being used by Information Technology (**IT**) specialists to refer to extra information about data, especially in the ever-developing Database Systems. Although Myers’ lawyers have tried to keep this word for private use, some consider it to have entered the public domain because of how widespread it now is in the IT world.

Tim Berners-Lee borrowed the metadata definition from the computer database world and used it as a tag in HTML (the <META> tag). This tag offered extra information about each webpage and for a while it was extremely popular, especially in the Search Engine field. Unfortunately, it was suggested that the META tag offered only limited information that a search engine could use, so most engines now ignore it completely and base their results on other metrics (Goodman, 2002). But this did not signal the end of metadata.

In March 1995, the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) held a joint workshop in Dublin, Ohio called the “OCLC/NCSA Metadata Workshop”. There, they discussed about how a core set of semantics for Web-based resources would be extremely useful for categorizing the Web for easier search and retrieval, which resulted in the creation of the Dublin Core Metadata initiative (**DCMI**). The DCMI has since defined a set of standards to describe resources, using tags such as Title, Date, Creator, Language etc. Their latest set is an official standard of the National Information Standards Organization (NISO) since 10 September 2001.

Apart from DCMI, other efforts have also consolidated towards the use of metadata. W3C's Metadata Activity initiated four projects: PICS, CC/PP (superseded by Device Independence), DSig and P3P.

The Platform for Internet Contents Selection (**PICS**) is a W3C project that aims to provide a labelling mechanism for Internet content according to predefined rating criteria. This effort wants to prevent minors from accessing inappropriate content on the Web and help parents have more control over the Internet experience of their children. This tagging mechanism was quite popular for a while but it has not managed to become a very widespread standard. Both **DSig** (also known as XML Digital Signature) and **P3P** (Platform for Privacy Preferences Project) are now being developed using XML to express their metadata. The latest version of PICS has been rewritten using the relatively new metadata framework called RDF.

**RDF**, which stands for Resource Description Framework, is a language for representing information on the Web. Using specific structures, it depicts objects and their direct relations to each other, so that indirect relations will be able to be inferred based on predefined logic mechanisms.

Since February 2001, RDF has become part of the **Semantic Web Activity** (Berners-Lee, Hendler & Lassila, 2001), which is a project that aims to model the meaning of all digital information. The leap between plain old custom metadata into a uniform well-defined metadata model was achieved with the help of a new markup language called XML.

**XML** stands for eXtensible Markup Language and is a general purpose markup language that provides the facilities for creating custom tags. This allows the creation of case specific tags or even entirely new markup languages. The Semantic Web Activity has used XML to standardise the languages RDF, RDF Schema (**RDFS**) and **OWL** (Web Ontology Language) (Alesso & Smith, 2006) to fulfil its needs of defining ontologies

**Ontologies** are collections of interrelated objects, interconnected based on their meaning. The word derives from the Hellenic words “ontos”, which means “whatever exists”, and “logia” which means “theory” or “science”. It has been stated that “*a formal ontology is a controlled vocabulary expressed in an ontology representation language*” (Pidcock & Uschold, 2003). A **controlled vocabulary** is “*a list of terms that have been enumerated explicitly*” and, most importantly, each list is controlled, updated and monitored by a registration authority.

Finally, adding to all the aforementioned technologies, Linked Data was introduced by Tim Berners-Lee in 2006 (Berners-Lee, 2006) in order to define specific Dereferenceable Uniform Resource Identifier addresses (URIs) for each piece of data. This way, each Semantic Web object (also known as “things”) can be defined by its own unique, unambiguous, Web address (Bizer, Heath & Berners-Lee, 2009).

### 2.1.4 From “Search Engines” to “Semantic Web Search Engines”

The idea of searching through a vast index of endless information can be traced back to the $19^{th}$ century philosophical movement related to the “Akashic records”. People with religious philosophy and metaphysics backgrounds discussed about an illimitable incorporeal library of information that can be accessed and searched through from any place in the world by using astral projection, a state of mind achieved with an out of body experience.

Many years later, near the end of World War II, Dr. Vannevar Bush, who was then the director of the Scientific Research and Development Office in the United States of America, suggested in his article “As We May Think” (Bush, 1945) the need to create a global record, indexing the body of knowledge composed of books and publications. At that point, the amount of written information had already increased very much, especially with scientific war innovations, so a centralised catalogue was required to help researchers find what they needed more quickly. Bush called this hypothetical system “Memex”.

In early 1960, Gerald Salton and his team at Cornell University created “SMART”, an information retrieval system that used many custom algorithms to search and index words in documents contained in the local machine. Additionally, Salton played an important role in information retrieval and in setting the foundations on word indexing for future search engines, with concepts such as Vector Space Model, Relevance Feedback Mechanisms, Term Weights and many others (Cornell University, 1996).

Almost in parallel with the SMART system, Ted Nelson founded “Project Xanadu” which aimed to create a network of interconnected machines specifically designed to store, link and search through documents (Barnet, 2005). He also defined for the first time the Computer Science terms “hypertext”, “hypermedia” and “virtuality”, which are still widely used today.

Many other information retrieval systems were developed in the late 1960s and early 1970s (Rowley, 1992, pp. 335,382) seeking to exploit the advantages of computer processing in maintaining large indexes, digital databases and information services. These systems were preferred because they offered a cost-effective approach to information control, permitting greater quantities of information to be sifted and directed appropriately to the end-users.

Shortly after the Personal Computer revolution, in 1990 a group of students at McGill University created “Archie”. This was a UNIX tool that connected to FTP servers, copied a listing of their contents and indexed them in local files for faster searching (Sonnenreich, 1997). Users could search for keywords in the filenames indexed from the FTP servers and then connect and download on demand the files found. Although this method did not scan or search inside the contents of the files, Archie is still considered to be the first Internet search engine.

It was not until the invention of the Web by Tim Berners-Lee (as mentioned in Section 2.1.2) in the early 1990s, that the global information system everyone dreamed of was finally created. At first, though, there was no way to quickly search through the constantly increasing number of Web documents (Schwartz et al., 1992).

The Web quickly became a “killer application” first for educational institutions and organisations and then for corporations. The number of webpages increased drastically through the years and new search engines appeared to assist the users track down documents faster.

Search engines used “Web crawlers” or “Web spiders” to index websites. These are automated applications that connected to webpages, scanned their contents, indexed their keywords and identified the links to other webpages, so that they could subsequently move on to those links. After this was complete, the search engine would search among the indexed keywords and present the resulting webpages based on the users' requests. Nowadays, these scanning applications (now also called "Web robots") have evolved to track down even links hidden inside scripts, avoid data duplication and even identify and remove possible malware websites.

The first WWW search engine is considered to be ALIWEB (Archie Like Indexing for the WEB), created in November 1993 by Martijn Koster (Sonnenreich, 1997). Koster also created the "Robots Exclusion Standard", a method to direct Web robots to specific pages of websites. In parallel with ALIWEB, "The Wanderer" Web indexer script, created by MIT student Matthew Gray, created the website index for the "Wandex" search engine. Both search engines did not become very popular.

The most popular search engines of the early Web days were "Yahoo! Search" (March 1995), "Lycos" (April 1995) and "AltaVista" (December 1995). Countless others started up as the WWW became more popular, but most have now ceased to exist, mainly due to high maintenance costs, company buyouts or a lack of resources.

Among many others, it is worth mentioning the "Open Directory Project" (also known as ODP or Dmoz), which was launched in June 1998 and has been since then trying to create a hierarchical ontology scheme to organise all site listings. At the beginning, ODP borrowed the hierarchical listing style of Usenet groups, but this changed in time. After the Semantic Web movement became more popular, ODP switched to using an RDF ontology, which is still constantly updated (Netscape Communications Corporation, 2004). ODP was created by two Sun Microsystems employees, Rich Skrenta and Bob Truel, and is currently owned by Netscape, which is owned by AOL, which in turn is owned by Time Warner.

In September 1998, Google was founded, based on a project started in January 1996 by two PhD students at Stanford University, Larry Page and Sergey Brin (Google, 2009). The indexing functions it used to store the webpages, in combination with its speed search algorithms and its proprietary result ordering ranking system, made it the top choice for users all around the globe. For the past 10 years, Google has dominated the search engine market and receives more than 800 million hits per day (Alesso & Smith, 2006).

With the rise of the Semantic Web movement, a new need was created for search engines that would search inside the newly created “ontologies”. One of the first functioning Semantic Web Search Engines was Swoogle, created in 2004 by the University of Maryland with funding from United States government agencies such as DARPA and the National Science Foundation. As mentioned in its initial publication (Li *et al.*, 2004), “*Swoogle is a crawler-based indexing and retrieval system for the Semantic Web. It extracts metadata for each discovered document, and computes relations between documents*”.

Also in 2004, Hakia was founded, a company that launched another Semantic Web search engine that uses a proprietary indexing technology to list results. With the guidance of Victor Raskin, a professor of linguistics at Purdue University, Hakia combines semantics with mathematics to provide faster and more relative results (Raskin et al., 2008). Hakia has also created its own custom Semantic API (Application Programming Interface) for programmers to use and to make the most out of their natural language searching platform.

In 2005, Powerset was founded and started creating a natural language search engine for the WWW (Helft, 2007). In order to achieve this, webpages are indexed inside Semantic Web ontologies which are then queried based on the users' requests. In May 2008, Powerset presented an online working version of their product, used to search faster inside the articles of the online user-updated encyclopaedia Wikipedia. In June 2008, Powerset was acquired by Microsoft and was eventually launched as “Bing”.

Also in 2005, Metaweb Technologies was founded (Markoff, 2007) and created Freebase, which was presented on the WWW in 2007. Freebase combines the concepts of Semantic Web ontologies with that of Social Networking communities. It aims to create a global database of categorised information, like ODP, which will be updated by its users, like Wikipedia. Metaweb and the Freebase project are currently supported by many venture capital companies.

Since then, a few Semantic Web search engines have been created, some supporting natural language search or just keyword search. Some examples, among other attempts, include “Watson” from the Open University, “Falcon” from Southeast University of China and “SWSE” (Hogan et al., 2007) from the Digital Enterprise Research Institute.

### 2.1.5 Blogs, Social Networking and the “Web 2.0” Trend

Soon after the Web became available to the public in the early 1990s, many users became interested in creating their own personal website to present themselves, their businesses or their interests. Certain companies began giving away size-limited Web space for free on their servers and in exchange they placed advertisements on top of each webpage. Some notable examples include “Tripod” (1992) which is now owned by Lycos, “Geocities” (1994) which is now owned by Yahoo!, as well as “Angelfire” (1995) which was subsequently bought out by Lycos.

Web “citizens” began interacting with each other, linking their websites, exchanging virtual “awards” and comments in electronic “guestbooks”, as well as publishing their personal thoughts, ideas, pictures or even poems and passages. Separated from the commercial websites, the personal websites had a life of their own, either organised in communities or independently posting articles and news of interest to this new “global village”.

Near the end of the 1990s, users started calling their personal websites “blogs”, a contraction of the words “Web” and “log”, leading to the creation of a new trend in the Web culture (Wortham, 2007). Many user websites changed form to look like online journals and became even more personal and popular, since the initial fear of the unknown, and sometimes dangerous, Web, had now been replaced by colourful blogs and online shopping.

This eventually led to the creation of the first “online social networks” in the early 2000s, a term also popularised as “Web 2.0” by Tim O’Reilly of O’Reilly Media (O'Reilly, 2005). Social networks reused all the well-known technologies of the Web but, instead of offering free Web space, they prompted users to upload all their information and material inside a standardised personal profile page. The website that became most popular for taking advantage of this niche was MySpace which was founded in 2003 and now counts almost 31 million active accounts. Other notable examples followed, such as Facebook which was established in 2004 and now has almost 1 billion active accounts and Twitter (est. 2006), which has more than 500 million active accounts.

MySpace became the centre of attention because famous music bands and movie stars promoted their work by creating personal profiles, although they were usually updated and handled by their managers. On the other hand, Facebook prompts its members to use their real names and only communicate with people they already know or they are interested in networking with. Both websites are still quite popular, although people have grown more sceptical concerning what information they should share with the rest of the world (Hogben, 2007).

### 2.1.6 3D Environments and Virtual Worlds

With the rise of the computer age in the mid-20th century, every type of business tried to get the best out of the new technological achievements. Even though at first computers only had text command interfaces, they eventually became able to present graphics. These graphics were initially plain, but as computer capabilities increased, so did the complexity of the designs.

From published photography to architectural plans, computer graphics are prevalent because of their accuracy and detail. Modern day applications offer easy-to-use tools to create effective graphic models quickly and inexpensively. Apart from common two-dimensional designs, three-dimensional graphics have captivated the users' interest and have been used in movies, product presentations as well as computer applications. These techniques are called Computer-Generated Imagery or **CGI**.

The computer game industry has taken advantage of 3D graphics and moved from classic point-and-click games to 3D gaming which allows freer camera movement, scale alteration (zoom in/out) and viewing angle switching. From first-person shooters (FPSs) to massively multiplayer online role-playing games (MMORPGs), 3D graphics improve the gaming experience by offering a more realistic gameplay to the users.

With the Social Networking niche in mind, companies decided to combine 3D gaming environments with online communities. Examples include Second Life, Active Worlds and IMVU, among others. Additionally, Sun Microsystems began working on MPK20, a 3D distance working environment that aimed to help employees working from home to feel closer to the company. These programs, though, require the users to download extra Windows applications, which execute separately from the Web browser in order to access the 3D environment. This fact not only prevents application mobility, but also isolates the actual program from the Web realm, leaving only the account management on the company website.

A notable attempt to combine browser based 3D graphics with a social network was Google Lively, which was launched, popularised and shut down in 2008. Lively used Flash, as well as a proprietary plug-in, in order to execute inside the browser and the final result was a fully in-browser 3D experience. Despite its initial success, Google decided to discontinue Lively shortly after its inauguration to focus more on their core search, ads and apps business (The Lively Team, 2008). A similar Flash application called “Smeet” still exists, created by a German company.

As far as 3D search engines are concerned, there have only been a few attempts. In 2005, a company called “INOZON” announced (Henderson, 2005) that they would be creating the first 3D based search engine, running inside a browser, but eventually the project fell apart for unknown reasons. Another example was “Ergo” (Invu, 2008) by Invu ('Invu,' 2012), launched as a beta version in 2007, running in an application outside the browser, offering a visual environment for searching information.

Although all the aforementioned attempts are recent, a Web language to express 3D graphics started to be formed as early as 1994 and was called **VRML** (Virtual Reality Modelling Language). Since that time, it managed to become a standard and reached version 2.0, before being succeeded by X3D.

**X3D** is an XML based language which aims to popularise a file format to display 3D graphics by using XML syntax (Geroimenko & Chen, 2005). This makes 3D graphics ideal to parse and present inside browsers or to use with any type of XML related API. At this moment, though, X3D is not widely used and its commercial utilisation is limited. Also, Web browsers require extra add-ons to display the X3D environment. Being standardised with HTML5, the latest version of HTML, will help this format to gradually gain more recognition, while the X3DOM project ('X3DOM,' 2012) is already trying to bypass the need for a plug-in.

HTML5 will also bring instant embedded support of WebGL, a language based on JavaScript that offers in-browser interactive 3D graphics, as well as CSS3 which will also offer 3D transformations. The basic aim of HTML5 is to be able to run in all types of devices and support the creation of Web applications without the need of extra plug-ins. This will bring the Web to an entirely new level.

## *2.2 Technologies and Techniques*

The Semantic Web science field is setting the standards for adding “meaning” to the Web content. In order to achieve this, XML is used for the basic syntax of the languages used to express the semantic models (Bratt, 2007). The most prevalent language to express these models, called ontologies, is RDF, which is responsible for defining the included objects and their relationships.

On top of RDF, RDF Schema (RDFS) is used to describe the properties of the ontologies, which are called their vocabulary (Antoniou & Van Harmelen, 2008). OWL has now prevailed instead of RDFS, because it adds more components and capabilities for expressing ontology properties.

In order to query these collections of interconnected objects, **SPARQL** is used, which stands for SPARQL Protocol and RDF Query Language, and is specifically designed to retrieve results from ontologies. Additionally, a newly proposed semantic standard called **RIF** (Rule Interchange Format) aims to standardise a format to express the rules and conditions of ontologies, so as to make their manipulation more portable.

Semantic Web search engines use ontologies to store their data objects and semantic languages to express their connections. According to a study (Andrei, 2002), search queries can be either: navigational (want to reach a specific site immediately), informational (want to find information from various webpages) or transactional (want to complete some Web related task). Normal search engines, as well as Semantic Web ones, usually bear in mind these types of queries and strive to fulfil the needs of all their users. Either by using plain keyword based text search or with the help of natural language processing text search, they provide users with the search results that fill the users' criteria better.

At the present time, though, no Semantic Web search engine has created a 3D visualisation environment to synthesise these queries or to display the final results. This can be achieved by adding digital graphics on top of the text search and making queries and results visual, rather than just textual.

There are currently many ways to add graphics to a search engine, depending on many factors such as scalability, network limitations, target audience, and so on. Most websites currently use Adobe Macromedia Flash objects to display lightweight graphics on their pages, which also allow user interaction. Other websites take advantage of the Java Virtual Machine and present Java Web Applets, which might take more time to load than Flash, but provide more functionality capabilities. In order for the graphics to be effectively three-dimensional without the use of these browser-embedded mechanisms, extra add-ons must be used.

With X3D, the successor of VRML, XML based 3D graphics can be created which can currently load inside a Web browser by using extra plug-ins. Since X3D is a fairly new standard, it is not yet widely used, so mainstream Web browsers have not added the needed plug-ins by default, but this might change in time as new X3D applications are created. Surely, it is worth exploring X3D in combination with Semantic Web ontologies, since they are both XML based and they were designed to present information through the Web in a portable manner.

## 2.3 Summary

In this chapter, the background information of scientific material and work related to this dissertation were analysed thoroughly, in order to help the reader understand the historical breakthroughs and scientific bases this thesis utilises.

# Chapter 3: Motivation & Problem Statement

## *3.1 Thesis Motivation*

From reading all the aforementioned historical breakthroughs, it is obvious that Computer Science has come a long way in the last half century. Its subfield of Web Science is setting the standards for Web communications and online document management. The World Wide Web has become accessible to almost everyone to read and download, to write and upload or to execute online applications.

New websites and webpages are added every second to servers connected to the, now, vast and immense WWW. Some may contain global news and facts mentioned inside objective articles. Others may have new pictures, sounds and videos from art exhibitions or landscapes around the globe. But most will probably be social communications of no general informational interest or subjective articles written to fulfil the needs and desires of specific individuals or corporations.

New search engines are also constantly created. They either add something new to the field or they just reuse old standardised technologies with the intent to promote sponsored products and companies. Sadly enough, a few years back, a new market arose where search engine rankings and result positions went on sale to the highest bidders, sacrificing relevant results for highly paid search engine marketing advertisements.

This marketing frenzy is also obvious from the great number of unsolicited advertisement messages transferred online (also known as “spam”), which keeps increasing through the years. By using e-mails, Web popup ads, instant messaging applications as well as many other mediums, spam is “flooding” the Internet, “drowning” useful information in its path.

There is a need for a search engine that will not only check indexed keywords for similarities to a given input, but will actually classify information to specific categories of a global catalogue. This catalogue should be produced by combining sub-catalogues, preferably maintained by consortia of experts on particular fields, thus “weeding out” any type of spam. These consortia, however, should not be too official so as to avoid update delays due to bureaucracy or excess formality issues. For example, organisations such as Wikipedia or Open Source Software communities, have managed to maintain quite a high level of data quality, even though they are not always edited by professionals, by managing the “wisdom of the crowd” (Kittur & Kraut, 2008).

At this point in time, organised catalogues can be found in the form of Semantic Web ontologies (Antoniou & Van Harmelen, 2008), even though no global standard exists for their form. The final search engine created should be able to use any given ontology, but using the proper ones will make it work more correctly. The Semantic Web is still growing so hopefully standardised catalogues will be eventually produced (Lenat et al., 1990, p. 48).

Adding a visual environment to this search engine is not a simple task, because it does not only concern building a typical 3D setting. Before the creation can begin, the visual elements that will constitute the search engine must be determined and finalised. These decisions will be taken based on the philosophy behind categorisation and classification, since everything – both material and immaterial – should be included and indexed. A concise interface should help the users set up a visual search query in seconds and the final result should inform them instantly regarding the outcome and also allow them to search even further, if needed.

In addition to the interface being concise, it also has to be easy to use and to navigate through. Usability will be an important factor of this new age search engine, in order for it to be accepted by the general public and integrated into everyday life. It has to combine simplicity with scalable complexity, based on customisable options, while maintaining the same effective search experience for all users.

The final result should be a full-feature **visually searchable catalogue** that will aim to produce useful and relevant results quickly and accurately. This means minimising unsolicited advertisements, preventing misinformation and yet presenting all possible views of each subject, both concisely and thoroughly, according to the users' needs.

## *3.2 The Future Web: Social, Semantic and 3D*

What the Web needs now is a new idea, to make users feel that all these advancements can indeed improve the current status quo.

Friendster, MySpace, Facebook, LinkedIn, Twitter and other well-known Social Web giants became increasingly popular because they responded to the need of people to keep in touch with their friends or acquaintances and participate into activities together, like groups or games. Many companies, such as YouTube, Flickr, Delicious and others, followed the Social paradigm by adding mechanisms to increase community participation and cooperation.

What all of the above websites have in common is that they categorise their data by using Folksonomies. This term was coined around 2004 ('Folksonomy Source,') from the words **folks** and **taxonomy** and basically defines the method of using keywords to describe the content of a data object. Blog articles, images, video and sound files stored all over the Web have been annotated using the words that uploaders chose to describe their data. These words may or may not be related to the content, depending on the perception (or mood) of each uploader. It may be an extremely fast way to annotate data but, due to the lack of standards, many tags are vague, misspelled or simply not appropriate.

When the Social Web eventually merges with the Semantic Web, the users will be the first to benefit because, after the semantics have been agreed upon, data sharing will be seamless and instant. It may take longer to annotate data than plain old keywords, but visualisation techniques can be used to speed up the process.

Introducing 3D visualisation techniques to the Social Web would be a start but by itself it would only appear to be a “cool” looking improvement of the graphics. There has to be essential change in the way we perceive data annotation and information acquisition. The Social community is a power strong enough to fill websites with a plethora of multimedia files gathered in many different ways. The users only need to become accustomed to a new method of tagging data, in order to make them part of a Semantic Ontology.

This is why the search engine we propose should be accompanied by the creation of a system that will have the properties of the Social Web, but with a 3D Web interface on top and a Semantic Ontology behind it. It will be a new age Web application, merging the aforementioned technologies into a 3D Social community, which will constantly be improving an underlying, unlimited, Semantic Ontology.

### 3.2.1 Creating an Information Driven Community

The proposed system will have a search engine which will not only check indexed keywords for similarities to a given input, but will also classify data to specific categories of an unlimited catalogue. This catalogue will be created gradually as users add new multimedia objects or change the already existing ones. Some organisations, such as Wikipedia, Dmoz or Open Source Software communities, have managed to maintain quite a high level of data quality, even though they are not always edited by professionals, but by earnest individuals who dedicate their time to adding and correcting the website material.

The users of this system will enjoy the advantages of a 3D Social community, such as 3D avatars and personalised virtual "lounges", and at the same time will contribute to the creation of an infinite and global source of knowledge. This way, not only they will be able to participate in Social Networking activities but they will also play an important role in rating the quality of the information available in the system by voting up or down additions and changes based on the correctness of the submission.

This participation will make the Web community more active and raise awareness against spam and phishing attempts because data will be immediately comparable due to the nature of ontology objects. Users will be required to make all changes and additions using their real name, which will radically reduce the cases of "trolling" (i.e. adding inflammatory or off-topic data).

Using an overall reputation based scheme, the level of user access to the system will be altered depending on how their additions and changes are ranked. Thus, material that contains mistakes or is miscategorised will be voted down and gradually replaced by correct ones, while at the same time the overall reputation status of the uploader will decrease, restricting their access to website features.

### 3.2.2 Making Data Management User Friendly

The most interesting factor of the system is that the users do not have to possess any type of Semantic Web or XML knowledge in order to use it. The 3D Web interface will help them to add, change or search for specific website objects by creating new compositions. In order to manage or look for data, users will select 3D objects and connect them to each other, thus graphically compose what they are adding or seeking.

Users can customise their graphical user interface by selecting which objects they would like to have immediately available, thus helping them to speed up their common or popular searches. An example interface, based on the possible selections of a user, is presented in Figure 6 in Section 5.1.

Registered users can select which objects they would like to see inside their custom workspace at all times. These private user workspaces are called “Home Lounges” and every registered member of the system has one. While in there, they can create compositions of objects to either search for existing entries in the system, or to add their own.

An example composition of three objects is presented in Figure 7 in Section 5.1. More specifically, it has been prepared to refer to a “Person” who is working for a

“Business” in a specific “Location”. The user added all the information they knew about the objects and they can now either search if the specific composition exists in the system, or add it using their account id as the original submitter.

We call this method “Semantic Synthesis”, from the Hellenic word “synthesis” which means “composition”. It is interesting to note that the term “Semantic Synthesis” was initially used by Igor' Mel'cuk in 1965 (Zolkovskij & Mel'cuk, 1965), for his Meaning Text Theory which was the suggestion of linking words of different languages together, based on their meaning (i.e. semantically). In our system the objects will be semantically connected to compose (synthesise) specific queries or additions. The associations between the objects will be assigned roles, based on a list of supported role objects for each relation.

In addition to the connection roles, extra information may be provided for each selected item, in a menu offering a choice of specific properties available for each object. As soon as the query or addition has been synthesised, there will be an option to either add it or search for it. Figure 8 (Section 5.1) shows possible search results next to the synthesised query.

The positive factor of showing results this way is that they can be selected and then further explored with additional queries. Specifically designed mechanisms will help the user quickly sift through the results. Some are currently being evaluated based on their performance for a variety of different queries. The whole search experience must be fast and always produce relevant results, which means that the total time spent and the quality of the results are the most important factors.

To create these mechanisms, information must be viewed not only as words and phrases, but as objects which are all parts of the same unified system. That is where the 3D Web interface will come in handy, to give users a look and feel of actually connecting objects.

Ultimately, the user may search through the information provided by our system and then explore links to external websites for extra references, such as images, videos and additional, not yet stored, material.

### 3.2.3 Social Semantics Drawbacks

Web citizens have been tagging online data sources for many years now. Apart from the obvious powerful virtual workforce of collective intelligence they comprise (Halpin, Robu & Shepherd, 2007; Robu, Halpin & Shepherd, 2009), there are also certain negative aspects to it. The current drawbacks of Social Semantics are summed up by three basic factors of human nature: responsibility, credibility and objectivity.

When people publish data on the Internet, they do not always take the time to label and organise them based on the existing standards for each category, thus making their meaning vague for a computer. This lack of responsibility can be avoided by utilising predefined methods for adding and manipulating objects so that they carry at least some basic annotation, which will be aided by using visual objects and Semantic Synthesis.

Even if this problem is surpassed, no one can guarantee that the categories and relations the user has selected are appropriate for that object, because not everyone is a field expert on everything they post online. That is why users will have the ability of voting additions up or down, which will affect the overall reputation of the uploader appropriately.

Finally, the biggest problem of all is objectivity. Even for an organised consortium of scientists and field experts it would be difficult to agree on a common methodology for uniquely characterising every possible piece of information (Aitchison, Gilchrist & Bawden, 2000, p. 173). The only way to prevent the debates from reaching a total deadlock is to try and present all opinions in a very specific and comparable way, so that different researchers can then determine what applies to their specific case.

Philosophy and Sociology will play an important role in the definition of most data categories (e.g. ideas and concepts) which are unsubstantial by nature. Gradually, all views and suggestions will be synthesised and depicted, allowing users and researchers to compare them side by side in a more streamlined manner.

## *3.3 The Bigger Picture: Metakosmos*

All of the above will be joined together in a system called “Metakosmos”, from the Hellenic words “meta”, which means “after”, and “kosmos”, which means “world” or “universe”. The name reflects the ultimate purpose of the system, which is to create a graphical virtual world of organised information about the entire real world.

### 3.3.1 Social Knowledge Management

Metakosmos will be a 3D Social Web community that will store its information by taking advantage of Semantic Web ontologies. The system will have two types of users: unregistered (visitors) and registered (members). Visitors will be only allowed to use the search engine, without being able to adjust the user interface or participate to the addition, categorisation or correction of data. Members will be able to use the website to its full extend.

Additionally, the members will be able to edit their user interfaces on the website by choosing which items should appear at their home screen, also known as their “lounge” (as mentioned in Section 3.2.2). They will select those items by searching through the available categories offered by the website or by adding new ones. They will also be able to add new items to the system, rearrange existing data or even correct mistakes that may exist such as, for example, object miscategorisation or misspelling.

Collecting consistent data is essential for any search engine, especially for one that needs to store them inside an ontology. Members will have to correctly synthesise the data, based on their understanding, in order to add them to the system in the appropriate categories. Other members with better understanding or expertise on a domain will be able to suggest corrections, if needed.

Members that offer a lot to the website, either by adding new material or by correcting existing items, will receive an increased reputation status which will allow them to influence the system faster. Clearly, members that do not cooperate respectfully will receive a decrease of their reputation status, giving them less and less access in influencing the system.

All the user interaction will bring out the social aspect of the system, uniting the members under the common cause of keeping everything organised and as accurate as possible. The additions or changes that are well defined and correctly categorised will be voted up by the members, thus increasing the overall reputation status of the person who synthesised them. In contrast to above, constant mistakes and sloppiness will cause members to vote items down, decreasing the reputation of the person responsible.

Five reputation levels have been created to be used by the system: Ignored (0%), Trainee (25%), Notable (50%), Popular (75%) and Perfect (100%). Newly registered users will start as Trainees and then, depending on how well they adjust, their reputation will increase or drop. If their additions or changes are constantly voted down they will eventually reach the Ignored status, which would mean that their new additions will be automatically pre-voted down and they will also be unable to make any changes to the system. More information on this mechanism can be found in Section 4.3.

The Metakosmos system will aim to provide simplicity to all users, whether they are technology amateurs or experts, as well as offer scalable complexity based on customisable options, while maintaining the same effective search experience despite of the settings.

### 3.3.2 Synthesising, Integrating and Sharing

By reintroducing the term of “Semantic Synthesis”, users will be called to select predefined objects, connect them in a way they consider meaningful, add literal parameter information and finally search for matching results stored in the system. Additionally, registered members will be able to synthesise additions or any needed changes to the system. All of the syntheses will be Semantic Web based, even though the website users will only see the 3D visualisation representation.

Metakosmos will aim to set certain standards that other developers will be able to follow so as to interact externally with the system or to implement additional capabilities for the user interface. Currently set as future work, is creating the mechanisms to integrate Semantic Syntheses from other websites, as well as to share local syntheses with external Web applications by exporting them in an appropriate form.

As time passes, more and more websites will appreciate the advantages of using Semantic Ontologies, making data sharing and integration simpler and instant. Semantic annotation will gradually replace the keyword cloud tagging of Folksonomies, allowing increased access to data of all types and formats (Google, 2012).

### 3.3.3 Adding a 3D User Interface

In order to make Semantic annotation easier and more straightforward, 3D visualisation elements will be used to depict the various Semantic objects that comprise each synthesis. Usability will be an important factor of this new age 3D community, in order for it to get accepted by the general public and integrated into everyday life. It has to combine simplicity with scalable complexity, based on custom options, while maintaining the same effective search experience for all users.

To achieve this, an extra layer of graphics is added on top of text searching, making search queries visually accurate and, as a result, retrieving more relevant search results. With the help of the underlying Semantic Web ontology, as well as with the introduction of 3D graphics, users will be able to query content both textually and visually, depending on their personal preferences and the nature of their search. This way, they will be able to retrieve highly relevant results, which can be presented in many different forms.

With the upcoming official arrival of HTML5 ('HTML5,' 2012) and the constantly increasing improvements of Web 3D graphics such as X3D, X3DOM, WebGL and CSS3, the Metakosmos system will obtain its final form and gain its place among the first 3D Social Web communities, while at the same time participating in the storing and categorisation of global knowledge on a large scale.

## 3.4 Summary

In this chapter, the motivation behind this PhD thesis was explained and justified and the need for a new, organised, 3D search engine has been identified, along with the first system to use it.

# Chapter 4: Analysis & Definition

## *4.1 Prologue*

William Gibson, in his short story from 1982 entitled "Burning Chrome", described a globally interconnected system of computers that graphically presented data to users. He called this system “Cyberspace”, from the Hellenic word “kybernetes”, meaning leader/ruler, influenced by Norbert Wiener’s work on “Cybernetics”. In his 1984 novel “Neuromancer”, Gibson popularised the term further by having his protagonists interact with the system more. A quotation from the book follows:

> "*Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts . . .* ***A graphic representation of data abstracted from the banks of every computer in the human system.*** *Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding . . .*" (Gibson, 1984).

Since then, the term “Cyberspace” has been loosely related to the Internet itself, because Gibson successfully managed to identify and predict the massive societal changes that interconnected computers would eventually bring to the world. The Internet, and its killer app – “the Web”, have indeed changed society in many ways, but there still is a lot more potential to be taken advantage of.

In 2000, Robert Putnam, a political scientist at Harvard University, wrote in his book "Bowling Alone": "*We must not assume that the future of the Internet will be determined by some mindless, external technological imperative. The most important question is not what the Internet will do to us, but* ***what we will do with it****.*" (Putnam, 2000).

This thesis aims to describe a universal system, based on Web technologies, which will use 3D Web graphics to allow its users to create queries and retrieve results in a visual, interactive, manner. Furthermore, this thesis will produce proof of concept that the internal searching mechanism can indeed provide a customisable, user friendly, 3D interface that simplifies searching for information and retrieves very specific, well-filtered and relevant, results.

## *4.2 The Universal System*

As mentioned in Section 3.3, this thesis defines a new, universal, system called "Metakosmos", aiming to take advantage of the new ideas suggested here. The system is Web based, uses 3D Web graphics and the Semantic Synthesis mechanism, and is "powered" by the Social Web, since the users will be the main contributors and sorters of the available knowledge.

Unregistered users will still be able to search for information in the system but they will not be able to add any new information or take advantage of the full range of interface customisations.

Registered users will be able to add new information, by using the aforementioned Semantic Synthesis technique (Section 3.3.2), based on their perception of reality and the things that comprise it.

Each user, however, will have different "effect" on the system, based on his or her, reputation based, access level.

## *4.3 The Reputation Levels*

Metakosmos will aim to take advantage of the Social Web principles of cooperation and peer-review to infuse and inspire the culture of expressing personal views and opinions using clear concepts and definitions. In order to maintain order in the system, registered users will have a specific reputation status level, which will increase or decrease depending on their interaction with the system and its users. In total, five different reputation levels will be used, ranging from 0% to 100%.

In specific, the five suggested levels are:

| Title | From (inclusive) | To (non inclusive) |
|---|---|---|
| `Perfect` | 100% | - |
| `Popular` | 75% | 100% |
| `Notable` | 50% | 75% |
| `Trainee` | 25% | 50% |
| `Ignored` | 0% | 25% |

**Figure 2: The five reputation level value ranges**

The reputation levels, which also correspond to specific colour ranges, will provide a visual representation of a user's progress and attempt to make the entire experience feel beneficial to users who actually help the system to improve and evolve.

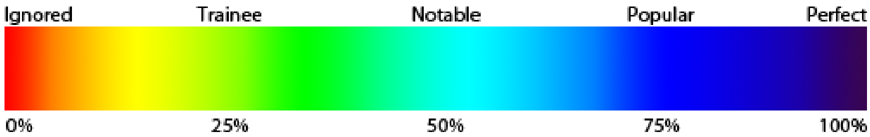



**Figure 3: The five reputation level colour ranges**

The titles have been selected to be equal in size, for reasons of symmetry, and to provide a semantically appropriate definition for each level. More specifically: “Ignored” means that the user’s additions to the system are, generally, not visible to other users. “Trainee” means that the user is still learning how to use the system, which is the initial level that all new users start at. “Notable” means that the user is no longer in training and has already offered useful additions to the system which are backed up by other users. “Popular” means that the user’s additions are becoming quite popular and well-known and many other users have reused them and backed them up. Finally, “Perfect” means that the user has reached the maximum possible level with an increasingly large amount of additions being widely accepted and reused.

These titles have been combined with colours in order to provide a more “graphical” representation. Since the system is meant to be used by anyone around the globe, a user study aiming to evaluate which colours all people like best would most likely be inconclusive, because every human has their own personal likings and the sample would never manage to be large enough to depict the preferences of the entire world. However, it would not be ideal to allow users to personalise the colours of the levels either, since a common standard should exist for referencing purposes.

This is the reason why the full spectrum of visible colours was selected to be used in combination with the reputation titles, by taking advantage of its natural frequency characteristic, in order to make the choice seem globally fair. Red, having the lowest frequency (as low as 400 THz), was used as the bottom of the reputation levels, whereas Purple (≈violet), having the highest frequency (as high as 790 THz), was used as the top (AK-Encyclopaedia, 2012).

This colouring decision is also further supported by the findings of Stone and English in their 1998 paper regarding the connection between colours and mood/satisfaction/ performance (Stone & English, 1998). The two researchers found that the colour red usually agitates people (unconsciously), making them feel uncomfortable, whereas blue calms people down, making them feel secure.

The "Ignored" (Red) status should feel unpleasant and unwelcoming, whereas the "Popular" (Blue) status should feel positive and desirable, thus subconsciously promoting constructive behaviour. The final "Perfect" (Purple) status exists to provide an ultimate "goal" that users can aim for, which would feel like "winning" the game. It should be noted that reputation levels are not forever, i.e. users who reach the maximum level may be downgraded back to Popular, Trainee or even Ignored, if their latest additions are not of high quality, thus keeping all members constantly vigilant.

This approach of combining words (the titles) with colours (the spectrum bar) aims to attract users of different ages, cultures and mindsets, making, in a way, the use of the system feel like a "game" that rewards you an appropriate title for your efforts. Even though many gaming websites and Web forums already use titles and colours to distinguish their members, this combination of titles, colours and scores is unique and harmonises with the globally constructive "culture" that Metakosmos supports.

## *4.4 The Universal Ontology*

Parts of this PhD thesis have already been published in 2011, including the aim of creating an unlimited Semantic ontology (Gkoutzis & Geroimenko, 2011) with the help of the Social Web and a 3D Web based interface to simplify the procedure (see Appendix B for the full publication).

It could be argued that mathematician Georg Cantor attempted to prove with his “naive set theory”, in 1891, that you cannot have a single, universal, set of everything, including itself (Cantor, 1891). But, as the philosopher and mathematician Bertrand Russell proved ten years later in 1901, Cantor’s argument contains a paradox – which is now called Russell’s Paradox or Antinomy – thus making Cantor’s naive set theory invalid (Russell, 1903; Russell, 1996).

This antinomy can be easily understood using “The Barber Paradox” (Quine, 1976, p. 12). In a town there is only one barber, and he is male. All the men of this town can **either** shave themselves **or** go to the barber and have him shave them. So, basically, a barber is the man who **only** shaves men who do not shave themselves. But then the question arises: **Who** shaves the barber? If he shaves himself then that would be against the rule of non-inclusion we just set; thus the paradox.

To put this in mathematical terms, let A be the set of all sets that are not also members of themselves. If A can be proven to also be a member of itself, then it would contradict its own definition. However, if A is not a member of itself then it would now qualify to be a member of itself, based on its definition, thus causing a contradiction once more. This paradox can be symbolised using the formula below.

$$let\ A = \{\, x \mid x \notin x \,\}$$

$$then\ A \in A \Leftrightarrow A \notin A$$

**Figure 4: Russell's Paradox**

Based on the above, Metakosmos can indeed define all the different possible opinions and perceptions of the world and allow its users to interconnect them as they see fit, without limitations.

## *4.5 The Specifics of the Search Mechanism*

It has been suggested by many researchers (Barrett, 1996, p. 52; Mizzaro, 1998; Saracevic, 1996) that the problem with the Internet is not so much in **finding** information, but in **sifting** it appropriately in order to locate the exact item that is actually required, among an unimaginably huge amount of other material. Based on the above, information can be divided into three categories: a) the relevant, b) the irrelevant and c) the arguably relevant. In this way, any search tool can measure its success by how quickly it can differentiate results belonging in these three sets and how small it can make the third set.

The Semantic Web project, as mentioned in Section 2.1.3, uses ontologies to strictly define and categorise every piece of stored information. Using an extra layer of graphical abstraction, where the users select **visual objects** masking the underlying ontology items, Metakosmos allows users to carry out Semantic searches without any knowledge of XML, RDF, OWL, or the querying language SPARQL.

Using visual objects as an abstract version of actual items is not a new idea in itself. Howard Rheingold mentions a story (Rheingold, 1994, p. 65) about Douglas

Engelbart who, in 1950, came up with the idea of using “symbols” in computers to help us think and organise information, in order to successfully cooperate and boost our collective intelligence. Engelbart later expressed this idea more analytically in a report he prepared for the Stanford Research Institute (Engelbart, 1962).

In pages 21 to 23 of this report, Engelbart discusses concept manipulation, stating that humans rose above lower life forms by developing abstractions and concepts and by differentiating between general concepts and specific instances, which consequently allowed them to predict new instances and associate other concepts. Additionally, by learning to represent particular concepts in their minds with specific symbols, humans added direct value to heavy thinking by mentally manipulating symbols instead of the more “unwieldy” concepts they represent. Finally, the author explains the importance of externalising the symbol-manipulation activity, particularly in graphical representation which supplements the individual’s memory and ability to visualise.

The search engine used by the Metakosmos system will be based on visual “symbols” which correspond to ontology entries. We call these symbols “**concepts**”. Each concept uniquely defines one, and only one, thing. This thing can be anything: material or immaterial. Concepts can also be based on other concepts if their definition includes in its entirety the definition of others.

Additionally, when users want to reference a specific instantiation of a concept, they create an “**object**” based on the concept they are instantiating. For example: while “Neuromancer” is a specific concept (A) which also includes the entire definition of the Book concept (B), it would not be informative enough to just claim a user owns A in general. Users would have to create a new object based on A, which they would then reference as something they own in their personal library, and they might perhaps add extra information that applies specifically to the copy they own, not A in general.

Both concepts and objects are visual items that appear in the graphical representation of the engine. Concepts can be independent, if no one has based them on another concept yet, but objects must always be based on a specific concept.

It should be noted that the visual concept representation created for and used by the system is similar to Joseph D. Novak's "Concept Maps", which he invented together with his research team at Cornell University in 1972 (Novak & Cañas, 2008). Additionally, the concept-object instantiation has been loosely based on the fundamental teachings of Object Oriented Programming (Budd, 1991).

In the end, after the visual items have been used in a search, the system can ask the user whether the results were accurate or not, in an attempt to determine the usefulness of the syntheses involved. The answers to this question will have a dual role: firstly, they can affect the ranking of the users who composed the returned syntheses (as mentioned in Section 4.3) and secondly, they can readjust the suggested items for each user in specific, thus learning from the user and personalising his or her overall system experience.

## 4.6 Project Sustainability

### 4.6.1 Economic Factors

For every project that aims to be launched officially someday, one must consider the overall economic sustainability of the venture.

According to an analysis regarding business models, specifically for businesses on the Web, the basic categories include: “Brokerage, Advertising, Infomediary, Merchant, Manufacturer, Affiliate, Community, Subscription, Utility” (Rappa, 2000). Without getting into too much detail about all the categories, based on this classification, Metakosmos can take immediate advantage of the Manufacturer, Affiliate, Community and Subscription models, and also of the Advertising and Infomediary under certain circumstances.

The Manufacturer (Direct) model refers to the ability of a manufacturer of a service to directly reach its audience. Metakosmos can allow external websites and services to query its system for a monthly, yearly or even hit-based fee, as Amazon and Google already do.

The Affiliate model defines the method of assisting other websites and services to promote their products via other websites. By allowing companies external to Metakosmos to synthesise their products into objects for users to search and buy, the system becomes the intermediary thus receiving an appropriate fee.

The Community model discusses asking the user base to assist with a specific goal, such as raising enough money to upgrade the infrastructure on a donation basis. Wikipedia has been using this method for many years with great success

(Wikimedia, 2012), but it should be noted that they have already built an extremely large and loyal user base willing to support them.

Finally, the Subscription model defines the process of charging users for access to specific, additional services, on a monthly, yearly or per service basis. Given the fact that the universal ontology of Metakosmos will gradually gather a lot of accumulated knowledge, it becomes obvious that many would pay to gain access to its more advanced synthesis tools.

In addition to the above models, the Advertising and Infomediary models should also be considered. If the system requires more income to survive, users could be asked to participate into optional surveys and their answers could be used for targeting specific advertisements to them, thus combining the two methods. This, though, should be a last-case scenario, to be used only if all the previous models do not manage to produce the required net income needed for the system to survive.

Regardless of what method the system uses to preserve its existence, the basic rule to be followed is: “respect the user, for the user to respect you”. A system without any worthy users left, will not be worth much to anyone.

### 4.6.2 Legal Factors

Neil Barrett, back in 1996, wrote that the Internet could have one of two potential futures: blossom, or death – or at least paralysis (Barrett, 1996, p. 220). This may sound like a generic argument but he continued to explain the causes that could make this a reality. He proposed three basic reasons why the Internet could be forced to withdraw: a) insufficient investment in infrastructure, b) excessive/stifling legislation, or c) legal challenges and prohibitions (Barrett, 1996, p. 221). He also noted though that, since most government funding has now been withdrawn, the Internet has become the domain of businesses that publish material or provide connectivity and communication services.

In addition to the above statement, Ben Shneiderman argues in “Leonardo’s Laptop” (Shneiderman, 2002) that the complexity of information communication technologies stems from the high degree of interactivity necessary for exploring available data, but the Internet addresses this issue by allowing for simplified interpersonal communications and decentralised initiatives. He continues to say that certain critics worry about the creation of an information-poor minority or even an “Internet apartheid”, due to the increased expansion of bureaucracy and legal limitations imposed by governments.

It is becoming obvious that the Internet and, by extension, the Web, are facing the serious and immediate danger of being destroyed by the same laws that are supposedly created to protect them.

If this system is to survive, it must be given the freedom to allow its users to express their opinions by synthesising their perspectives and combining them with each other. To achieve this, an international legal team should be employed to carry out all the necessary tasks of handling any freedom of speech issues that may occasionally arise, as well as to monitor and criticise any questionable laws such as, for example, Wikipedia did in its recent actions against the dangerous “Stop Online Piracy Act” (Wikipedia, 2011).

The legal factors will be further discussed in the future work section (7.3).

## *4.7 A Summary of New Ideas*

As mentioned in Section 1.3, this thesis has introduced a few new ideas to the Computer Science field. A summary follows.

### 4.7.1 Merging of three different fields

The Metakosmos system takes advantage of the capabilities of the Semantic Web, Search Engines and (X)3D Graphics, and attempts to add an extra layer of abstraction to conventional plain-text searching. There has never before been, until now, a Semantic Web Search Engine with 3D graphics; especially one that uses the XML compatible X3D standard, which is also browser friendly since it will probably soon be integrated into the next official version of HTML.

A graphical depiction of this combination can be found in Figure 5 below.

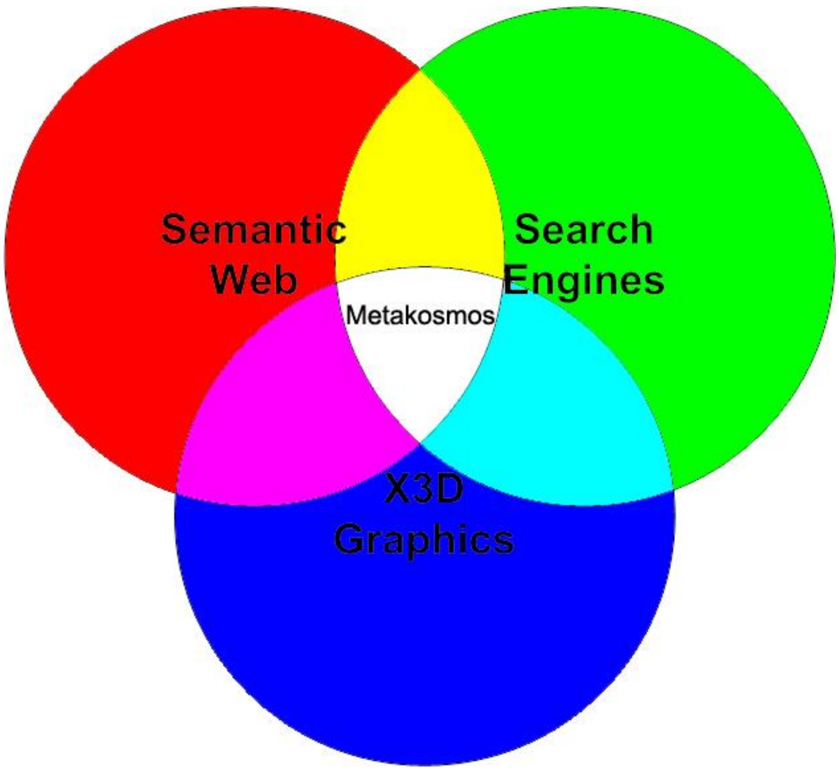



**Figure 5: Where Metakosmos stands**

### 4.7.2 Concepts, Objects and Semantic Synthesis

As mentioned in Section 4.5, this thesis combines Joseph D. Novak's "Concept Maps" (Novak & Cañas, 2008) with the fundamental teachings of Object Oriented Programming (Budd, 1991), to create the new notion of handling visual items representing concepts and objects, that are stored in the system based on the perception of the users who added them.

Concepts uniquely define only one thing and they can be based on other concepts if their definition includes the entire definition of others. Objects are the instantiation of concepts, when a user needs to reference a specific manifestation of a concept.

There is a close relationship between Object Oriented Programming and the Semantic Web (Knublauch et al., 2006), which has made it extremely easy to transfer this logic into the mechanisms of the system.

To serve the purposes of this system, a new technique has been specified called Semantic Synthesis. This is a 3D based visual method of composing combinations of graphical items that aims to encapsulate the underlying Semantic Web system by adding a layer of abstraction on top of the searching and addition of information.

The items on the screen take advantage of direct manipulation principles (Shneiderman & Plaisant, 2004, pp. 192-241) by offering virtually tangible items that can be handled both my mouse clicks and keyboard shortcuts. From the smallest grapheme to the largest possible discrete idea, the concepts and objects of the system can be easily visualised, searched, found and retrieved.

### 4.7.3 Personalised Search Interfaces

Following from the previous two contributions, in order to add an extra layer of visual abstraction to searching via the Semantic Synthesis mechanism, an appropriate visual interface is required.

Since the system aims to be used by anyone around the world, its interface should to be able to “adapt” to the needs of each user in order to be as convenient as possible, regardless of the educational or IT background of the different users. For this reason, Metakosmos will ask its members if the search results that were returned were appropriate (i.e. semantically correct) for their search query. The answers of the users will affect the popularity of the returned syntheses, while also affecting the ranking of the members who created them, which is discussed in the next section (4.7.4).

Based on the usual searches of each user and the items they usually pick to define first, the system can readjust its suggestions in order to assist its members to compose their syntheses faster. These suggestions will be part of the “Search” button, which is defined in Section 5.1.4.

### 4.7.4 Custom Reputation Levels

Although this is not a new idea per se, since it has been widely used by role playing computer games for a long time already, applying “usefulness”-based reputations to this knowledge-management system promotes a reciprocal culture of mutual development, while at the same time it makes the experience more recreational and interesting.

It is important to note that this mechanism will not be used to punish users for expressing their sincere opinion, whatever that may be, but for intentionally devastating and flooding the system with non-existent trash entries or harmful materials such as viruses and unsolicited spam advertisements.

The reputation mechanism is not being used by the core search engine which is described in the remainder of this thesis, but it is an essential factor for the full system to operate correctly on a universal scale.

## *4.8 Focusing on the Core*

### 4.8.1 General Description

The search engine of Metakosmos will derive its information from a universal Semantic Web ontology and will offer a 3D interface, inside the Web browser, with the help of X3D technologies. The reasoning behind choosing Semantic Web and X3D Graphics for this Search Engine has been described in previous sections (2.1.4, 2.1.6, 3.3.1, 3.3.3).

The system will have two types of users: unregistered users (Visitors) and registered users (Members). Visitors will be allowed to use the search engine without being able to adjust the user interface or participate in adding, categorising or correcting data. Members will be able to use the search engine to its full extent, which includes personalising their user interface as well as being allowed to add and categorise data.

Visitors should be able to register for a new account by entering their personal information (full name, birth date, etc.) including their e-mail address. The password should be sent to their e-mail address and they will then be able to log into the system using their username and password combination.

After logging in, Members will be able to edit the user interface of the website by choosing which items should appear on their main search screen. They will pick those items by searching through the available categories offered by the website, based on their personal preferences and requirements.

Members will also be able to change the personal information that they submitted. If they choose to change their e-mail address, a verification e-mail should be re-sent in order to check the validity of the new address.

Members will also be able to add new information to the system, rearrange existing data or even make suggestions for the correction of mistakes that may exist (including miscategorisation, spelling, etc.). Members that contribute with continuous and correct data entry and item categorisation should occasionally receive an increased in their reputation status which should allow them to influence the system database more instantly/directly. Of course, Members that are not cooperating peacefully should receive a decrease in their reputation status, giving them decreased ability to influence the system.

Members will also be able to logout of the system in order to avoid identity theft or misuse of their account.

All this user interaction will bring out the social aspect of the system, uniting the members under the common cause of adding and categorising everything as strictly as possible.

The system will be able to process the Semantic Syntheses that Members prepare, and successfully import them.

Based on the above, the main system functionalities can be divided into four sections: Synthesising, Storing, Querying and Displaying.

### 4.8.2 Data Retrieval

Collecting consistent data is essential for any search engine, especially one that needs to store them inside an ontology. Automatically retrieving data from Semantic Web ontologies would probably be faster but most information on the Web is currently in plain text Web pages, scrambled within the layout, or hidden in the "Deep Web" of scripts and login-only communities. Additionally, even the available ontologies should first be fully "normalised" before being imported into the system, which means being simplified enough for every item to strictly mean one and only one thing. It therefore has to be performed by a human who understands logic.

Members will be able to manually add information from the plain text Old Web into the system. They will have to correctly synthesise the data, based on their understanding, in order to add them to the appropriate categories. Other Members, with more expertise or understanding of a domain, will later be able to suggest corrections or alterations, if needed.

### 4.8.3 Local Data Storage

The retrieved data will be appropriately reformed in order to be capable of being integrated with the rest, in the universal ontology of the system. This "ultimate" ontology will be stored on hard disk drives local to the system, as further discussed in Chapter 6.

### 4.8.4 Query Handling

The queries will be produced based on the selections of the graphical user interface. Each item will be a parameter for a query which will be executed in the database and the results will be returned to be displayed by the 3D visualisation environment inside the Web browser.

Synthesised compositions will act both as queries and results, thus maintaining a sense of coherence. The returned results will be based on the information currently stored in the system, added by the registered users.

It is important to note that the search results can be considered as “semantically” related both to those of Wikipedia – an online encyclopaedia – and to those of Google – a Web page finder. This happens because the concepts of Metakosmos resemble the entries of a global encyclopaedia or lexicon, whereas the objects of the system are instantiations of concepts that can refer to actual existing things like, for example, webpages.

In this way, when a blogger of an external website decides to write an article about cats, not only he/she can use the unique identifier of the “Cats” concept in his/her article, but he/she can also use a very specific object identifier in order to explicitly refer to his/her Siberian pet cat called “Furball”.

In addition to the above, users can execute parallel searches and then combine their results to identify and highlight any common items that exist in both result-sets. This creates the effect of applying the mathematical union (U) operation of set theory, assisting and enhancing combinatorial thinking.

This makes Metakosmos a unique resource, both as a concept referencing system and as an object catalogue indexer.

### 4.8.5 Visualisation

In a 2004 study regarding information representation in Web directories (Chen, Magoulas & Macredie, 2004), it was noted that “it is important to consider versatility in the design”, in order “to allow for use by a variety of individuals, rather than a particular user group”. In a similar manner, others have argued that “there is no panacea” in software visualisation (J Stasko, 1998, p. 479) and that “success will come to designers who provide compelling content, relevant features” and “novel social-media structure support” (Shneiderman & Plaisant, 2004, p. 223). Users of information and communication technologies have an essential role to play, by defining what they want and need from any system claiming to promote universal usability (Shneiderman, 2002, p. 36).

By offering a customisable 3D user interface where users can select the options they prefer, people from all around the world, with different likes and dislikes, can make the system match their personal style and preferences.

Visitors will use the website with a default interface offering the basic options. Members will be able to have full scale access to all the functionalities and experience the Web application in its entirety, which also serves as an incentive for someone to become a registered member.

Members will be able to select a style theme from the already available categories and should also be able to produce and apply custom layout themes based on their needs and likes. Using their personalised interface, they will be able to faster synthesise queries in order to find what they are looking more swiftly.

The search results will be visually displayed to the user of the system based on the pre-selected settings. Visitors will be presented with a default linear result view, whereas Members will be shown their pre-selected view based on their personal settings. Four different example result views are suggested, by default, by the system: compact, linear, spiral and torus.

The compact mode merges the results based on their commonalities with the original search. It clearly displays the connection between query and results but it requires careful item spacing and size management.

The linear mode presents the results one after the other, as if browsing through a paginated colour sample book. It is the simplest mode and it will be available by default to non-registered users.

The spiral mode (loosely based on the golden ratio, Fibonacci spiral) presents the results in a spiral which can be easily manipulated and explored in a continuous manner while maintaining a sense of serialisation and ordering. This mode can be selected by registered users.

The torus mode (loosely based on the tube torus) presents the results spread around the user in a sphere, with the user’s perspective being from the middle of the circle. This method allows for larger result sets to be more easily displayed and explored. This mode can be selected by the registered users.

“Digital data architects” should be able to create and apply their own result representation themes, based on scientific research or their imagination. The default suggestions of the system are expected to still be available, but the presentation will differ depending on each user’s preferences and likings. Customised user templates will make sure that the system will take into account diversity and inclusion issues. This is a really essential modification to the entire system which has been marked as future work (Section 7.4). Even if not every single person in the world ends up using

Metakosmos, different abilities and needs should be taken into consideration because this system is meant for everyone to use.

It should be borne in mind that the system graphics should not and will not promote a feeling of “hyperreality” (Baudrillard & Evans, 1991), which means that they should not aim to make the users feel as if they are still in the “real” world. Users should be made to feel natural and comfortable enough to use and directly manipulate the controls and adjust the interface to fit their preferences and idiosyncrasy, but this ease-of-use should not replace the real world experience with its digital, virtual, definition. As mentioned in Section 3.3, Metakosmos is a “world” that comes after, and is about, the real world. Without experiencing the real world that has been catalogued inside the system, users would be reading about the existence of concepts and objects as if they were reading a science fiction novel. In order to appreciate the usefulness of the information provided by the system, users should feel the need to research and double-check the existing syntheses, as well as the ones they add themselves.

## *4.9 Summary*

In this chapter, the analysis of the system was presented, which shows how the system should generally operate and how the specific search engine should work. The main challenges have been identified and they have been divided into categories based on their functional similarities.

# Chapter 5: Design & Modelling

## *5.1 Preliminary Interface Designs*

During the design phase, certain preliminary draft designs of the interface were created, presenting the initial 2D interface models before their transformation into X3D (to be presented in the next chapter). The models were produced to depict an initial view of the system in order to visually identify its distinct parts and spatial requirements.

It should be noted that these designs have also been published and presented at the 15th International Conference on Information Visualisation (Gkoutzis & Geroimenko, 2011).

### 5.1.1 Initial User Interface 2D Models

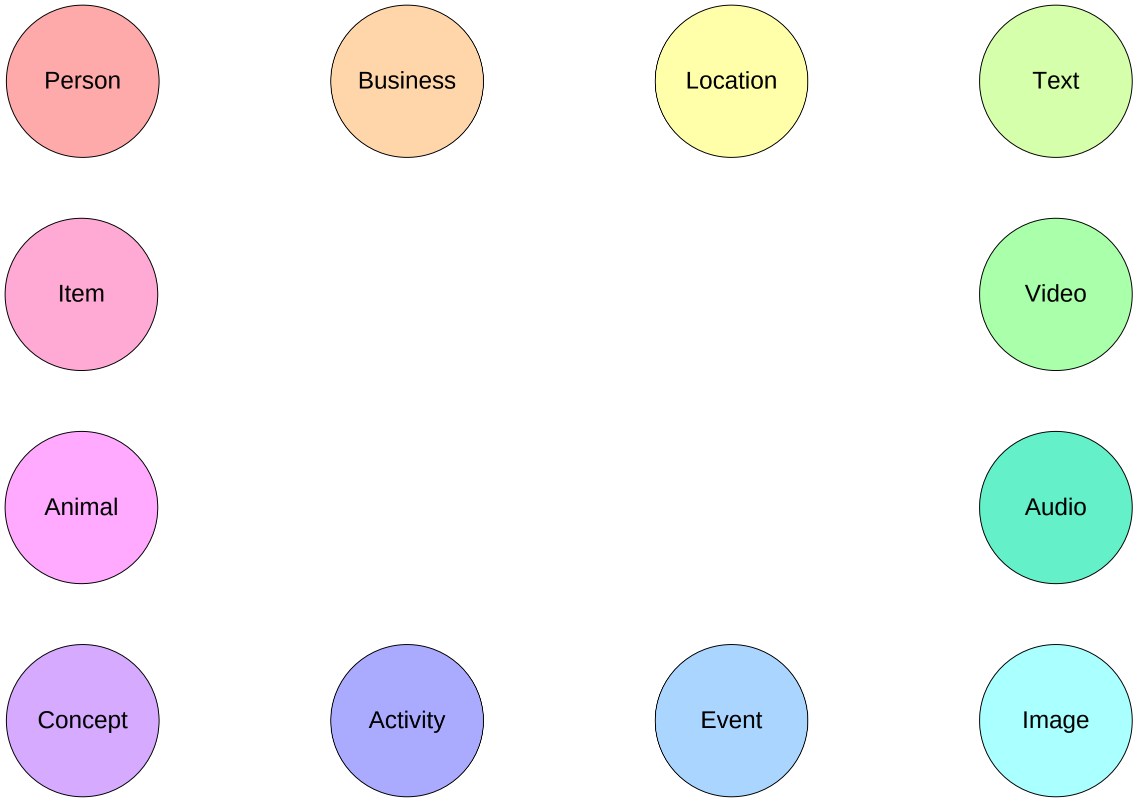



**Figure 6: A custom user interface with twelve items immediately available**

In Figure 6 above, a custom user interface is presented, with twelve visual items immediately available to the user. These items can either be concepts or objects, depending on what the user usually searches for.

If users need an item that is not immediately available on their interface, they can use the search menu (not depicted in these images) to recall the item from the system. If the item they need does not exist in the system yet, they can use the menu to add it.

Users can manipulate the visual items on their interface to synthesise queries that can be executed by the system in order to retrieve matching results. An example of a synthesis follows in Figure 7.

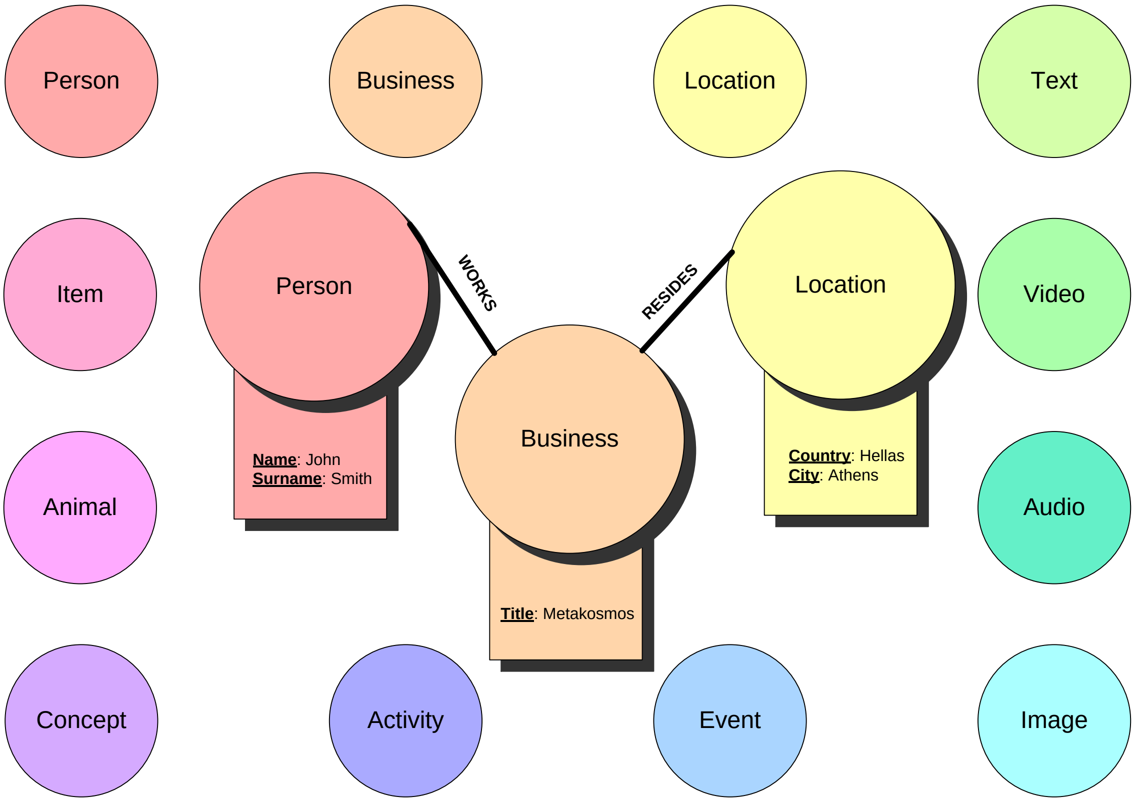



**Figure 7: Draft layout of a composed synthesis in a custom user interface**

The user selected a Person concept, combined it with a Business concept and also with the Location concept. Then, certain values were given to the sub-concepts used as parameters, based on the information that the user already held about the synthesis.

After all the known information has been filled in, the synthesis can be queried by the system to find possible matches containing **at least** the data already entered in the composition by the user. A possible result-set can be found in Figure 8 below.

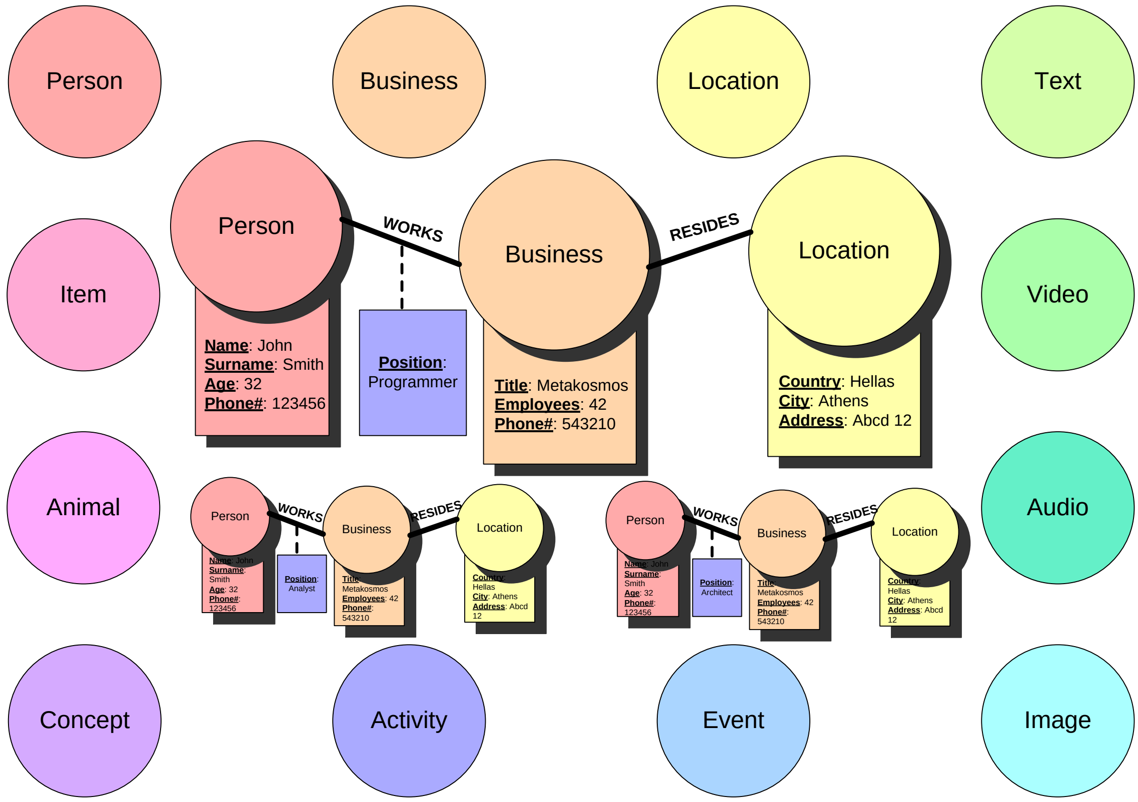



**Figure 8: Searching for the synthesis returns all matching results**

The results that have been returned are highly relevant to the user's search because they contain the very specific concepts and explicit parameter values that were defined in the original synthesis. The user can then sift through the results to find those of specific interest.

It should be noted that if the search comes up with no results – but the user is certain that the composed synthesis actually exists and is valid – then it can be added into the system with the user referenced as the original adder. In this way, gradually, the system will include all possible interpretations of reality based on the users' perception. This is further discussed in Chapter 7.

### 5.1.2 Initial User Interface (X)3D Screen Captures

After the creation of the 2D models above, the first attempts to transform them into X3D began. Using only Notepad and Don Brutzman's book on X3D (Brutzman & Daly, 2007), the first .x3d files were created. In order to view the output, special X3D viewers were used, further mentioned in Chapter 6.

The initial 3D version of the 2D model presented in Figure 6, can be found in Figure 9 below.

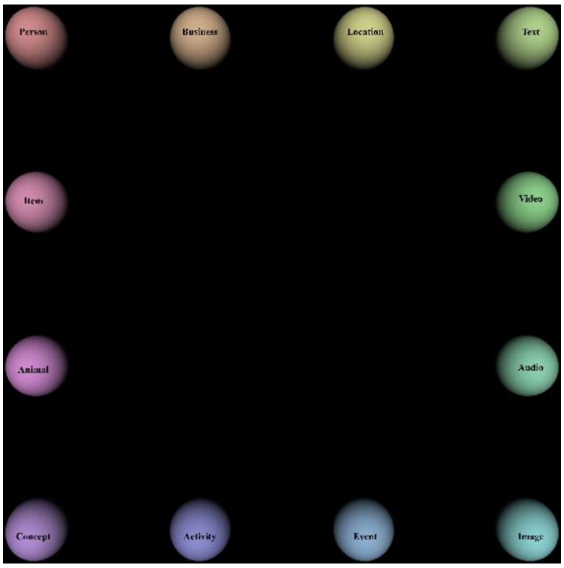



**Figure 9: A custom user interface (in X3D)**

In order to perceive how the synthesis technique spatially fits into the interface, Figure 7 has also been converted to X3D and can be found in Figure 10 below.

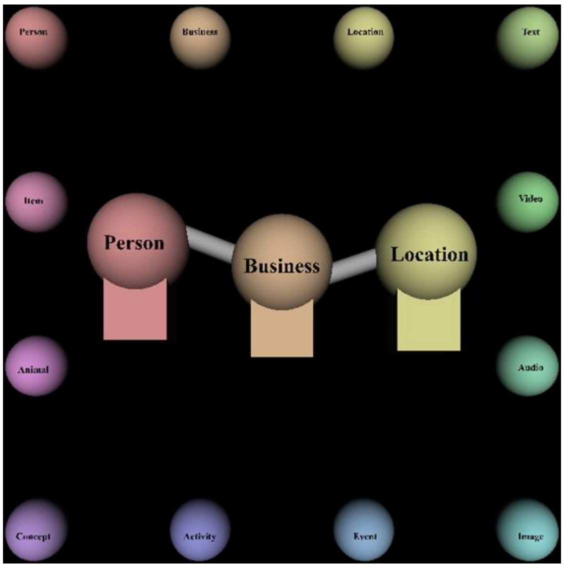



**Figure 10: Draft layout of a composed synthesis (in X3D)**

Finally, Figure 8 containing a mock-up query example with linear result view, has also been translated into X3D and can be found in Figure 11 below.

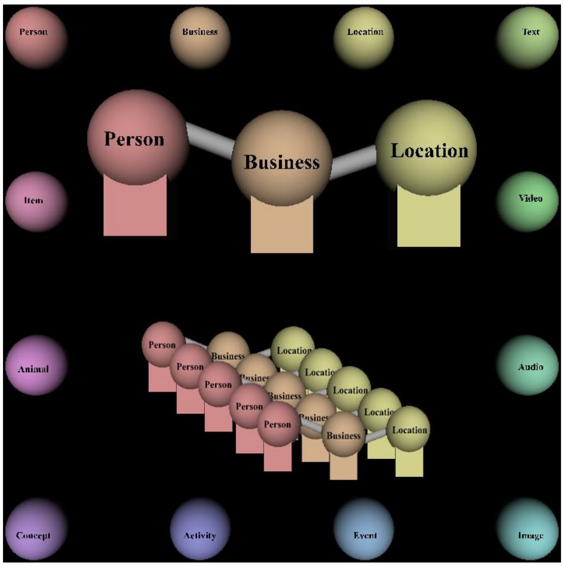



**Figure 11: Searching for the synthesis (in X3D)**

### 5.1.3 Anatomy of a Visual Item

As mentioned before, the system uses graphical representation of concepts and objects to provide users with visual items that they can directly manipulate in order to synthesise queries and additions to the system.

The general layout of the first version of such an item can be found below in Figure 12. This contains the basic elements that should be present on an item in order to provide faster access to item actions. This suggested design is bound to change in future versions, as new system needs arise.

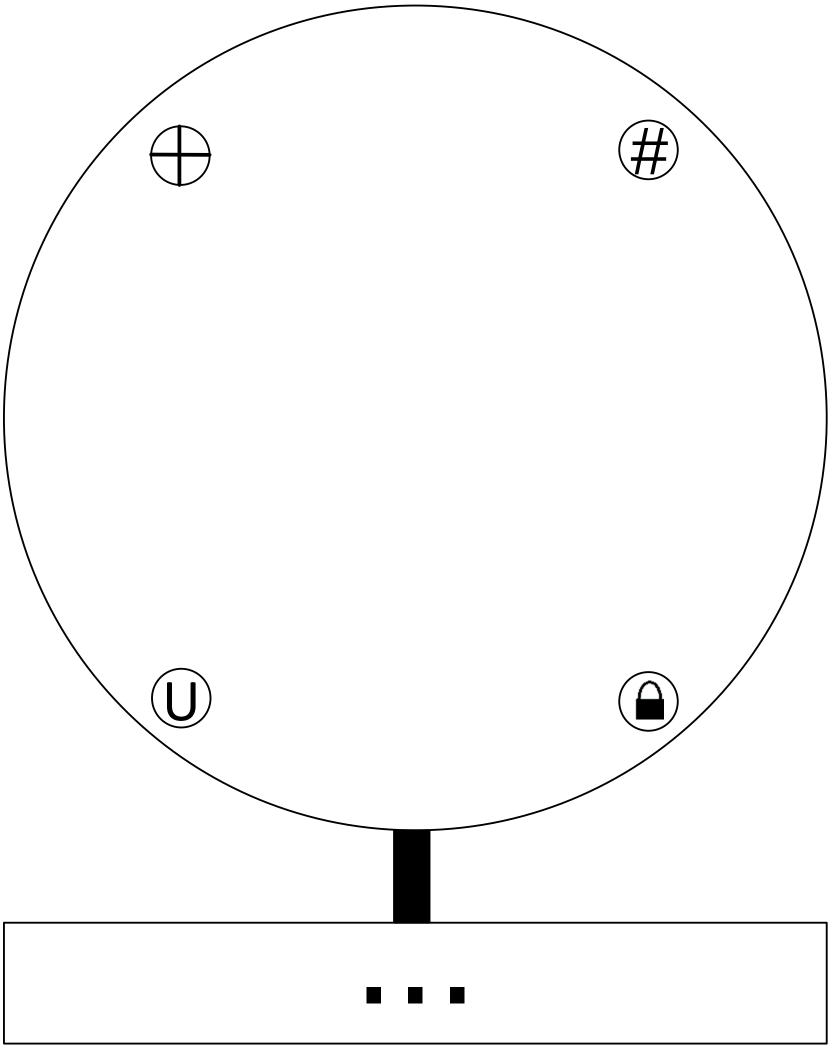



**Figure 12: The schematic layout of a visual item**

In the figure above, certain elements essential to the use of the system are visible.

Top Left: This is the expand/retract option, which allows the user to expand the full contents of the item. The alternative version contains a minus instead of a plus sign. This aims to help users manage the space that their queries and results require in order to be displayed and converts syntheses from/to the “Compact” mode.

Top Right: This is the unique identifier option, which shows a Dereferenceable URI that uniquely identifies this specific item or combination of items. This URI can then be copied to memory or bookmarked for future reference. This idea has been borrowed from the Linked Data movement (Watson, 2010).

Bottom Left: This is the “find common (=mutual) items” option, which adds the URI and its items into the comparison container. The U symbol stands for “union”, a term borrowed from set theory, as mentioned in Section 4.8.4. This can be useful when combining two different searches into one.

Bottom Right: This is the lock option, which locks the specific item to show whenever it appears in syntheses instead of, for example, ending up in the consolidator bar or getting buried deep within the result-set. This option will “teach” the system about the specific preferences of each user, allowing each user account to offer different item suggestions depending on its personalised settings.

Bottom: This is the consolidator bar, the place where extra items related to the main item are hidden, in order to avoid cluttering the interface with too many graphics. This menu expands and retracts vertically below the main item and contains all the extra interrelated items that have been set with specific values about each item.

In the middle, the label of the visual item is displayed. This label readjusts its angle depending on the camera viewing angle, in order to offer legibility regardless of the rotation. This provides a “2D feel” to the overall design, while at the same time maintaining the advantages of a 3D environment, like, for example, the extra space for additional items on the z axis.

In the next chapter, these elements have been converted to X3D and their functionality is demonstrated.

### 5.1.4 Anatomy of the Companion Menu

In order for the users to use the system more efficiently, a companion menu has been added to assist them with common tasks. A draft model of the menu can be seen in Figure 13 below.

| Information bar | | |
|---|---|---|
| search | move up | add new |
| move left | move down | move right |
| home | reset view | settings |

**Figure 13: The schematic layout of the information bar**

The information bar shows useful messages from the system to the user like, for example, a read-only plain text description of their query or the number of results found.

The “search” button either queries the composed synthesis, or searches for an item in the system in order to extract it onto the interface. Suggested items appear first, as defined in Section 4.5.

The “add new” button adds a composed synthesis into the system.

The “home” button returns a user to his/her lounge (=home).

The “reset view” button resets the camera view back to the default.

The “settings” button opens the settings menu which contains the system options that the user can effect, such as for example: set lounge wallpaper, set button colour, set default result mode (compact/linear/spiral/torus), show/hide search result axes, set keyboard shortcuts, define visible result depth, etc.

It should be noted that the companion menu follows the user – hence the name – and rotates depending on the camera angle (like the visual items in Section 5.1.3) in order to maintain legibility.

## *5.2 Evaluating the Semantic Synthesis Mechanism*

One of the scientific contributions of this PhD thesis is the creation of the “**Semantic Synthesis**” mechanism for visually composing search queries. In the previous chapters, Semantic Synthesis was defined (3.2, 3.3) and explained (4.2, 4.4, 4.5, 4.7, 4.8), while Figure 6 (5.1.1) displayed a graphical example of how it works.

The mechanism of Semantic Synthesis plays an important role in the development of the system described in this dissertation and it was therefore decided that it should be evaluated via a small-scale user study. The procedure to be followed involved: deciding on the form and the questions of the study, successfully gaining the approval of the Research Ethics Committee, and finally carrying out the study itself. The steps taken are described in the next sections of this chapter.

### 5.2.1 Creating and Planning the User Study

The creation of a user study for a proposed mechanism that has not been implemented before is a very challenging process. Even though Semantic Synthesis has been based on well-known and well-established IT paradigms like Concept Maps and the foundations of Object-Oriented Programming (4.5), the novel combination of these ideas could be initially difficult for a new user to grasp. With this in mind, the user study was planned and created to also include a tutorial of the system, as if the participants were actually using Metakosmos, even though the procedure was carried out on paper.

The “Practical Guide to Ethical Research Involving Humans” (Vinson & Singer, 2008) provided a set of extremely useful guidelines on how to design this user study. The evaluation begins with anonymous demographical questions, in order to acquire a general understanding of the prospective user’s background. After this step is completed, the users “enter” the system and are prompted to choose whether they would like to view the tutorial guide or not. If a user chooses the option no, he/she will be directly taken from slide 8 to slide 20. If a user chooses the option yes, he/she will view a brief guide of how the system works. If the users who chose no change their mind afterwards, they can ask the researcher to go back to the guide again.

This introductory guide aims to fill the knowledge gap that exists because of the uniqueness of the visual searching mechanism. Towards the end of the tutorial, the participants are asked to answer a quiz about Semantic Synthesis based on what they just learned, in order to check if the learning process was successful or not and to determine whether the guide achieved its purpose.

After the participants complete the tutorial, they “enter” their personalised interface in the Metakosmos system. First, they are asked to answer two untimed questions and their answers are checked and corrected by the researcher. The researcher will ensure that the users are stress free and aware of the fact there are no right or wrong answers because the mechanism being studied is a new method of searching so they are not supposed to know how it works beforehand.

The rest of the questions are timed and after each question is answered by the candidates, a suggested answer is then presented. At this point, the researcher informs the participants whether their answer was “Correct”, “Semantically Correct”, or “Wrong”. The term Semantically Correct means that the answer matches the logic of the mechanism, even if the wording is different, because the system can easily deduce that certain items are interchangeable or identical, based on the interrelations of the underlying ontologies. Additionally, in the case of any answers where the users skip the step of drawing certain labels in order to save time, the answers will count as a Semantically Correct since the actual system would have provided these labels in a list for the users to choose from.

If the users fail to produce a Correct or Semantically Correct synthesis, they will be informed of their mistake (or mistakes), so that they realise what went wrong, and then the researcher will pose the next question. It is important for the researcher to maintain a neutral attitude towards the participants, as much as possible, to give the candidates the feel of a “talking computer guide” that could be carrying out this evaluation on the Metakosmos website.

After the timed questions are answered, the user will be “logged out” of the system and asked to answer the final questions regarding his/her personal liking of Metakosmos and the Semantic Synthesis method. Finally, the researcher will gather any additional answer sheets and keep them in a safe place until they are processed.

Please note that the final version of the aforementioned User Study can be found in Appendix E.

### 5.2.2 Ethics Procedures

After any user study has been created, it must be approved by the university’s Research Ethics Committee before the researcher can proceed with the interviews. For this purpose, all the necessary application forms were completed, based on the formal scientific guidelines (PU, 2012b; PU, 2012a), and were then submitted to Plymouth University for evaluation so that the next step could be pursued.

The Ethics Committee was informed that this user study will evaluate a mechanism designed for the purposes of a PhD thesis entitled “A Semantic Web Based Search Engine with X3D Visualisation of Queries and Results” and that the thesis aims to define a new type of search engine which uses Semantic Web Ontologies to store its data, in order for the information to be interlinked based on meaning (=semantics) rather than just lexicographical value. Additionally, it was mentioned that this search engine will also provide a Web browser-based X3D user interface to its users so that they are able to visually handle and manage the information stored in the system.

This brief introduction to the system paves the way for the description of the topic of the study itself. It was explained that the mechanism created for and utilised by this visual searching system is called “Semantic Synthesis” and because it constitutes the “heart” of the overall design, it has been decided that it should be evaluated via a small-scale user study.

For the purposes of this study, 5 to 15 “prospective users” of the system are to be selected as the sample population, in order to briefly examine and evaluate the potentials of the mechanism in practice. As mentioned above, these participants are to be asked a series of questions and their answers will be recorded in answer sheets (if needed) and then analysed.

The Ethics Committee approved the application for the user study and requested a minor change in the last slide (#42). The original phrase: “Please remember that you may always opt out at any given point before your data have been processed and aggregated with the rest” was revised to “Please remember that you may always opt out at any given point” on the committee’s request. After gaining the approval, the next step was recruiting and interviewing participants.

### 5.2.3 Interview Sessions

The interview sessions involved private, one-to-one, meetings with each participant, at a location and time selected as per the convenience of both the participant and researcher. Silence was maintained in the surrounding environment as much as possible to help participants focus on absorbing the new information from the tutorial guide. The user study was carried out over a period of four (4) days with a total of nine (9) participants.

All nine individuals participated freely in the study and no one opted out afterwards, even though they were informed that they could. Each interview lasted, on average, approximately thirty-two (32) minutes. All the answers of the participants have been illustrated in a table in Appendix E.

The final step was that of analysing the data acquired from the user study in order to evaluate the Semantic Synthesis mechanism, and deriving conclusions regarding its usability and possible improvements.

### 5.2.4 Results Analysis

Because the study was anonymous, no private information was recorded. The codenames “Participant #1” – “Participant #9” were used instead of real names in order to distinguish between the participants. Both men (44%) and women (56%) participated in the study, as depicted in the figure below.

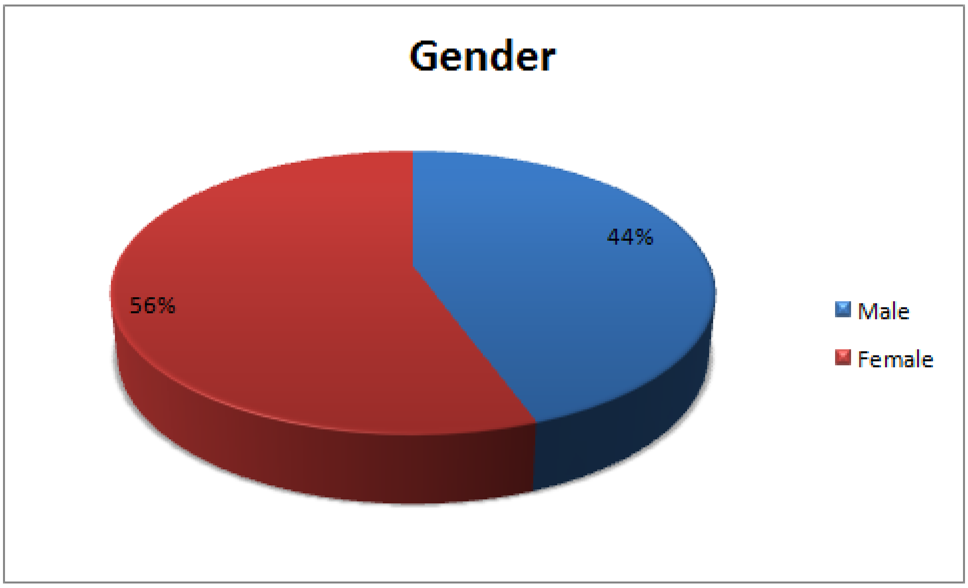



**Figure 14: The genders of the participants**

The participants belonged to different age groups: 56% were aged between 18 and 24 years, 11% between 25 and 30 years, 22% between 31 and 40 years, and 11% between 41 and 50 years.

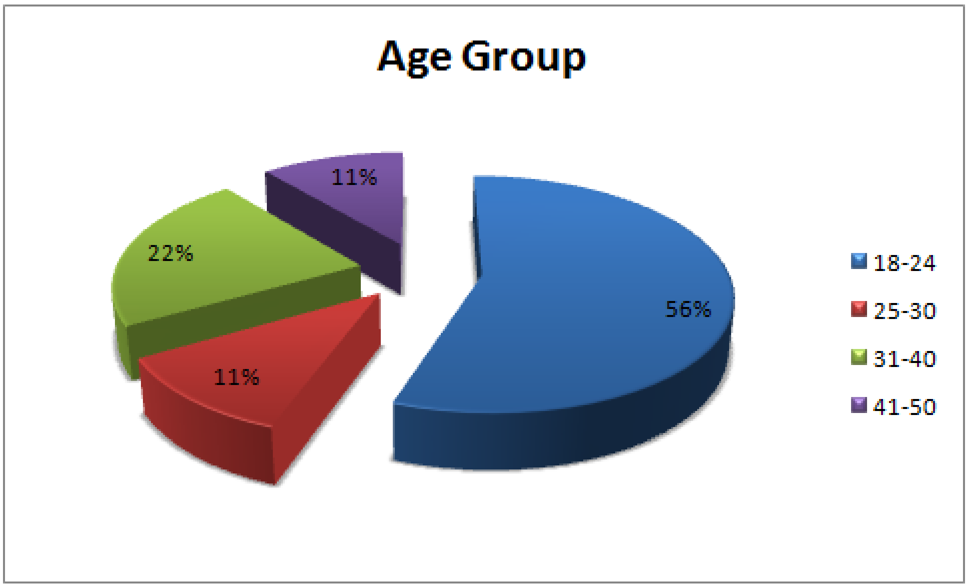



**Figure 15: The age groups of the participants**

The participants had different educational backgrounds: 22% had completed further education, 22% had achieved undergraduate degrees, and 56% had achieved postgraduate degrees.

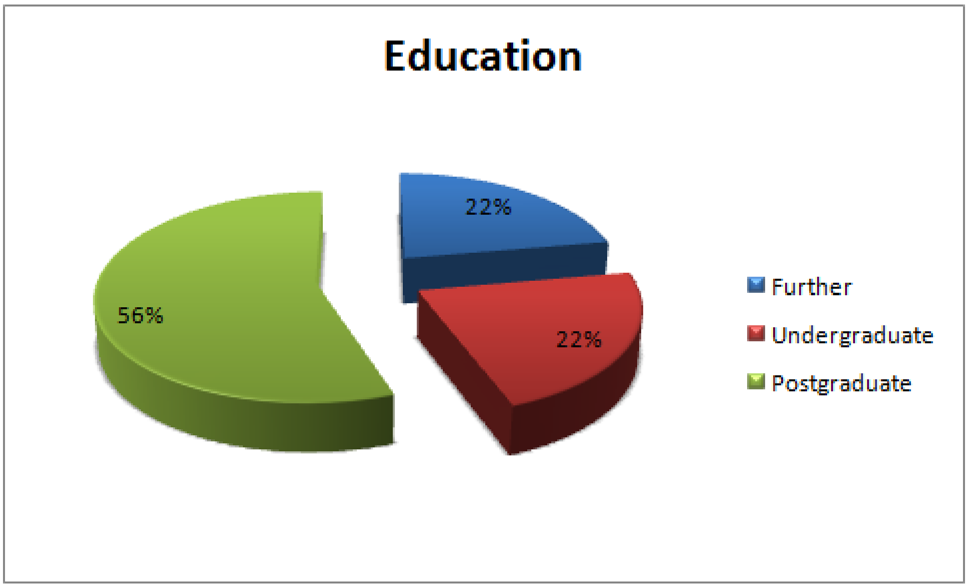



**Figure 16: The education levels of the participants**

The IT knowledge of the participants also varied: 44% had basic computer knowledge and Internet skills, 44% had relevant computer certificates (ECDL/Microsoft/Other), and 12% had a relevant (to computers) tertiary education degree.

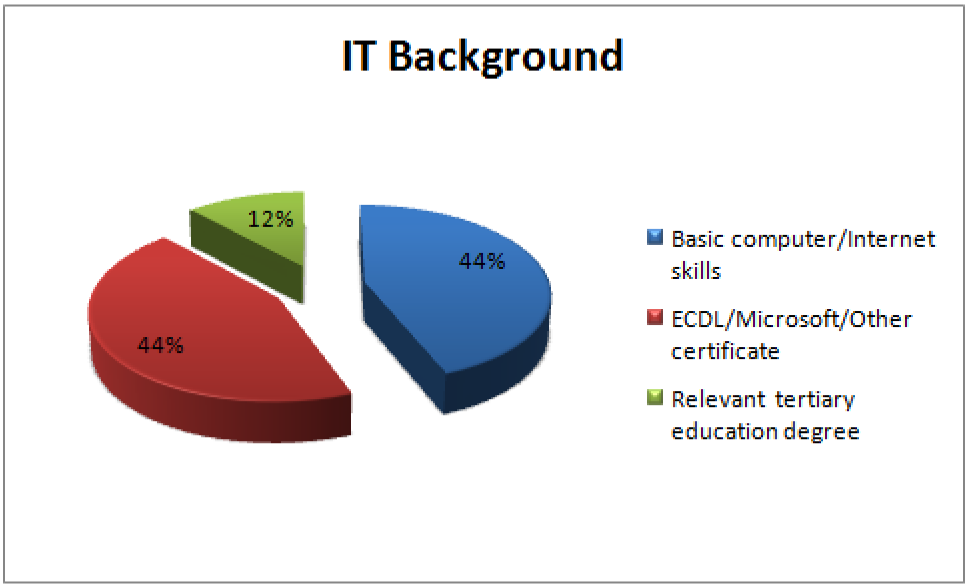



**Figure 17: The IT background of the participants**

All the participants were then asked whether they would like to view an introduction Guide to the system (slides 9 to 19).

Interestingly, one participant (Participant #5) answered no, even though she did not know how to use the Semantic Synthesis mechanism. After she failed to answer Questions #1 and #2 correctly, she asked to go back and view the Guide. She then successfully answered the Quiz and the rest of the synthesis questions (#3, #4, #5), giving either Correct or Semantically Correct answers.

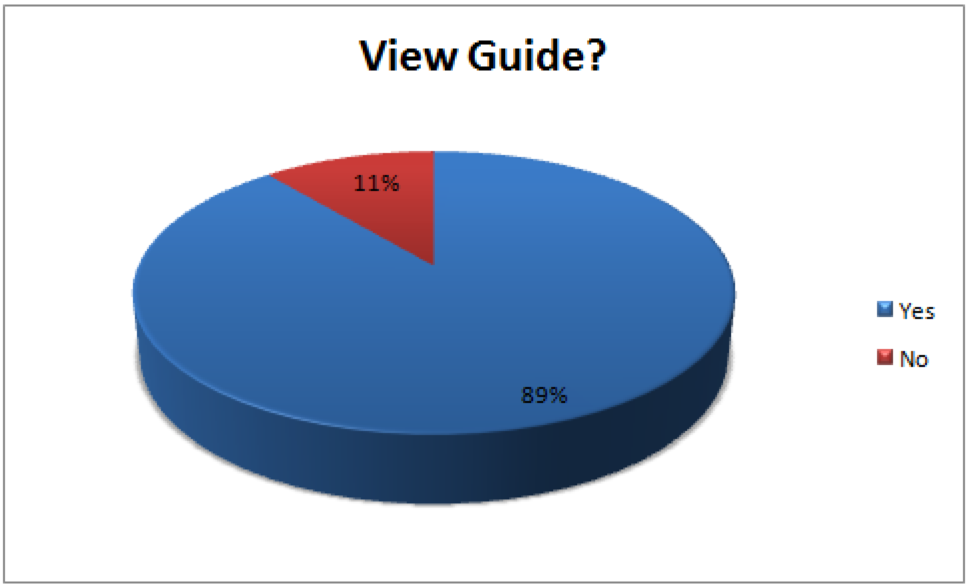



**Figure 18: Did the participants view the guide?**

The participants were asked to answer a **Quiz**, to determine whether the learning objectives of the tutorial guide were achieved: 67% of the participants answered it correctly, with the remaining 33% failing to do so.

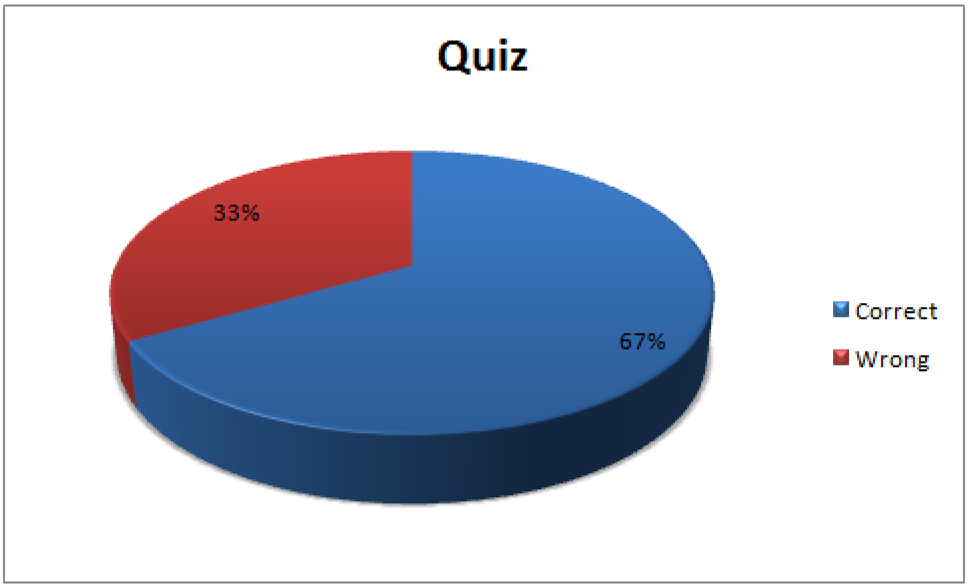



**Figure 19: The answers to the Quiz**

**Question #1**: The participants had to actually create a Semantic Synthesis from scratch. Even though there was no time limit, 67% of the participants got their first answer Wrong, while the remaining 33% provided a Semantically Correct answer. No one gave an entirely Correct answer.

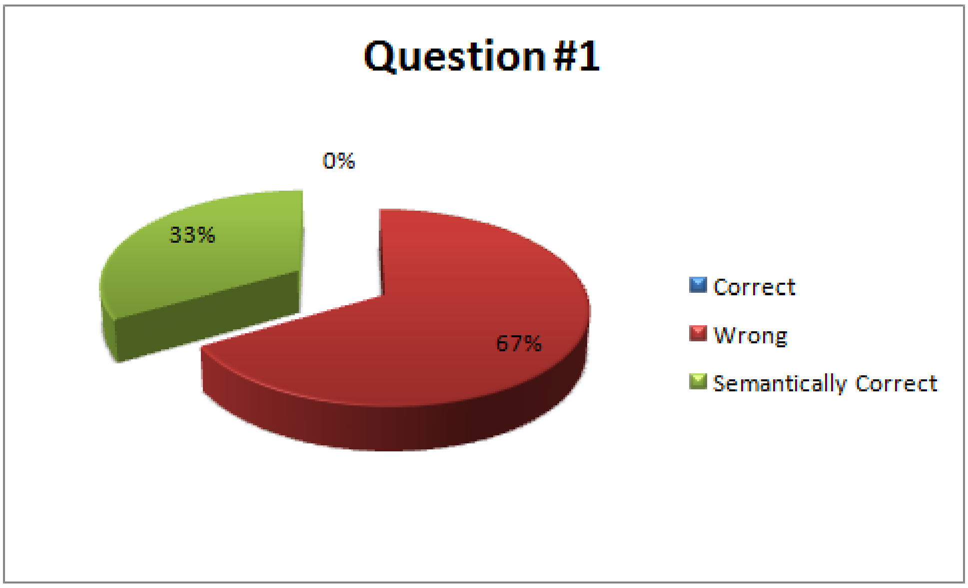



**Figure 20: The answers to Question #1**

It should be noted that out of the five participants who had previously answered the Quiz correctly, three did not manage to successfully create a synthesis from square one for this question. Interestingly, one of the participants (Participant #3), who did not answer the Quiz correctly, managed to create a Semantically Correct synthesis for Question #1. The numbers mentioned in this paragraph intentionally do not include the details for Participant #5 who answered the Quiz correctly only after she had failed Questions #1 and #2.

The answer sheets of the participants who answered Question #1 incorrectly showed that **all** of them had the same mistake: instead of linking all the characteristics directly onto the Person item (first name, surname, gender), they chained them together in various orders, thus searching for nonexistent entries. The mistake was explained to all the participants before proceeding to the next question.

**Question #2**: The participants had to create a Semantic Synthesis from ground up. This time, there was no specific query. The participants were asked to create their own synthesis of whatever they preferred.

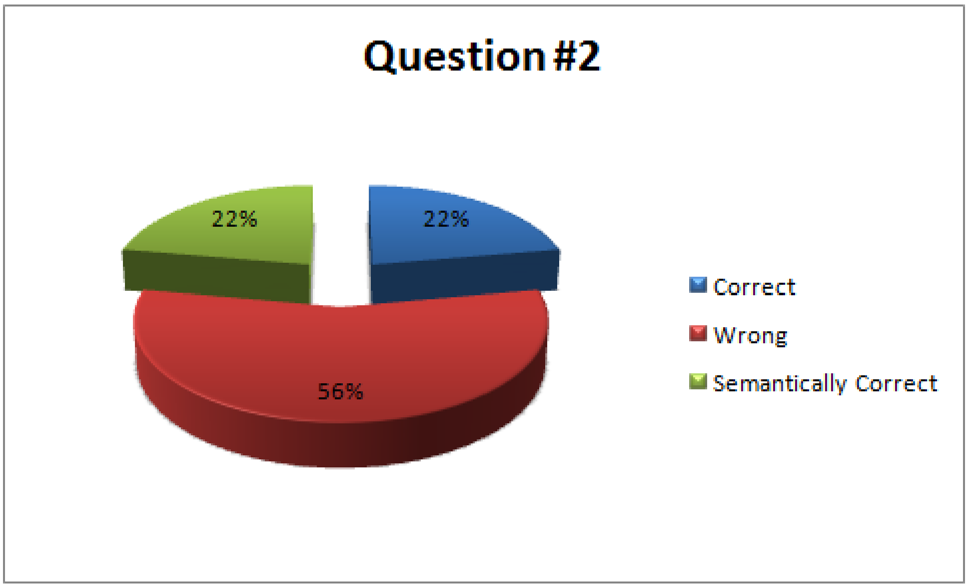



**Figure 21: The answers to Question #2**

In all, 44% of the participants provided Correct or Semantically Correct answers to their own syntheses, with the rest 56% not managing to do so. Excluding the participants who once again failed to connect the items properly, the remaining mistakes came from not setting all the necessary intermediate steps like, for example, skipping straight from a Location item to the name of a city, without using a City item in-between to specify that it is not the name of something else (e.g. a country or county).

**Question #3**: This was the first question where the participants were timed. Interestingly enough, the amount of Semantically Correct answers reached the record peak of **89%**, even though there were no entirely Correct answers. The remaining 11% (one candidate: Participant #4) attempted to approach the answer differently, possibly because of his advanced IT background, but the results were unfortunately incorrect. Additionally, it should be noted that all participants completed the answer within 1 minute, except for Participant #9 who needed 2 minutes.

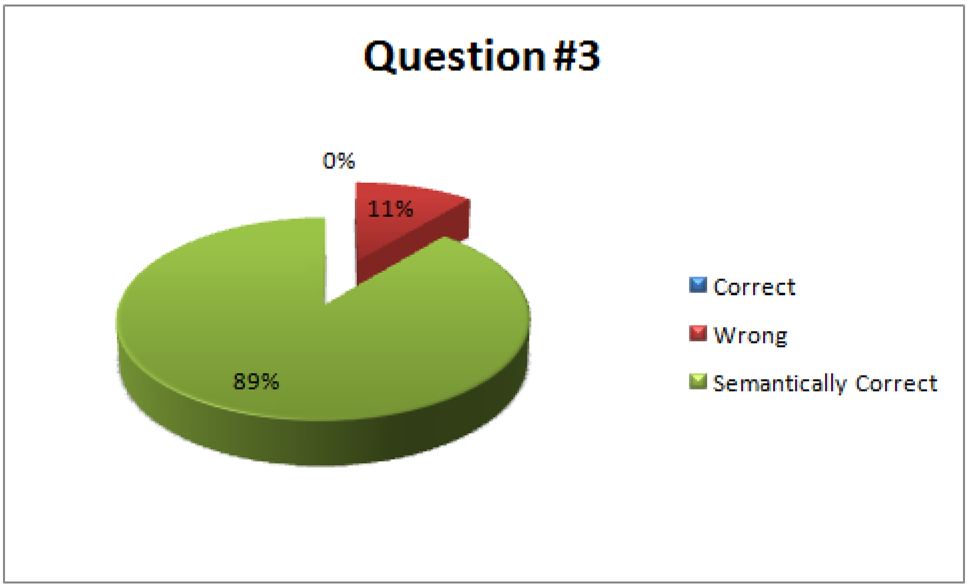



**Figure 22: The answers to Question #3**

It would be interesting to find out whether timing the participants had inculcated a sense of formality to them, making them take the test more seriously and answer the questions more carefully. However, this cannot be verified by this specific user study.

**Question #4**: The participants were once again timed. This question required a more complex answer than the previous ones, hence the increased number of Wrong answers (56%). However, this time, two participants gave entirely Correct answers and another two gave Semantically Correct answers. On average, it took the participants 1.44 minutes to answer this question.

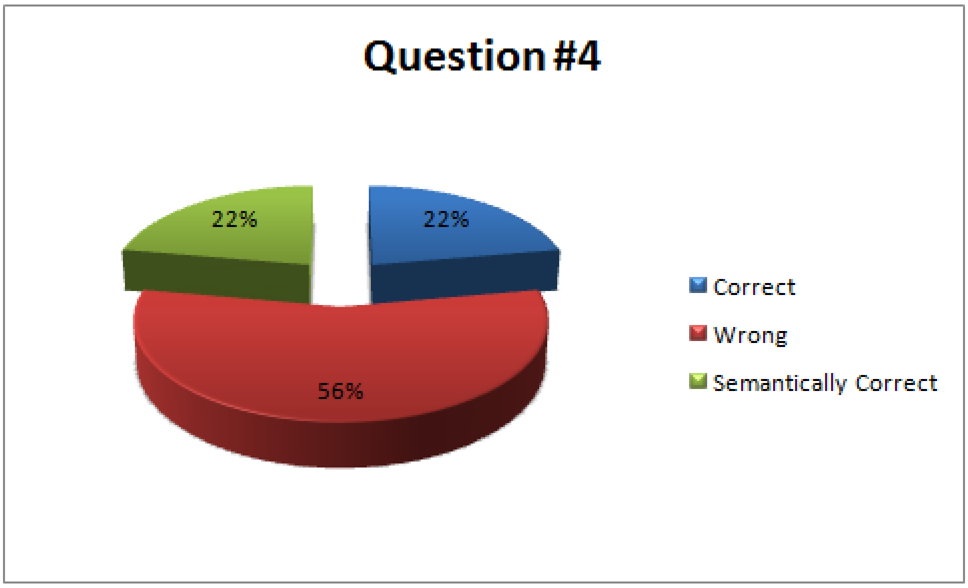



**Figure 23: The answers to Question #4**

**Question #5**: This was the last question where the participants were timed and had to draw a synthesis. This question was more of a reality check to determine whether the participants remembered what they had just learned in the previous two questions. It basically asked the participants to draw the answers of the two previous questions combined, without looking at their previous answers. This time, all the answers obtained were either Correct or Wrong.

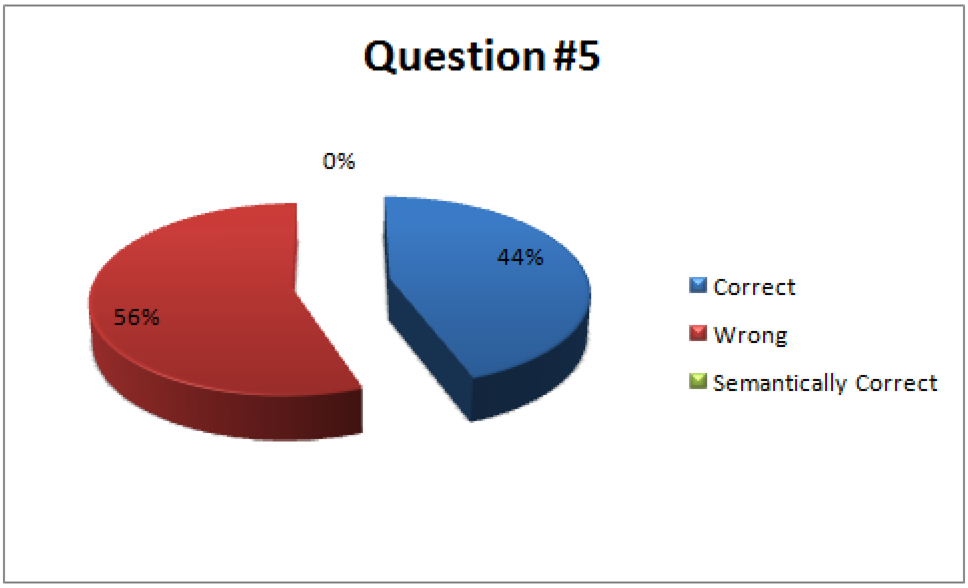



**Figure 24: The answers to Question #5**

Out of the four participants who gave the Correct answers, one of them, Participant #4, had actually given wrong answers to **both** the previous questions, but he remembered his mistakes and provided the Correct answer to this question.

**Question #6**: The participants were asked to use the Wikipedia website to find the actual results for Question #5, so this answer could only be Correct or Wrong.

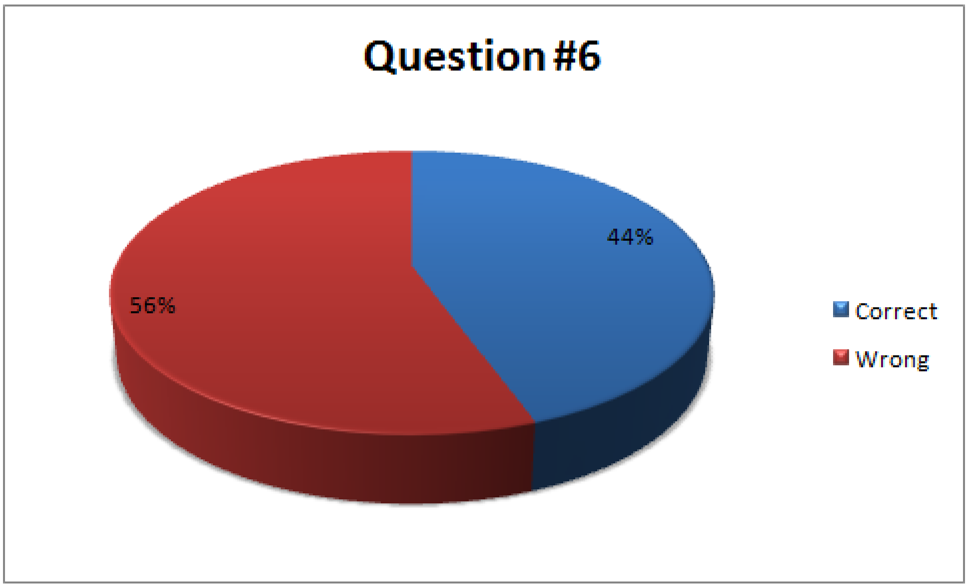



**Figure 25: The answers to Question #6**

It is notable that 56% of the participants failed to give an entirely Correct answer, even though the procedure involved merely comparing two webpages of Wikipedia. Out of the five participants who gave the wrong answer, one added an extra entry by mistake and the remaining four failed to realise that there was a second result sub-page in one of the webpages. Oddly enough, the statistics for this question are identical to those for the previous question (5 Wrong, 4 Correct), with only one participant (Participant #7) answering both questions correctly.

However, the main objective of this question was not to check whether the users managed to use Wikipedia properly, but how long it took them to manually find the required answer. The following diagram illustrates a comparison of the time (in minutes) it took the nine participants to answer Questions #5 (blue) and #6 (red), considering the system data retrieval turnaround times as negligible.

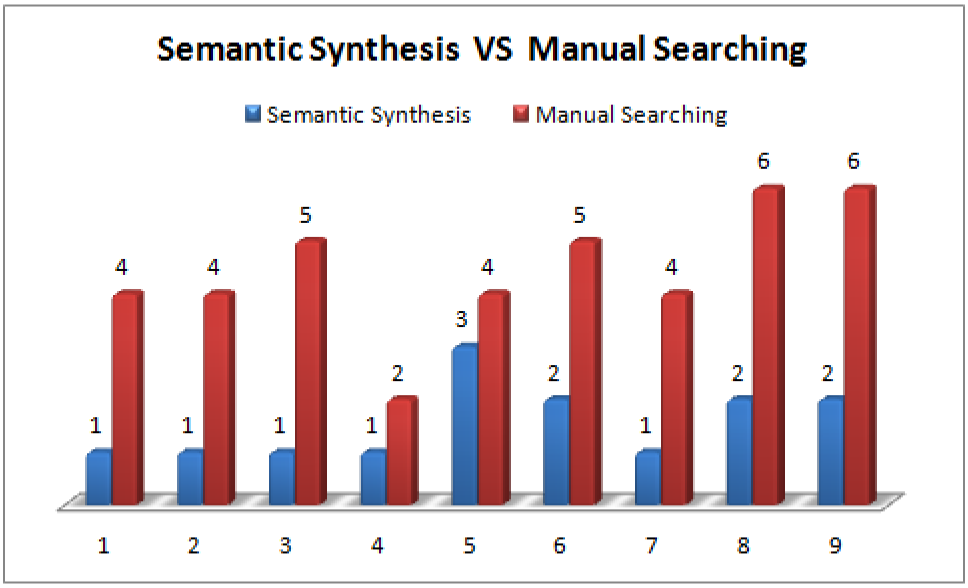



**Figure 26: Semantic Synthesis vs. Manual Searching**

Because both questions have the same number of Correct and Wrong answers, we can consider them as equal and compare their durations. On average, to create the Semantic Synthesis query the participants needed 1.44 minutes, whereas for Manual Searching they needed 4.44 minutes, thus saving each participant approximately 3 minutes. If we do not consider both questions as equal though, then we can only use the timing of the answers of Participant #7 (because both her answers were correct) but, once again, the difference was equal to 3 minutes. This difference can be expressed as a 75% increase in the speed of searching for information, even though the success rate of the query remained stable at 44% (4 Correct answers both times).

At this point, it should be noted that the participants were provided with the URLs of the two Wikipedia lists in order to exclude from the final statistics the amount of time it would have taken them to find the proper pages to read. Since Semantic Synthesis retrieves its results from the Semantic Web ontologies of the Metakosmos system, the mechanism itself acts as a "**filter**" of information – not a Web Spider or Smart Agent looking for results "in the wild". Therefore, asking the users to use the Web in order to track down the required – very specific – list of six results (i.e. a needle in a haystack) would be unfair in terms of timing and also outside the context of this research, because the benefits of using ontologies is not the subject of this user study and has already been researched in the past (Li *et al.*, 2004).

**Question #7**: This question showed the participants four different result-set display modes and asked them to state their personal preferences by saying which one they liked best. Their answers are presented in the following figure.

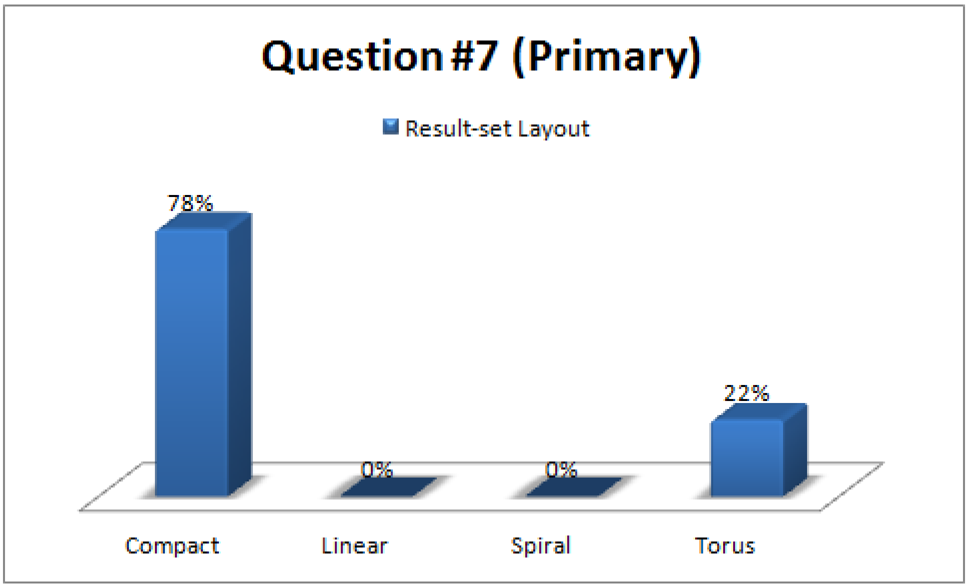



**Figure 27: The primary answers to Question #7**

Almost everyone selected the "Compact" mode, with only two participants selecting the "Torus" mode. After the participants gave their initial responses, they were asked if they would use the same display mode for a large numbers of results (e.g. more than a hundred). Their second answers are presented below.

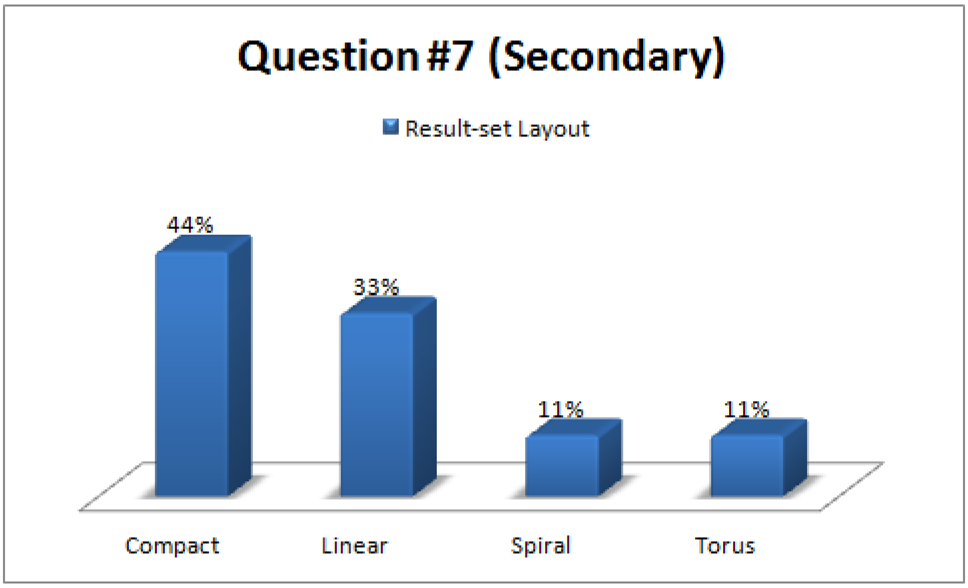



**Figure 28: The secondary answers to Question #7**

Four participants selected the “Compact” mode for the second time, but the rest changed their answer to “Linear” and “Spiral”. Additionally, one of the participants who had selected “Torus” had then switched to “Linear”.

**Question #8**: This was the first of the three questions that used a 1 to 5 scale, with 1 being “Not at all” and 5: “Yes, absolutely”. The participants were asked whether Semantic Synthesis was easy to understand. On average, the overall rating of the participants was **3.78**.

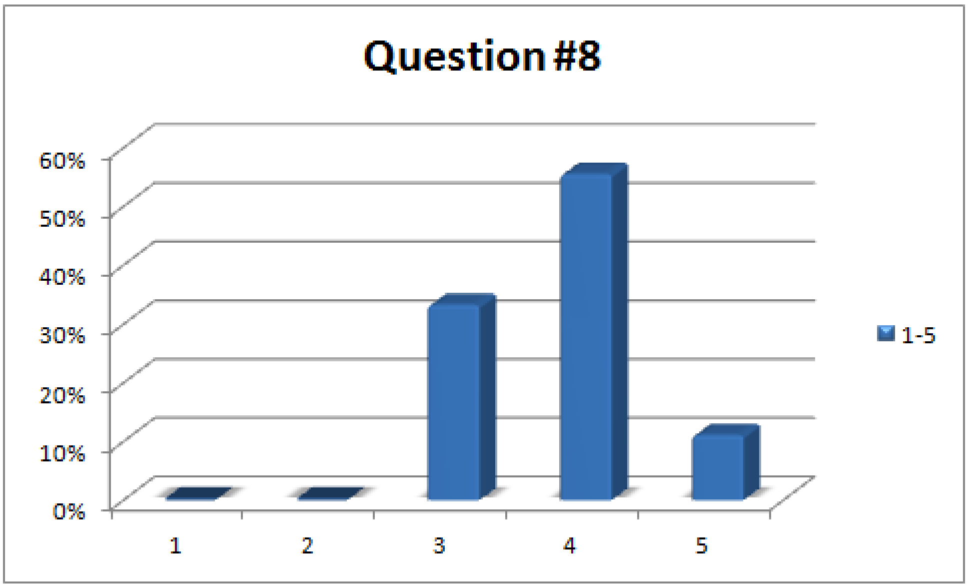



**Figure 29: The answers to Question #8**

**Question #9**: The participants were asked whether Semantic Synthesis was easy to use, again using the same 1 to 5 scale. This time, the average rating was **4.11**.

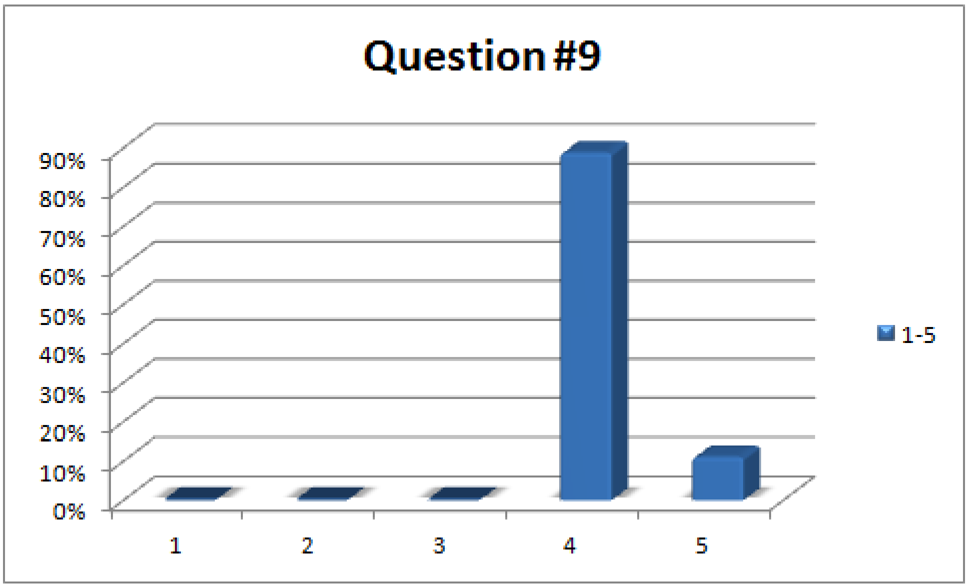



**Figure 30: The answers to Question #9**

**Question #10**: The participants were asked whether they would use a system like Metakosmos with Semantic Synthesis and the average rating was, a very promising, **4.33**.

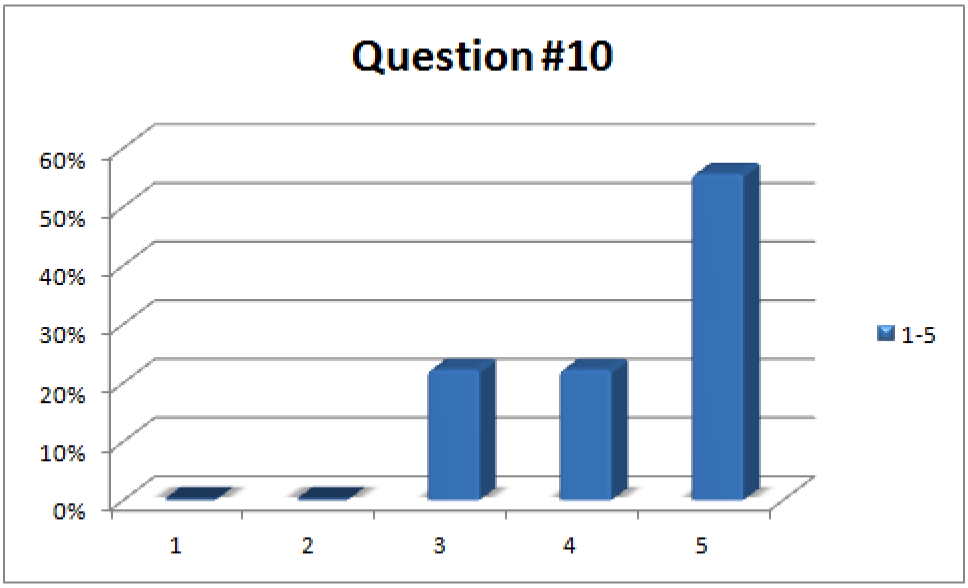



**Figure 31: The answers to Question #10**

In addition to the data recorded above, the participants also mentioned to the researcher that they would definitely use Metakosmos for complicated searches that combined many different criteria, because it could give very specific results. Furthermore, certain participants made positive comments regarding the personalised nature of the system and how appropriate it would probably be for visual learners.

### 5.2.5 Evaluation Outcomes

Based on the results mentioned above, it becomes obvious that the Semantic Synthesis mechanism can be successfully taught to users, thereby helping them take advantage of the speed increase it offers for complex searches, while providing them with a useful multi-criteria search tool – as soon as they understand how to control it. However, the results also show that there is room for change and improvement.

Although none of the participants had used the Semantic Synthesis mechanism before, only with a short ten-slide tutorial they managed to learn how to use it and successfully created syntheses to query the system. Even though all nine participants gave incorrect answers at some point, all of them gave at least one Correct or Semantically Correct answer. This fact supports the idea of further creating guides and tutorials which need to be embedded in Metakosmos for future users to watch, read, and interact with, in order to improve their understanding of the system and its underlying mechanisms.

Via these guides, the importance of proper item linking should be emphasised, illustrating the logical significance of connections, thus clarifying their purpose: to associate items in a coherent manner, as per the rules of logic, in order to search for, and return, all known syntheses of the system which share the same basis.

If the guides are successful, the system could attract users looking for answers to complex questions with multiple criteria, as well as other inquiring individuals. In other words, Metakosmos should initially aim to attract users by advertising the speed increase it offers for complex queries, in order to introduce new users to the system who may then use it even for simple queries, because they would eventually get accustomed to and familiarised with its mechanisms.

In light of the common mistakes observed in the study, it becomes obvious that for the system to be easy-to-use, commonly used or popular items should appear first when users add new items to their syntheses, thus minimising illogical queries while also saving look-up time. This can be further personalised based on the usual searches of each individual member of the system, thus making the interface adapt to the searching style of each user.

Finally, the information bar could be used to display a read-only text version of the currently composed synthesis, in order to help users realise what they are composing and to offer a "plain text search engine" feel, thus aiming to bridge the gap between text and visual searching. The appropriate section of this thesis (5.1.4) has already been updated with this change.

## *5.3 Summary*

In this chapter the initial models of the design of the resulting interface were presented, as well as the user study that was used to evaluate the "Semantic Synthesis" mechanism.

# Chapter 6: Development & Implementation

## *6.1 A Semantic Web Based Search Engine*

Many different tools, methods and combinations have been tested and reviewed during the creation of this thesis. Even though Semantic Web ontologies have been around for more than a decade, no definite method for storing extremely large data collections has been agreed upon.

The latest Oracle 11g Release 2 has many Semantic Web technologies embedded by default (Lopez, 2010). As the authors mention, Oracle can load, query and inference Semantic ontologies while offering SQL, SPARQL and Java (Jena & Sesame) based adapters. This is a very interesting choice, especially due to the high level of support offered by the company, but other factors also have to be considered, such as licensing costs in a future wide-scale commercial launch of the product.

Another suggestion for the ontology (Section 4.4) is the D2R Server tool ('D2R Server,' 2012) which reads data stored in relational databases and then publishes them to ontologies using the RDF language. This provides a useful solution for systems that are already using relational based systems and want to switch to the Semantic Web but, since the system in this thesis is new, this tool would not serve its purpose.

In the book “Querying the Semantic Web” (Chebotko & Lu, 2009), the authors have explored new methods for faster querying of Semantic Web ontologies. In the end they formalised the relational algebra based semantics of the SPARQL language in order to bridge the gap with SQL. Their findings could prove to be useful in a commercial launch of the final system as defined in Section 3.3.

In addition, Liyang Yu mentions in chapter 2 of his book (Yu, 2007) a few methods of attempting to store, retrieve and query Semantic Web ontologies. He also suggests in later chapters that Web Services should be used for massively querying different ontologies. This is a quite interesting concept that could also be adopted and it is currently set as future work (Section 7.4).

In a recent paper (Curé, Lamolle & Le Duc, 2011), the authors suggested using the NOSQL (Not Only SQL) database model for creating and storing customised ontologies, while eliminating scalability issues. Their ideas could also prove useful for the needs of the universal ontology of the system.

Finally, Bhavani Thuraisingham has proposed using XML Databases to store Semantic Web data since RDF is also XML based (Thuraisingham, 2002). This appears to be a very fitting idea for such a system, even though XML Databases are not very popular any more.

## *6.2 With X3D Visualisation of Queries and Results*

In their 2003 book “Visualizing the Semantic Web”, Vladimir and Larissa Geroimenko stated that since X3D and SVG are both written in XML, they can be easily integrated with each other, as well as with other XML based languages and services. This means that these two standards will play an important role in visualising information and creating graphical interfaces for next generation Web applications (Geroimenko & Chen, 2006, p. 132) such as Metakosmos. This is the reason why X3D was chosen as the standard output language of the suggested system interface presented in this thesis (5.1.2 and 6.3).

After the creation of the initial X3D model in Notepad (Section 5.1.2), it became obvious that better and easier methods needed to be employed in order to achieve faster model creation with better quality and control. At first, a combination of NetBeans with the X3D-Edit add-on was used, but the product at some point became unresponsive, perhaps due to a faulty update, and had to be set aside. The next attempts were with the Blender 3D creation software, which uses the BS Exporter plugin to export designs to X3D format. After some time, the student version of Autodesk 3D Studio Max with the Instant Reality plugin was also used, which finally became the software of choice for this thesis.

It should be noted that in order to achieve optimum, Web friendly, results, the 3D meshes had to be optimised by using low poly standards, as if working with handheld devices. The polygon count of each mesh had to be kept as low as possible – but not too low, because the graphics must show up clearly for the user to see. The ideal threshold depends on the desired outcome which is based on the number of items immediately visible on the interface.

In order to view the results, a variety of players was tested. The BS Contact player was very easy to install and use but the constant watermark of the free version interfered with the output. The Octaga X3D player was also simple to use, but it seemed to translate the output differently from the other players. FreeWRL was easy to set up, but it would not present complicated X3D files in their entirety. Finally, FluxStudio was selected, because it overcame all the aforementioned issues and it also offered a working browser plugin.

Currently, in order to view the results inside the Web browser, a plugin is required, because X3D support has not yet been widely popularised. With the upcoming official release of HTML5, X3D will probably be standardised as an embedded Web feature, which will render the X3D plugins redundant. In this way, X3D will probably become the official standard for displaying 3D graphics inside the Web browser.

Until then, “shortcuts” must be used in order to bypass the need for a plugin. The X3DOM project ('X3DOM,' 2012) has managed to use JavaScript in order to directly integrate the X3D graphic nodes as DOM elements inside an HTML5 webpage (Behr et al., 2009). They are still trying to transfer fully all the features of X3D, but it looks as though it is the best solution at the moment in order to avoid extra plugins.

X3DOM can possibly be further enhanced by adding parallel JavaScript functions. Intel has implemented RiverTrail (Intel, 2012), an addition to JavaScript that lets the system use all the cores of the CPU, thus allowing for parallel programming and execution. There are obvious differences between serial and parallel graphics representations. In Figure 32, the CPU usage in an 8 core CPU without RiverTrail (left) is only 14%, but with RiverTrail enabled (right) it reaches 91%. This can prove to be useful when displaying a large number of items on-screen.

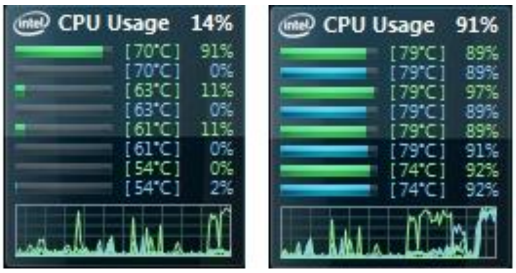



**Figure 32: CPU usage before and after enabling RiverTrail**

The frames per second in the above example increased from 1 to 10, making the movement of the graphics more fluid and natural. Parallel JavaScript scripting can help projects such as X3DOM achieve better results and some are already looking into combining them (Klein et al., 2012). Please note that the tool used to measure the CPU usage in Figure 32 was created by OrgLog (OrbLog, 2009), specifically for Intel CPUs.

Combining HTML5 with X3DOM and parallel JavaScript can boost the X3D environment to high speeds and quality, while at the same time avoiding the need for an extra browser plugin. This combination can be further enriched by taking advantage of the HTML5 Fullscreen API ('HTML5 Fullscreen API,' 2012) which would allow the system to execute in true fullscreen mode (as, for example, games or movies do), thus enhancing the overall experience.

In its final form, the system will resemble a HTML5 app, ready to be used by the Web OSes of the future.

## *6.3 The Resulting X3D interface*

All the system requirements, analysed in previous sections, have been implemented into X3D models. Screenshots from a suggested resulting interface can be found below.

Please note that the information about the grape varieties that is visible in certain screenshots has been obtained from the corresponding Wikipedia article (Wikipedia, 2012). Any other data depicted in the screenshots are just dummy values. The same data were used for the user study that evaluated the Semantic Synthesis mechanism (Section 5.2).

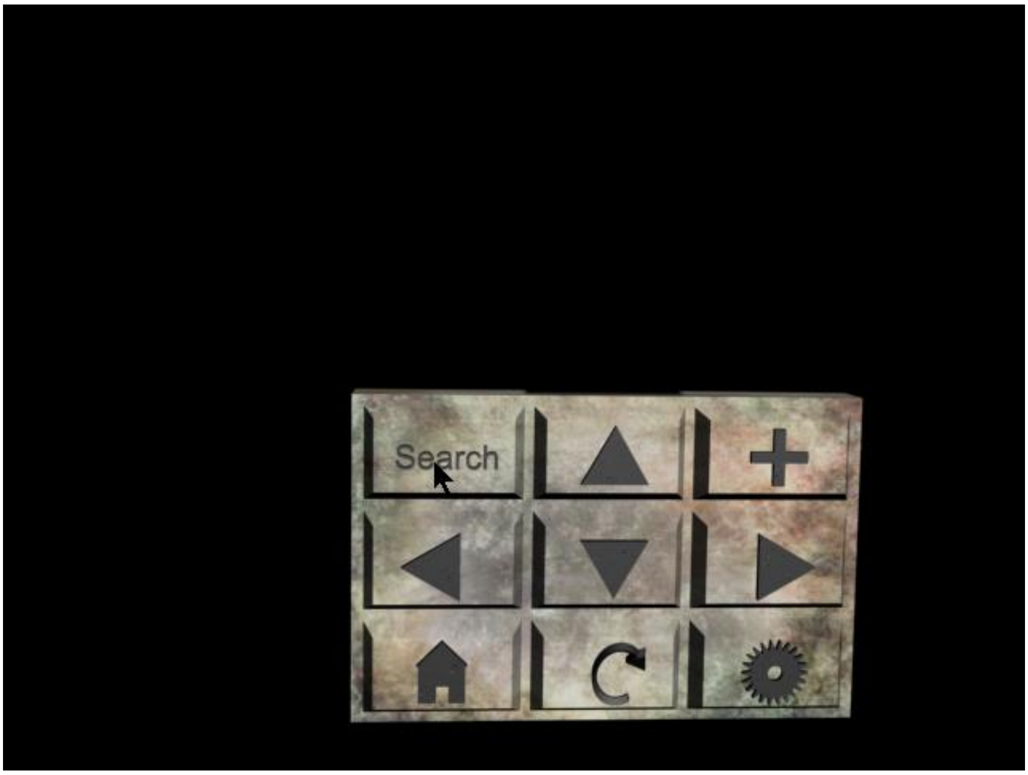



**Figure 33: An empty user interface containing only the companion menu**

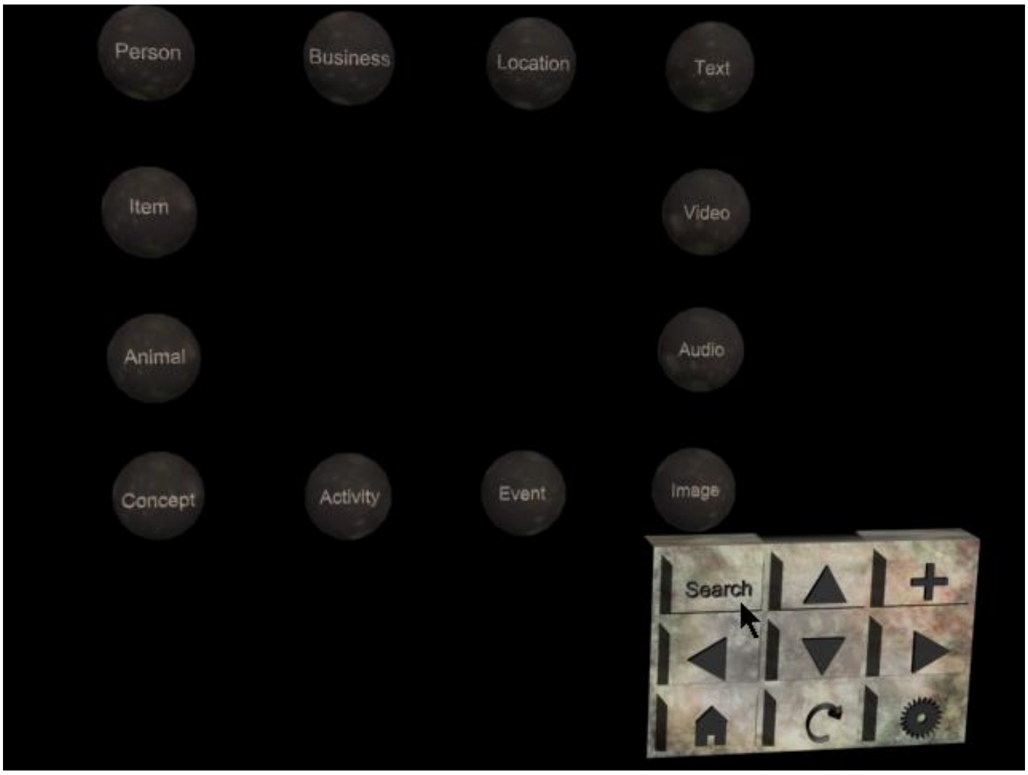



**Figure 34: A custom user interface with 12 items immediately available (3D)**

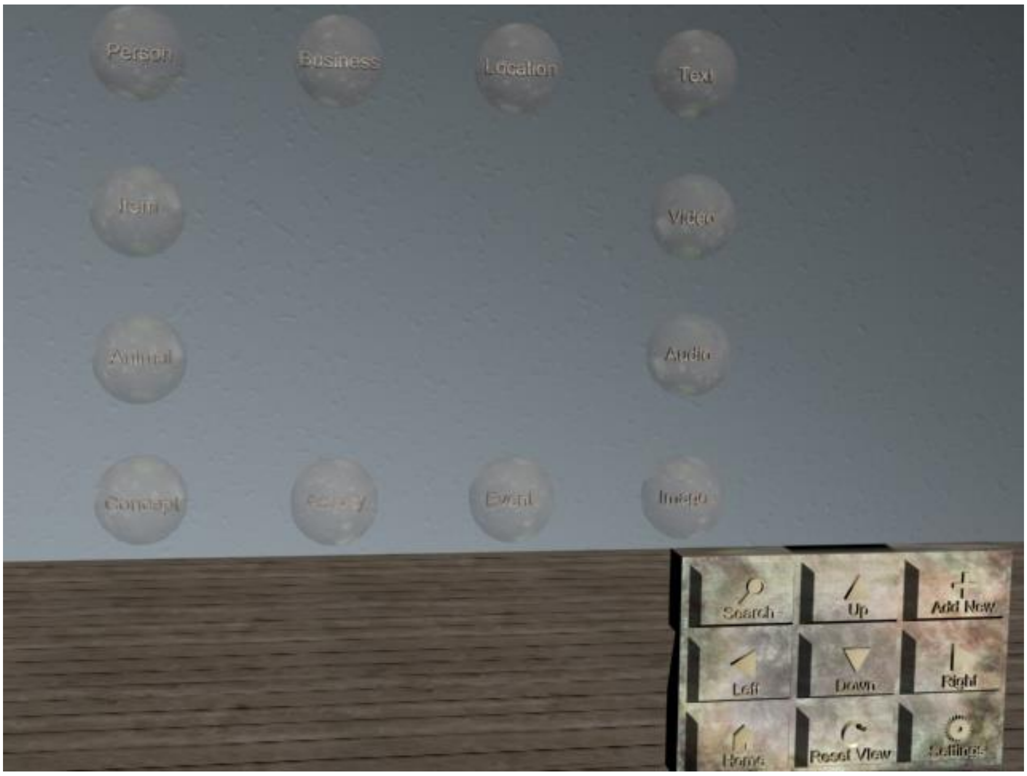



**Figure 35: Custom interface with a custom wallpaper**

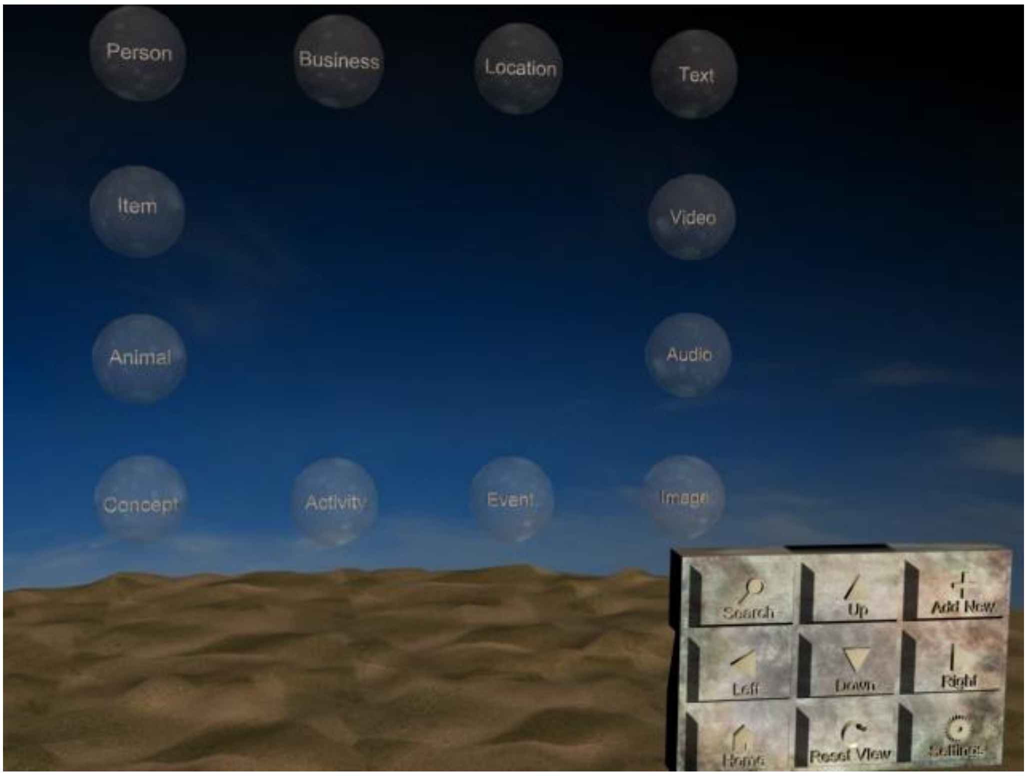



**Figure 36: Custom interface with another custom wallpaper**

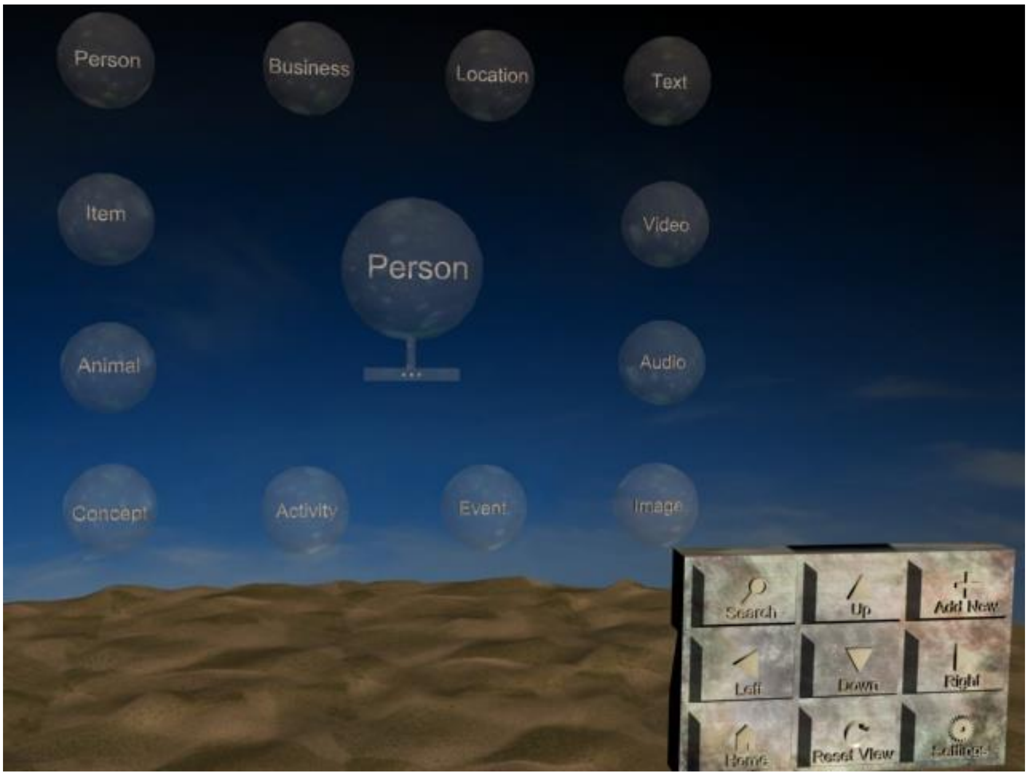



**Figure 37: The user selects a Person concept**

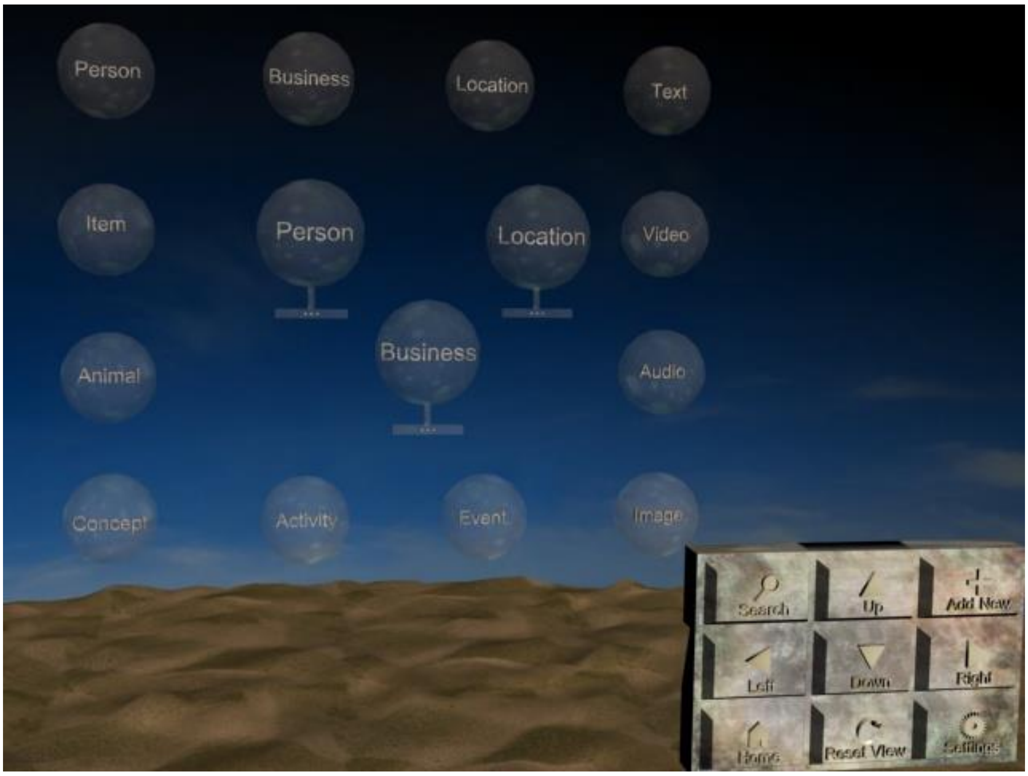



**Figure 38: The user also selects the Business and Location concepts**

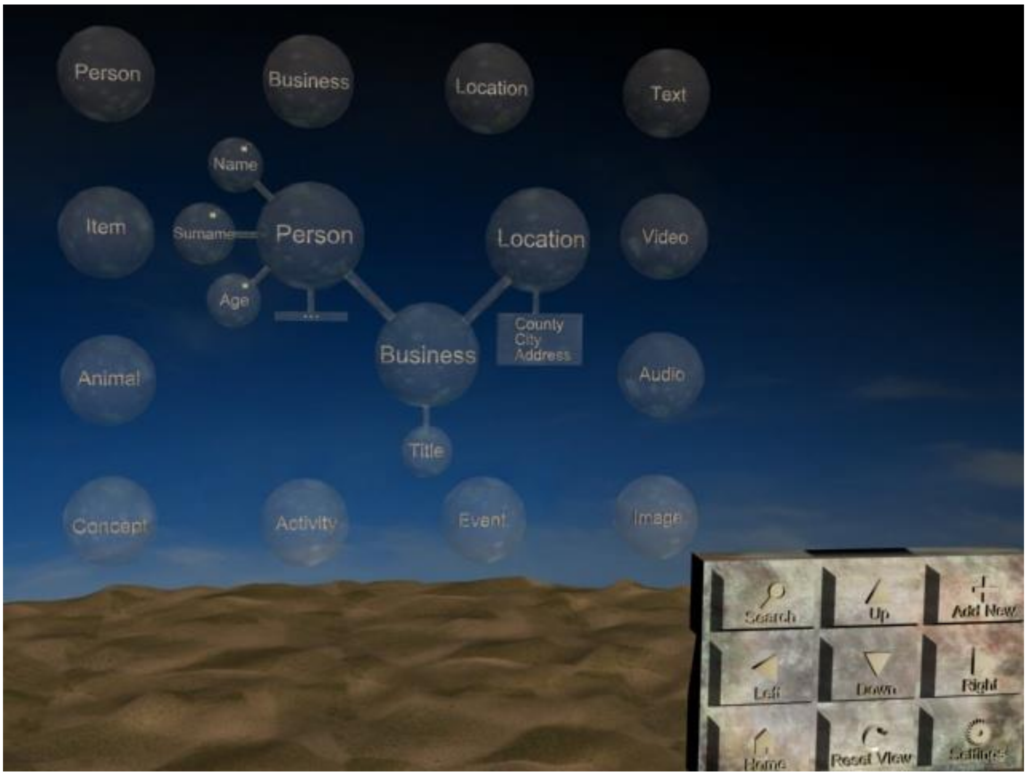



**Figure 39: The user adds more fields to the selected concepts**

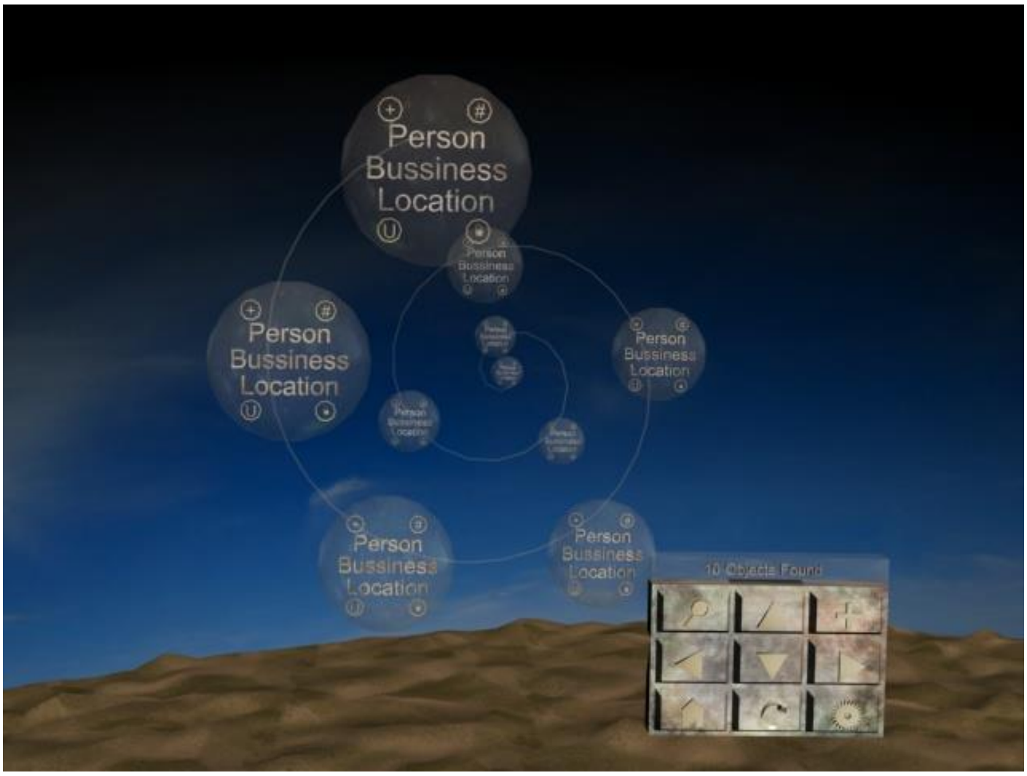



**Figure 40: The results are returned in spiral format (axis showing)**

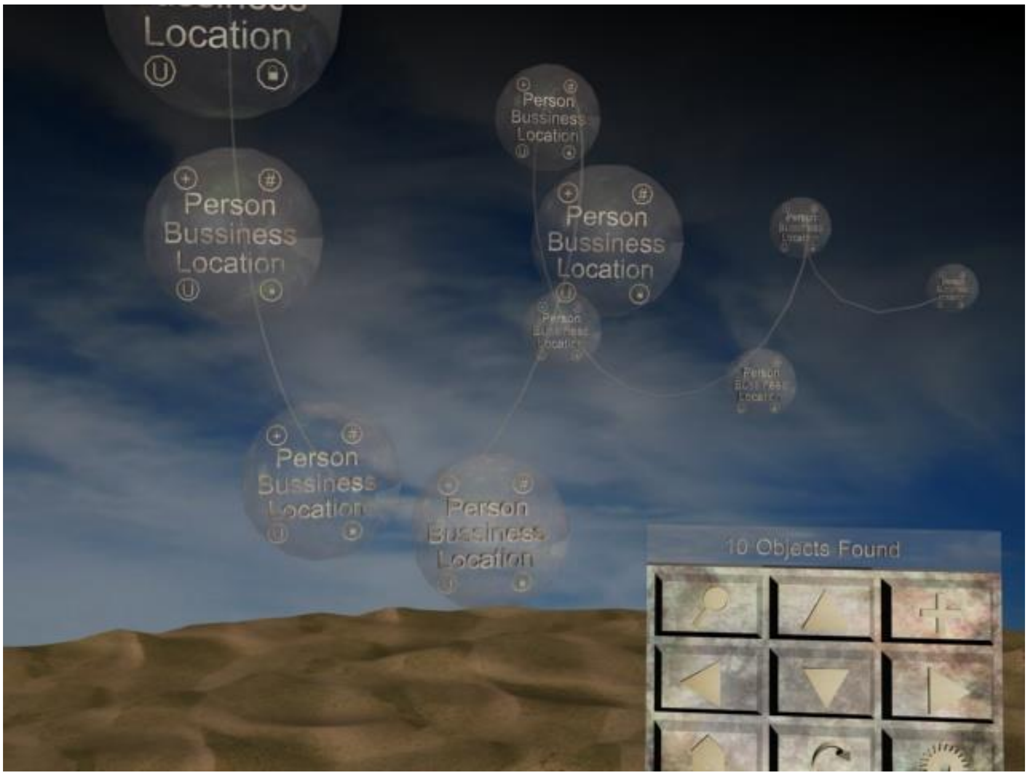



**Figure 41: The spiral from a different angle with the rotated labels**

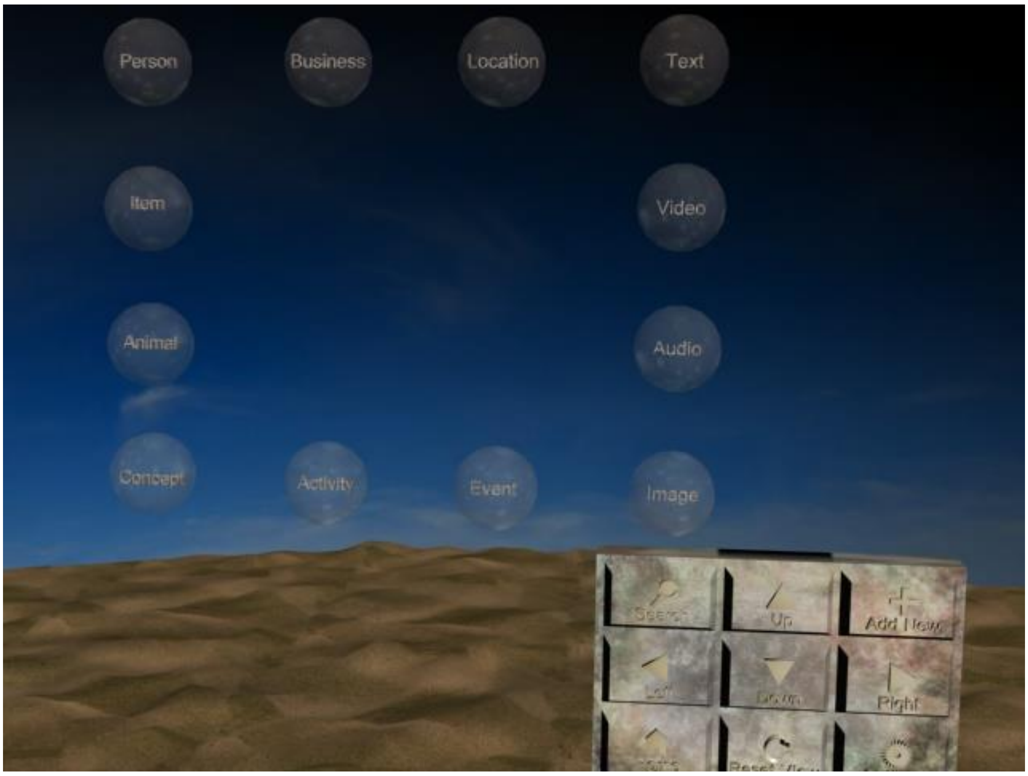



**Figure 42: The interface has been reset/cleaned**

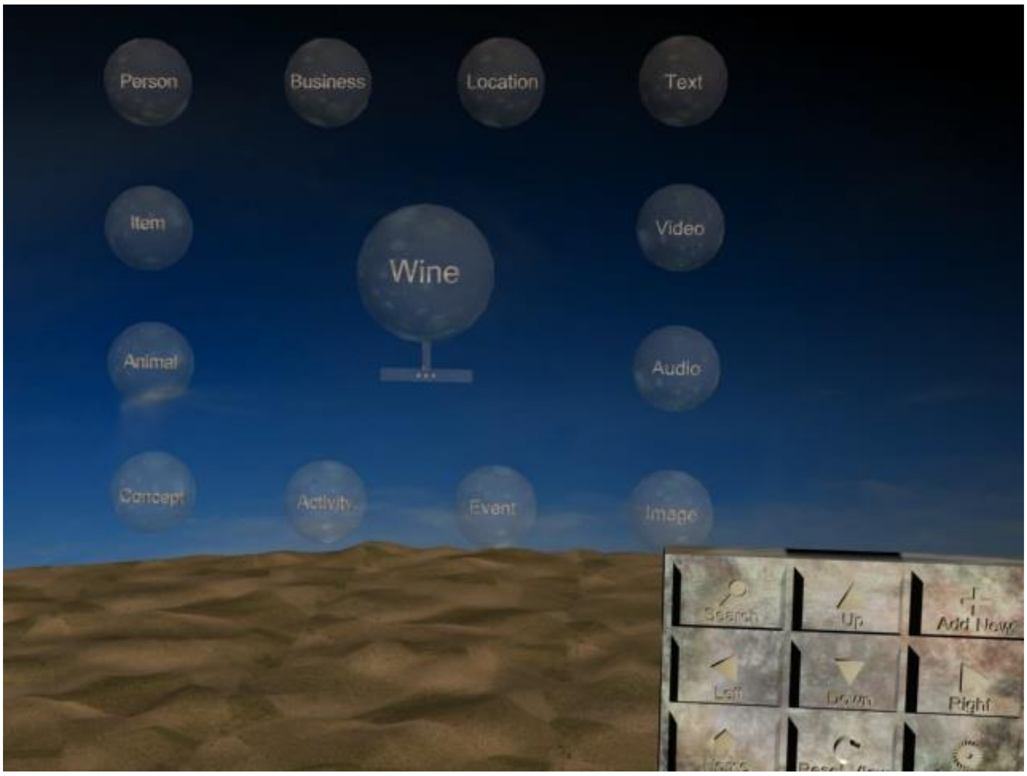



**Figure 43: Using the search button, the Wine concept is retrieved**

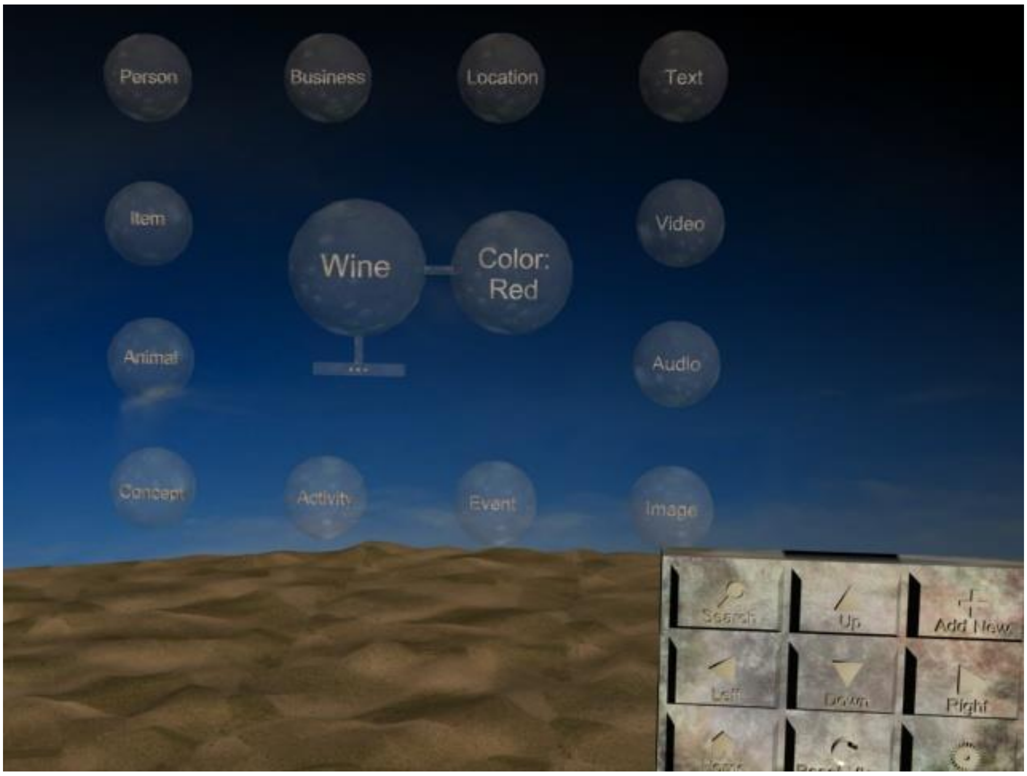



**Figure 44: The Color concept is attached to the Wine concept**

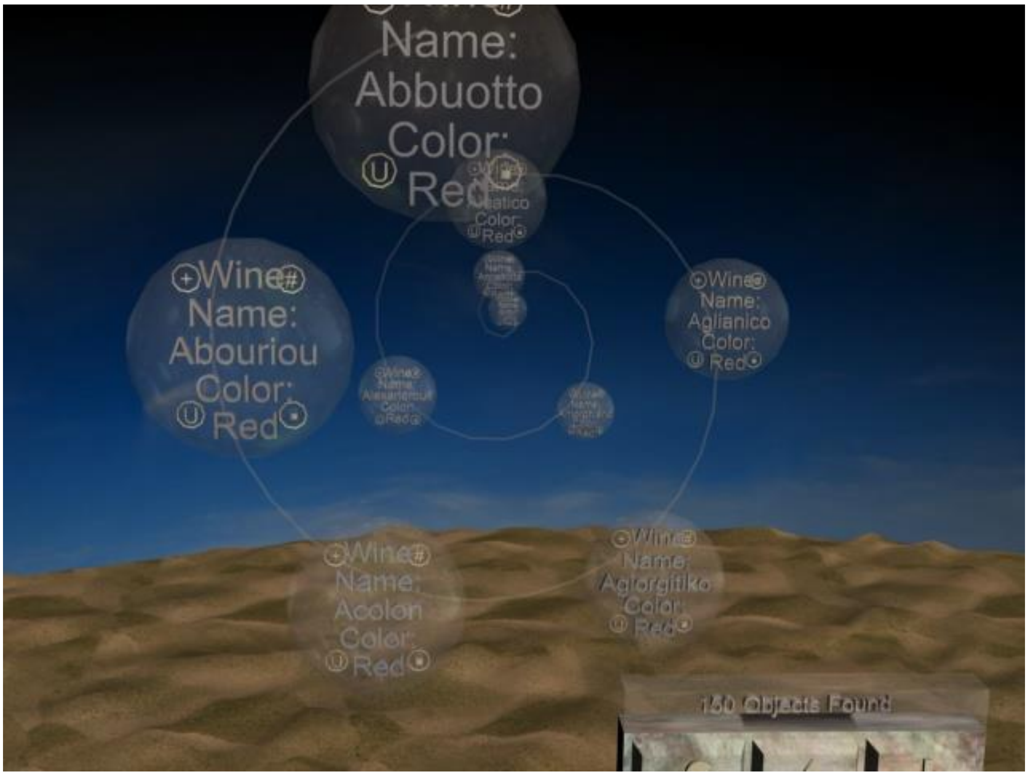



**Figure 45: The search returns all the red wines contained in the system**

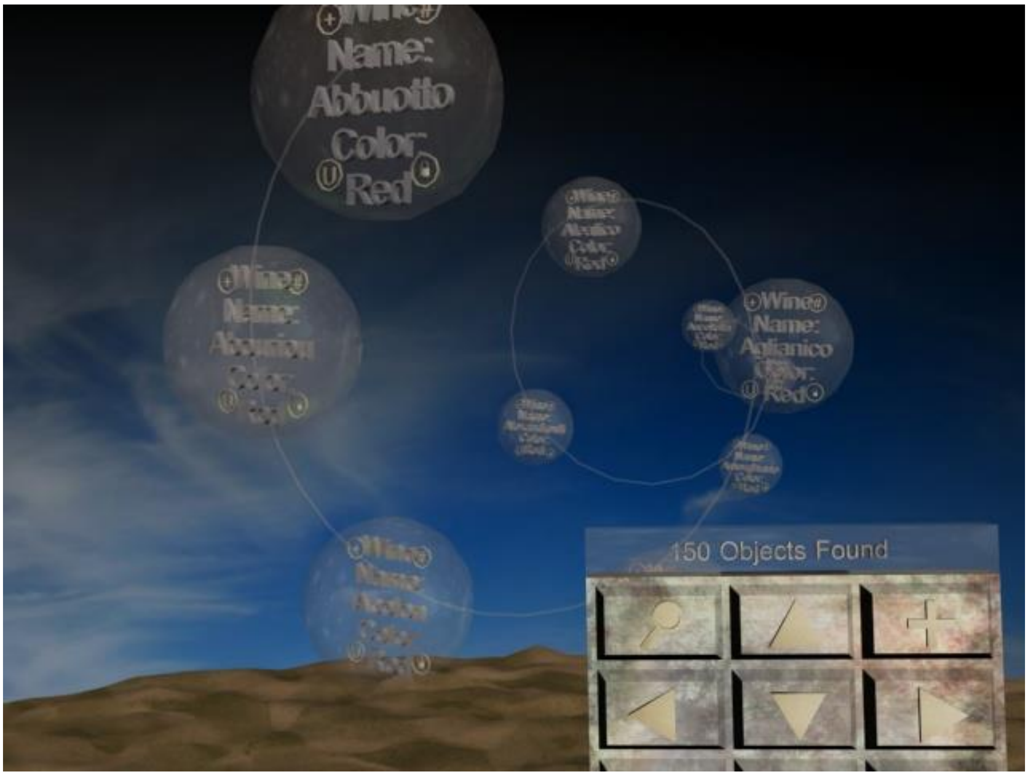



**Figure 46: The information bar text displays the total number of results**

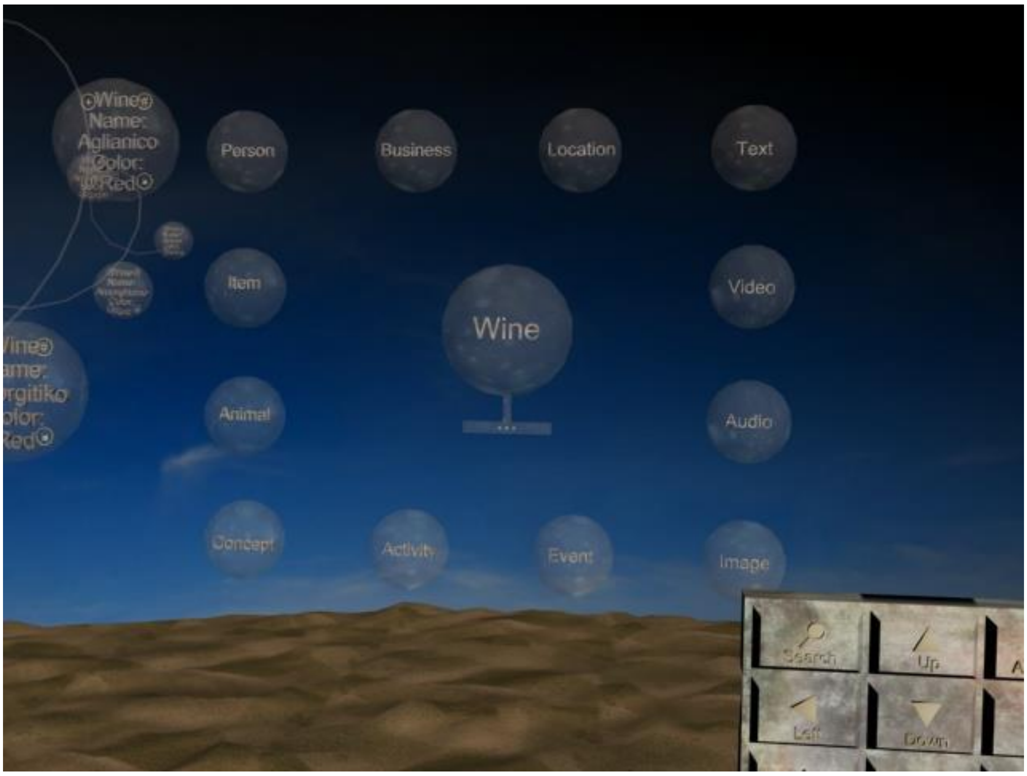



**Figure 47: The user creates a new search parallel to the previous one**

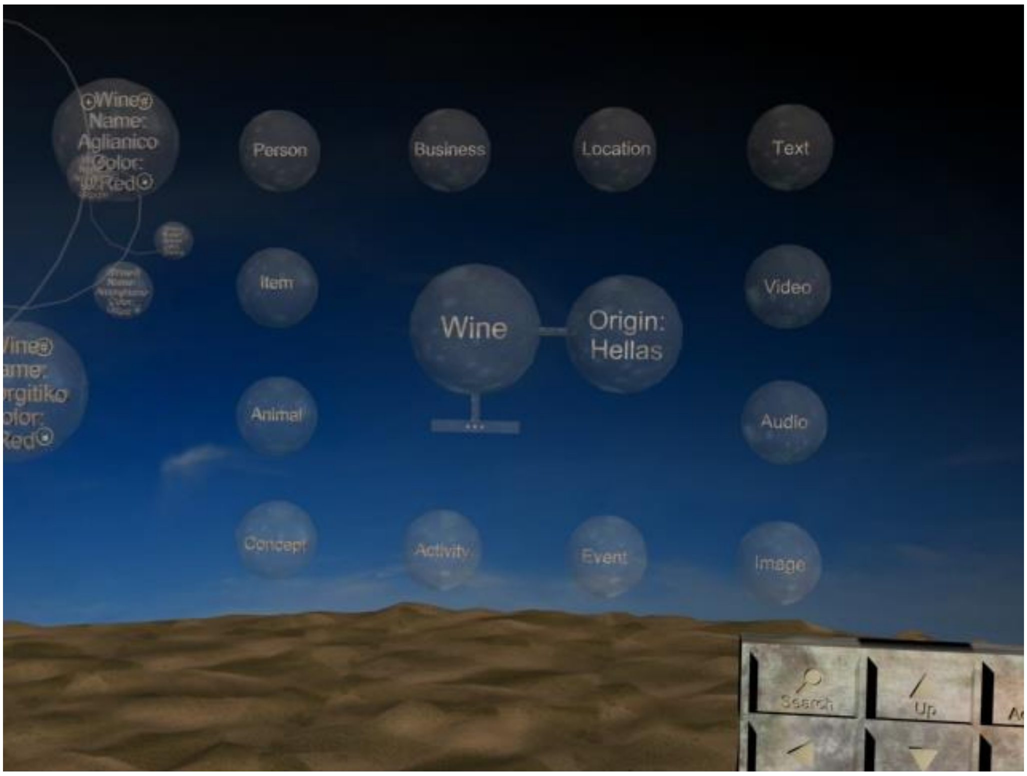



**Figure 48: This time the user searches for Hellenic wines**

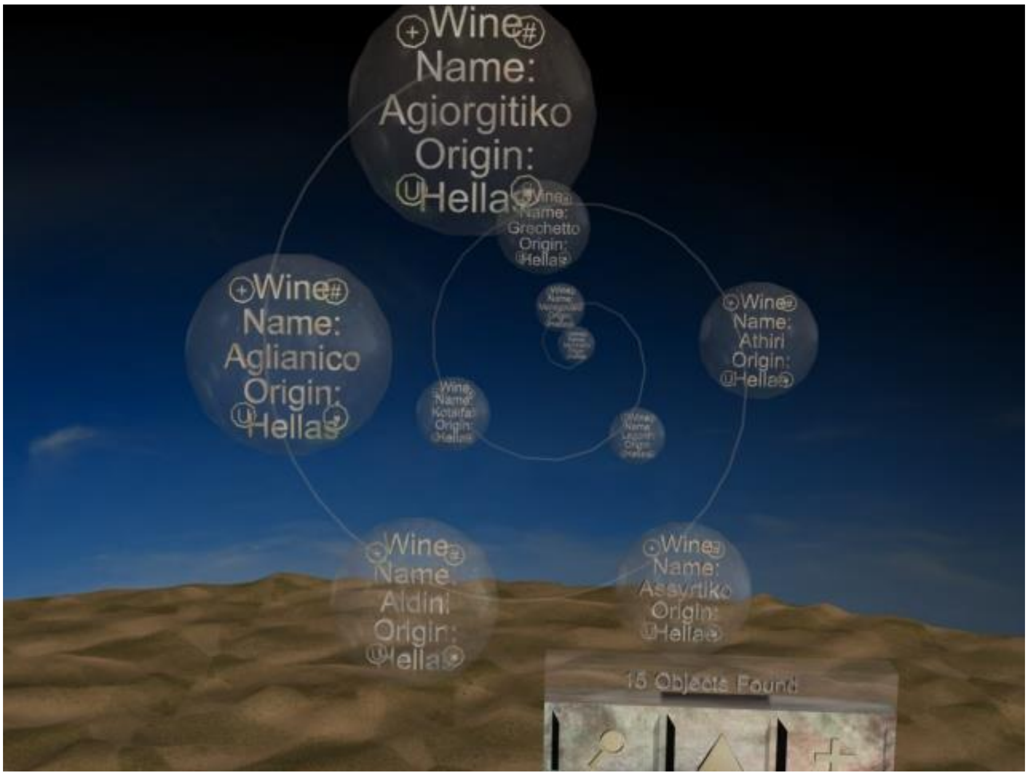



**Figure 49: The Hellenic wines are returned in a second spiral**

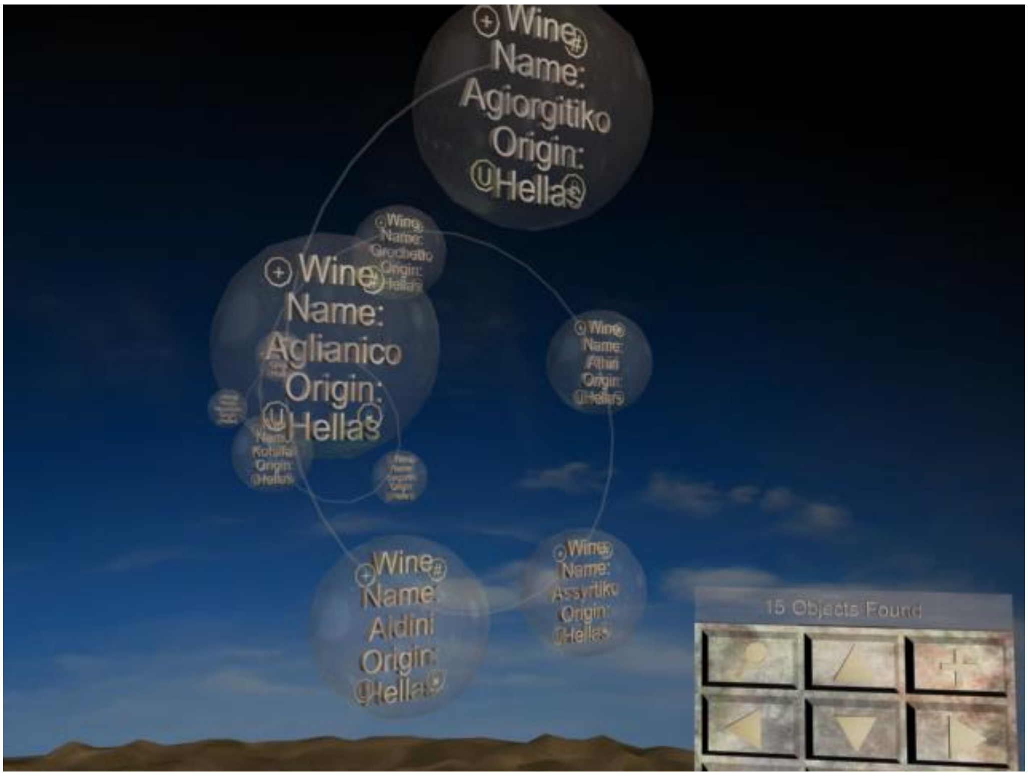



**Figure 50: This spiral has fewer results than the other one**

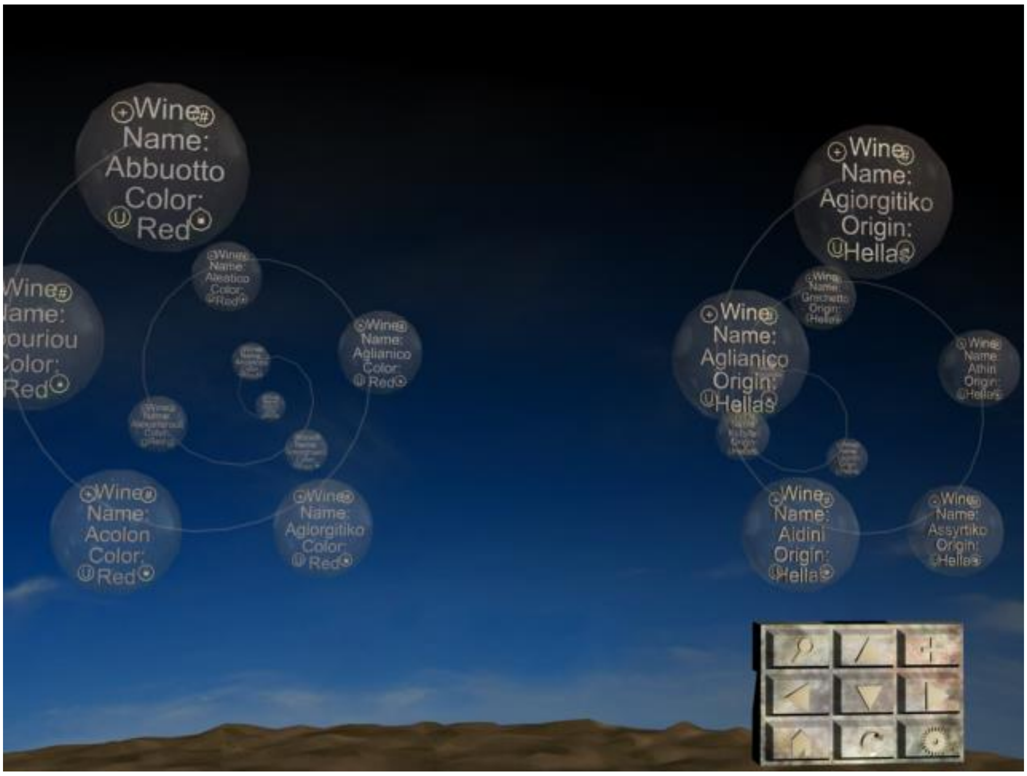



**Figure 51: We can see the two result-set spirals side-by-side**

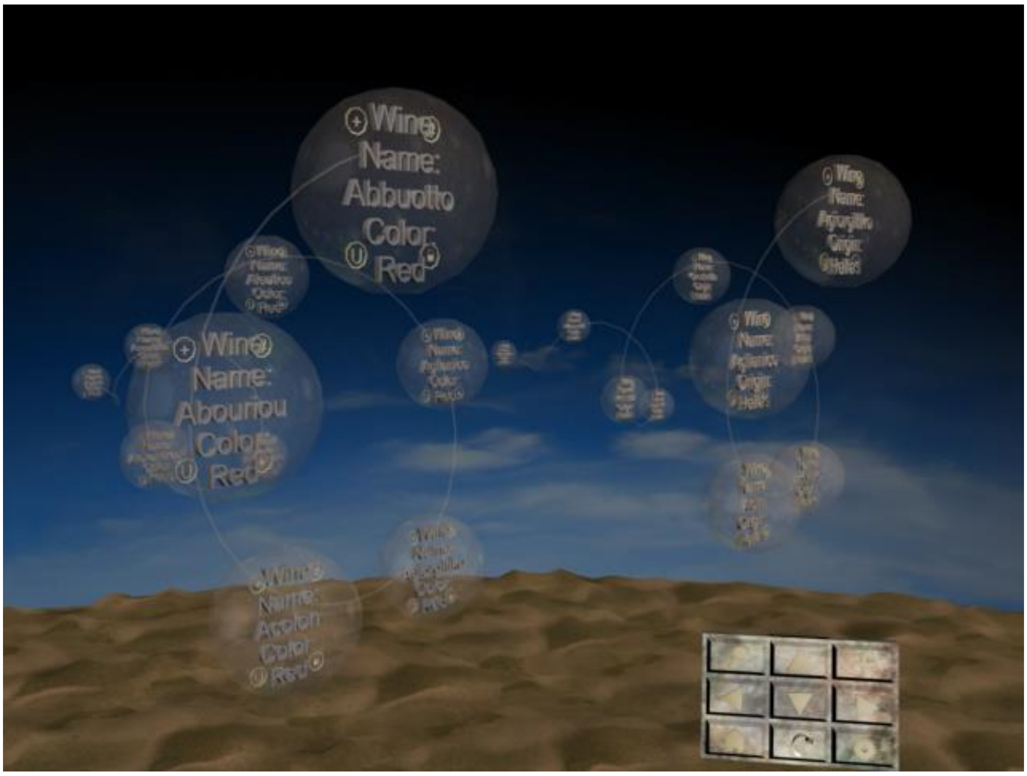



**Figure 52: We click the U (union) button on their top elements**

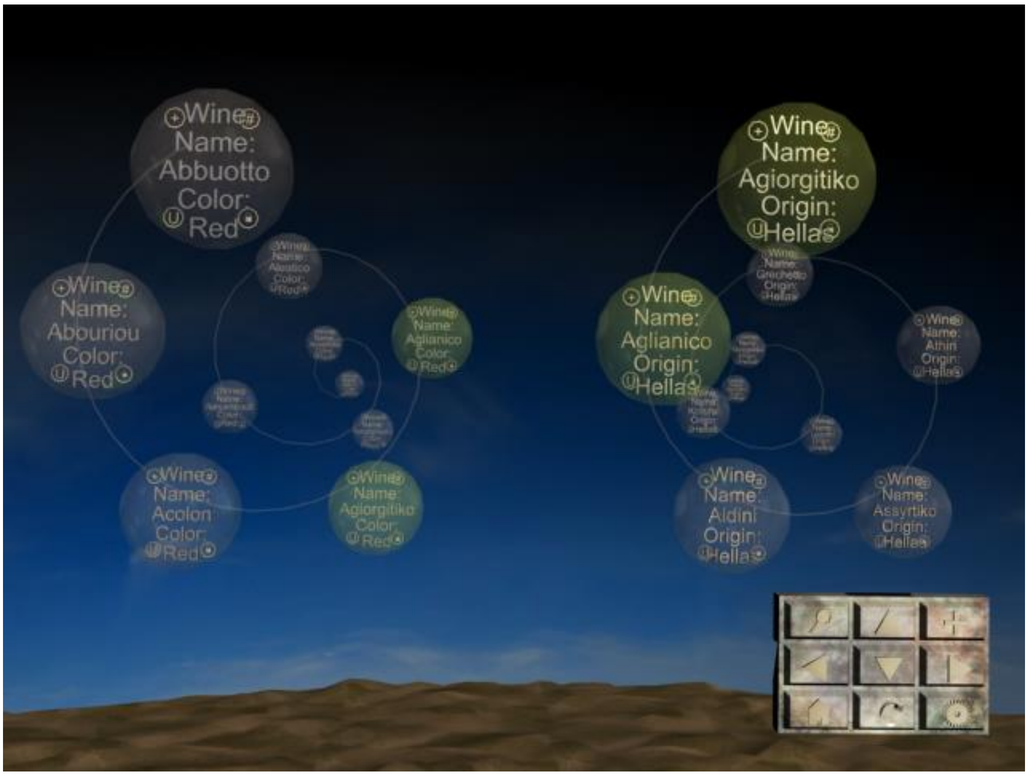



**Figure 53: Their common items are highlighted**

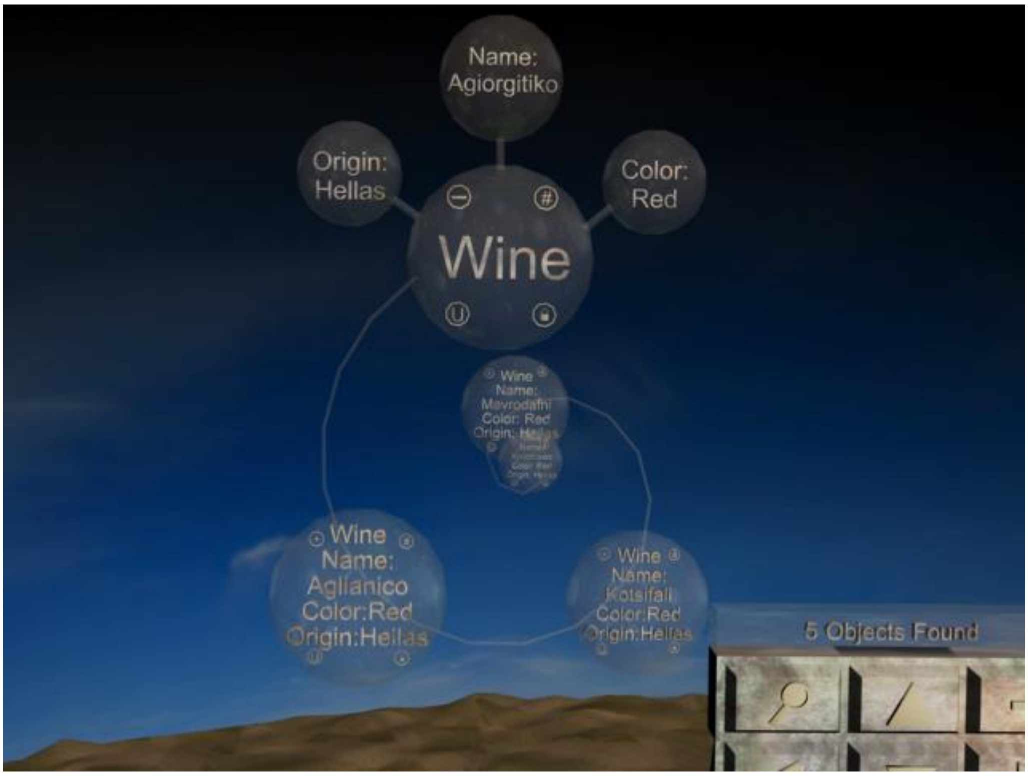



**Figure 54: We can extract their common items in a new spiral result-set**

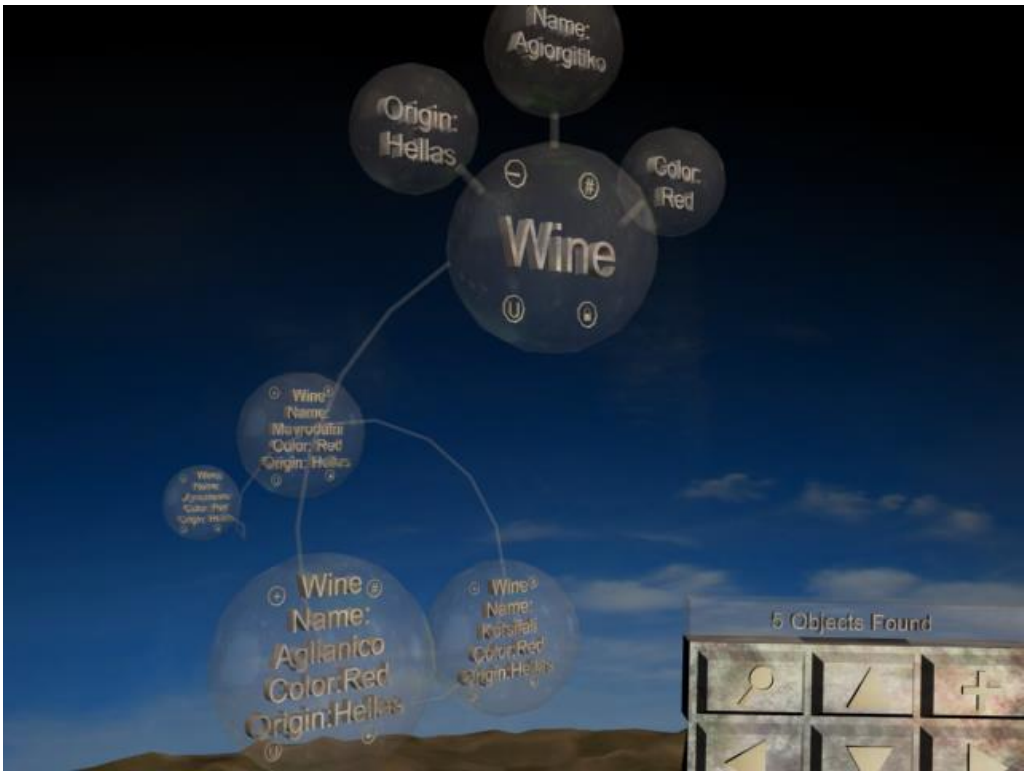



**Figure 55: Notice how the top item has been expanded**

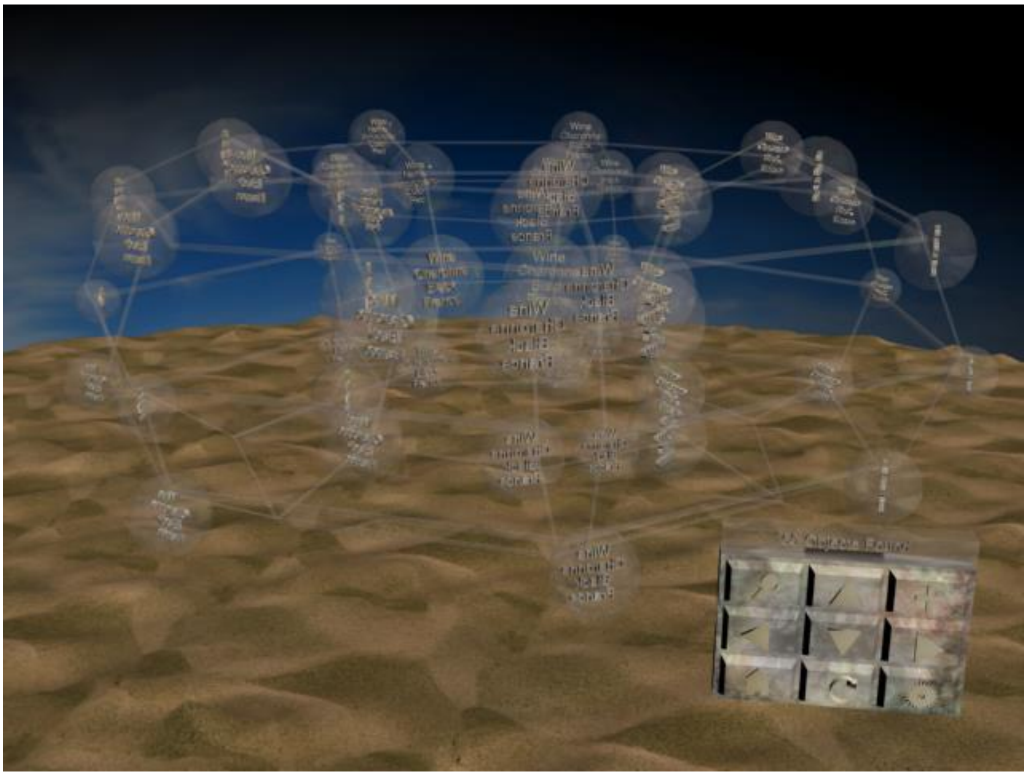


**Figure 56: Users may also select the torus result view**

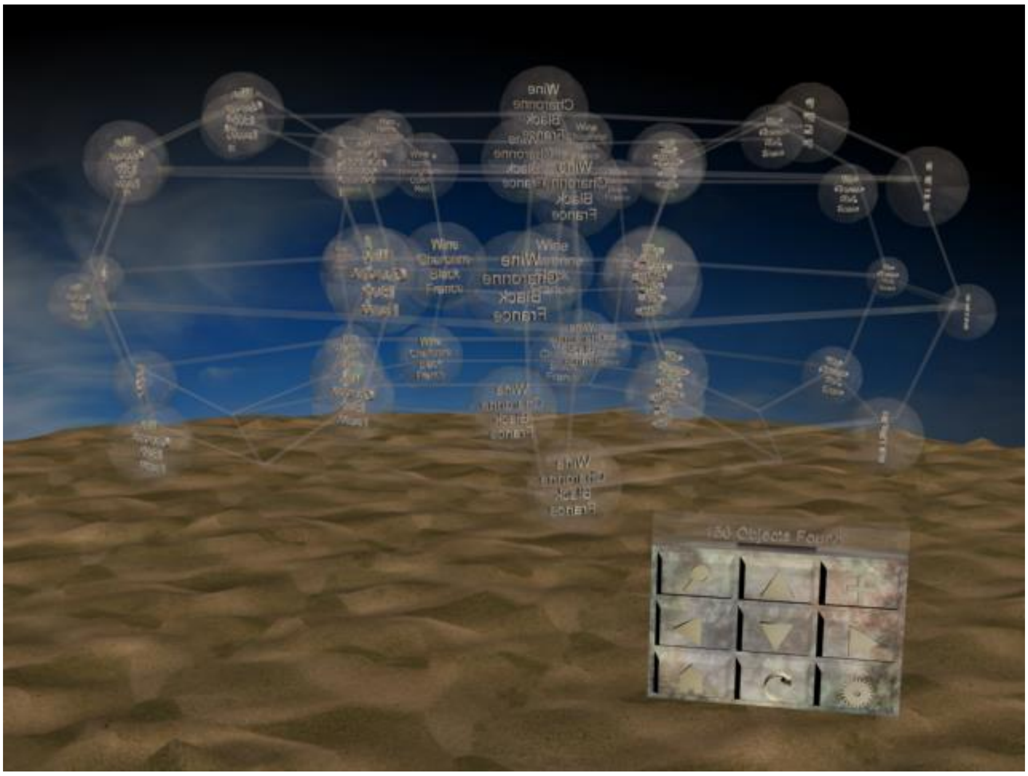


**Figure 57: It is a more complex view but it can easily show many results**

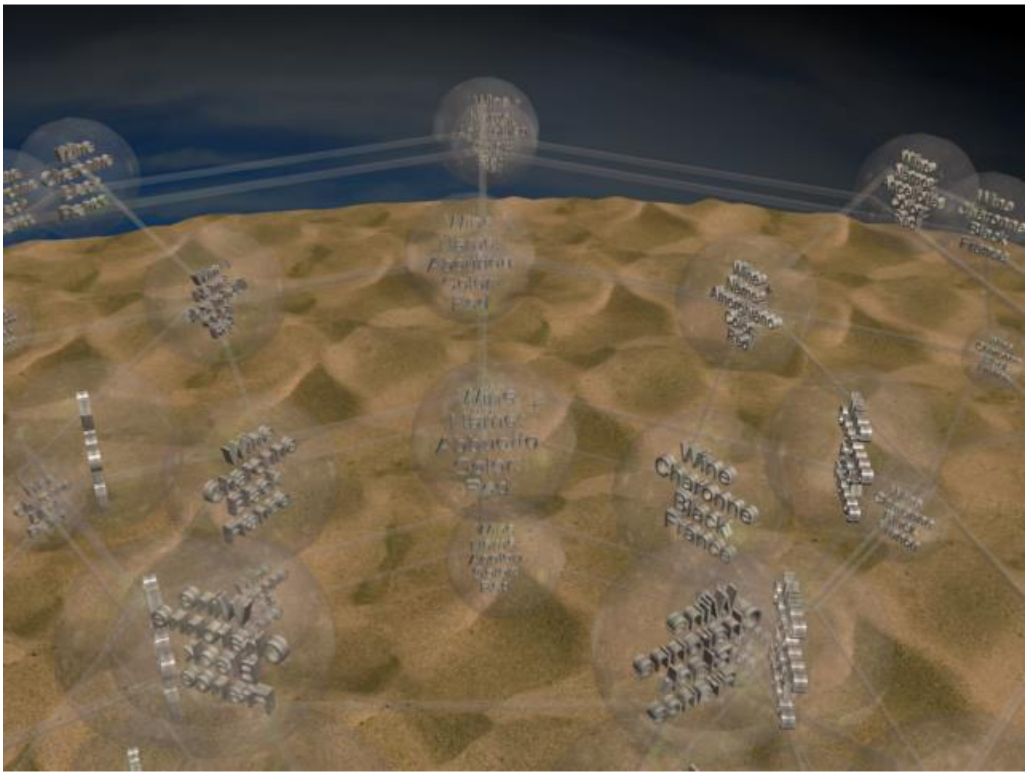



**Figure 58: A different camera angle for the torus result-set**


**Figure 59: A close-up on an item (axis showing)**

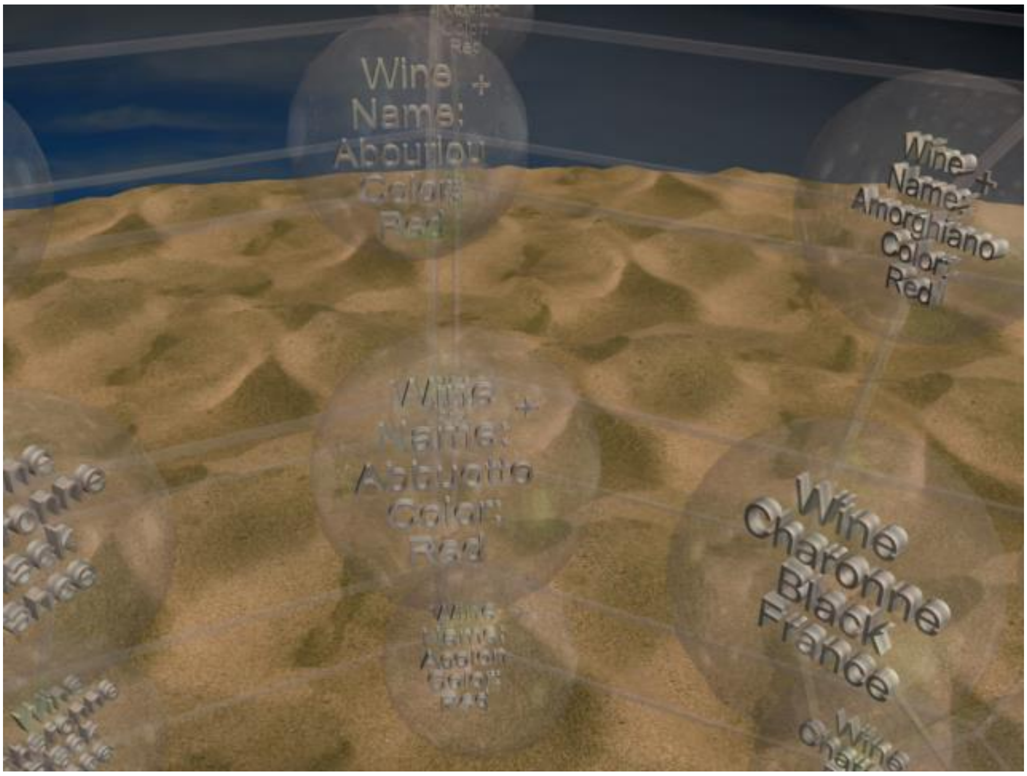



**Figure 60: Many items, side-by-side**

## *6.4 Future Integration with Metakosmos*

The search engine described in this thesis and the new ideas it introduces can be embedded into the Metakosmos system and launched on a commercial basis. A login and reputation mechanism has to be added in order to allow users to identify themselves, enter new syntheses and rate the quality of the content of the system. This mechanism can also include cross-site login, such as OpenID, ClaimID and BrowserID, thus allowing users to identify themselves as users of other websites and merging their information in a more centralised manner.

However, certain factors, mentioned in Section 4.5, have to be taken into consideration beforehand. After the initial legal issues are resolved and decided upon, the server and hard disk cost has to be examined. It should be noted that because of cloud storage technologies, hard disk space costs have dramatically decreased, dropping as low as $0,05 per month per 1 GB (Johnston, 2012) which will prove to be really useful for a system like Metakosmos.

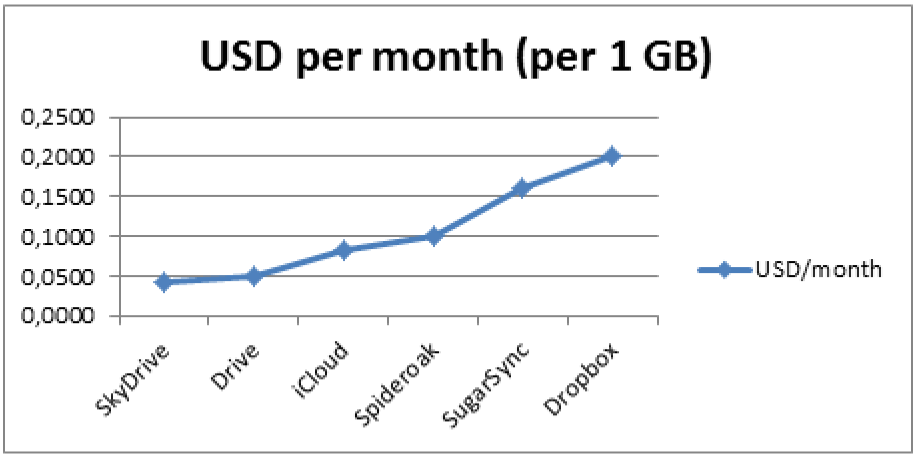



**Figure 61: Hard disk space costs per month per 1 GB**

Additionally, a thorough marketing strategy has to be implemented, aiming to attract different types of users of various backgrounds, in order to make the knowledge added to the system be as diverse as possible. Based on the user study that was carried out for the embedded item manipulation mechanism of Metakosmos, people are willing to try this new method of searching when they believe they will have a benefit from using it, like, for example, if they save time when looking for complicated queries (5.2.5).

The marketing strategy should stress how personalised the system can become and the amount of time it could save users for complex searches. It should also point out that it can be a “fun” system, because of the game-like ranking system (4.3) that offers a goal to the members of the system, and also the way Metakosmos learns from the actions of its users (4.5), attempting to speed up their searching.

Finally, the last suggestion of this thesis is the use of the “**mtk**” prefix, a virtual host subdomain identifier that will help differentiate between 2D pages (WWW) and 3D (MTK). Every website can create a new, X3D based, version of its synthesised content and include it in their own Web hosting space, under the mtk subdomain. Metakosmos will link to this content via the stored syntheses, which will act like indexes pointing to external Web links. In this way, the Web can gradually move from the 2D-only websites to a hybrid 2D/3D format and eventually convert entirely to 3D.

## *6.5 Summary*

In this chapter the development and implementation details have been identified and analysed, providing different possible scenarios and outcomes. Many X3D based screenshots have also been included, with more following in Appendix D.

# Chapter 7: General Discussion

## *7.1 Conclusions*

Based on the previous chapter, as well as the additional material attached in Appendix D, it can be concluded that, overall, the final result is a well-defined Semantic Web based Search Engine, which successfully adds a visual X3D based 3D environment on top of the old-style simple text-search.

With the help of the underlying Semantic Web ontology as a basis, as well as the introduction of the X3D graphics on top, users are able to query content using a visual/textual combination, maintaining the best parts of both methods. In this way, users manage to retrieve highly relevant results that directly answer the specific query they searched for. Users can also adjust their personalised settings, making result viewing faster, swifter and easier, based on their personal preferences and needs.

With the help of this graphical interface and the underlying ontology, the individual opinion of each user can be synthesised and connected to the rest of the system, possibly introducing new parameters and relations, or just aligning with the already existing information. This consolidation can be used to prove or disprove the validity of each synthesis based on the current data, while at the same time making available all the underpinning concepts and objects that the knowledge of the system is based on, thus supporting freedom of opinion and expression.

Day by day the “puzzle of knowledge” that this system is slowly trying to put together, by combining all the different perceptions of people, will attempt to allow us to take a peek into the revealed parts of the one, great, objective, truth. This truth will be gradually derived from all the individual subjective opinions and observations, and turned into a series of indisputable facts that will bring us true wisdom.

## *7.2 Scientific Contribution*

As originally mentioned in Section 1.3, as well as multiple times throughout the dissertation, four main scientific contributions were presented in this PhD thesis.

First of all, an additional visual searching layer was introduced, on top of Semantic Web ontology queries, meant to offer a graphical representation of the data stored in the system. The purpose of this extra layer has been to provide an alternative to a plain text search, by allowing users to manipulate information in the form of virtually “tangible” items. As presented in previous chapters, these items can be interconnected in order to generate search results or new additions to the ontologies of the system, which is achieved by the second contribution.

This new visual searching layer called for the creation of a new mechanism to add, search and retrieve graphical items by allowing users to compose visual queries. The new mechanism was entitled “Semantic Synthesis” and its purpose has been to provide for the seamless and unified visualisation and manipulation of search queries and results. The visual syntheses composed by this mechanism can be converted to computer instructions (e.g. SPARQL queries), which can be used to return a selection of filtered results from the ontologies of the system, based on the criteria provided by the users via their interface described in the third contribution.

Since this mechanism required users to handle visual items in order to operate, the customisable visual searching interfaces were introduced, set up based on the style and preferences of each unique member of the system. The system learns from and adapts to the searching habits of its users, personalising the popularity of items on a per user basis, thus suggesting different items to different people. In this way, members can set up their personal virtual “home lounge” with the items they use more frequently, and also retrieve more tailored item suggestions when composing new syntheses, based on their searching habits. These syntheses are then rated for their usefulness by other members, thus determining each user’s overall rank, as mentioned in the fourth contribution.

To maintain order in the system, a game-like user ranking scheme that rewards constructive system use was suggested. This scheme assigns appropriate percentage levels, titles and colours to each member, depending on the usefulness of their syntheses, as rated by the social community of the system. The rank of each user will define his or her access when adding new syntheses to the system, and will also affect the popularity of the compositions they have already inserted. The ranking scheme does not base its levels on whether the users “Like” or “Dislike” a synthesis, but on whether the compositions are useful (i.e. not spam) and semantically correct. This provides the members of the system with a goal: to make correct and constructive additions, thus helping Metakosmos produce accurate results.

## *7.3 Evaluation and Limitations*

Metakosmos, and its search engine described in this thesis, aim to gather and connect all the possible different human perceptions in one universal system, capable of showing where their differences and similarities lie, in order to promote healthy peer collaboration and the derivation of straightforward logical conclusions.

This unified representation of interconnected items does not flatten nor blindly merge all the contained knowledge of the system or the world. It merely shows the vast interrelations among the presented information items, while trying to capture, and perhaps combine, all the possible different interpretations of each concept and object that the synthesising users wish to express. In addition to the above, the 3D representation attempts to simplify the search process while at the same time avoiding excessively oversimplifying the procedure, because that could lead to stultifying its users.

Even though the underlying source code is not directly visible to the users, this does not mean the logic has to be hidden as well. The Semantic searching mechanism used by this system does not only speed up the procedure of finding specific information, but also reveals the reasoning behind its results for human users to see and understand. Computers are already faster than us; if we allow them to also become smarter than us by leaving all the thinking to them then, instead of promoting science and increasing our understanding of the kosmos, we will end up bringing the world one step closer to a technological singularity.

This attempt to interrelate the knowledge of the entire world – by storing, linking and graphically presenting all the different human perceptions of concepts and their instantiations – in a single system, should not be mistakenly compared to “rebuilding” the Tower of Babel (Genesis 11:1-9). The true aim of Metakosmos, and its visual Semantic Synthesis mechanism, is to help people successfully communicate their understanding of reality in a comparable and universal manner, in order to more easily match and combine corresponding pieces of information.

The system has been designed in such a way that it allows for the representation of all different opinions by freely permitting the combination and interconnection of different “things” in any imaginable way. Although it can be argued that this could produce chaos, with the help of the built-in rating system it will gradually become clearer what is, and what is not. In this way, by offering the necessary mechanisms for users to interrelate information objects as if they were puzzle pieces, the unified ontology of Metakosmos will be synthesised bit by bit in an attempt to “paint” the true picture of the kosmos as it actually is – rather than just how we perceive it.

Other limitations of the system have been mentioned in Section 3.2.3, focusing on the responsibility, credibility and objectivity of its members, who are the main source of information for Metakosmos. It is hoped that, with the help of the embedded ranking mechanism, the social community of the system will be able to depict all of the different opinions of its members, while educating non-constructive members against being destructive. In this way, the most accurate syntheses will gain popularity, even if there is more than one opinion (i.e. version) for the same issue. As long as the syntheses of the different views are well-defined, they can be easily compared and users can then derive their own conclusions.

Even though Metakosmos aims to be used by anyone, anywhere in the world, achieving its use by the entire population of the planet would be impossible, since not everyone is connected to the Internet or necessarily interested in using such a system. This could mean that certain data will probably not be included in the ontologies of the system, thus limiting the available offered information. The same rule, however, applies to every single Internet-based system in the world, regardless of the amounts of funding and advertisement backing it up. The only way to handle this reality is to attempt to attract as many users as possible, and to give them the capability of adding all their knowledge and ideas in a logical and comparable manner, hoping that, in this way, the lost knowledge will eventually be derived by the cooperating users of the system.

Funding will of course play an important role in supporting a system of this scale, as mentioned in Section 4.6.1. If the system fails to attract enough users, or if it manages to attract too many too soon, it will probably not manage to economically sustain itself. In order to attempt to avoid any such incidents, when the Metakosmos system is being prepared for commercial launch, an analytical financial plan has to be designed, aiming to cover the possible scenarios for different occurrences.

As time passes, ontology standardisation will play an increasingly significant role in information categorisation and exchange. Scientists will be able to compare their research with others faster and in great detail due to the specific nature of ontology definitions. All individuals can benefit from the advantages of Semantic annotation, for example while searching for music that sounds like their favourite band or for a movie to watch based on films they have enjoyed so far ('Jinni,' 2012). It is only a matter of presenting the idea of well-organised data to the Social Web communities in a useful and inviting way.

## *7.4 Future Work*

The Metakosmos system and the underlying visual search engine analysed in this thesis will be used as the basis for the creation of a 3D Semantic Web based Social Knowledge community.

The merging of the Social Web with the Semantic Web is an inevitable advancement which will lead to increased information organisation and universally useful data mining (Gkoutzis & Geroimenko, 2011). The addition of the 3D Web interface on top of the ontology management mechanism provides an extra level of simplicity in order to make the experience more user-friendly. Although keyword search still remains a part of this system, the proposed visual search environment is only a glimpse of what the future of the next generation Web holds.

Currently identified as future work is the creation of tutorial guides aiming to help users improve their understanding of the system, as well as the implementation of mechanisms to integrate Semantic Syntheses from other websites or to share local syntheses with external Web applications by exporting them in an appropriate form.

The first goal can occur by employing a group of professional tutors, trainers and academics, and asking them to generate educational content appropriate for many different ages, educational backgrounds and previous IT knowledge. These tutorials will be then offered to the users for “beta testing”, asking them to declare if their understanding of the system increased after using them, in order to determine the effectiveness of the created material.

The latter goal can occur via the use of special purpose Web Services working on either a “quid pro quo” basis, or a predefined fee. Additionally, digital data architects can create customised style themes that will act as templates for better looking user interfaces and improved presentation of the concept and object results.

Eventually, the Metakosmos system can go live and gain its own faithful community. Before this happens though, an expert legal team has to analyse the potential issues that may arise and find the appropriate methods of handling them before they occur. Wikipedia and Google can be used as examples of what happens to organisations handling large, international, datasets.

In addition to the conventional input methods described in this thesis, such as handling visual objects with the mouse and keyboard shortcuts, or typing in the required values of the various properties, the system can be further expanded to use alternative input and output methods.

With HTML5 revolutionising how we use Web browsers on all possible devices, we can soon attempt to use Metakosmos by wearing virtual cyber-gloves and 3D augmented reality glasses. Even blind people, who lack visual perception, might someday be able to use the system, with appropriate alterations in their refreshable Braille displays, allowing them to utilise the new ideas Metakosmos introduces.

The possibilities of input/output combinations are endless, with science fiction films such as Johnny Mnemonic and Minority Report paving the way for a new kind of future for the Internet, **where information is both free and organised**.

## *7.5 Summary*

In this final chapter, the thesis was summarised, the results were evaluated and the future work that can be done on this subject has been discussed. This concludes the thesis. Thank you for your time and attention.

K.

# Appendix A: Description of Methods (RDC.2 Report)

## *About the PhD work plan*

Up until now, during the MPhil stage, the need for an efficiently organised 3D search engine has been identified and justified. The theory behind this PhD thesis has been read, analysed and explained within Chapters 1 and 2. This includes the background philosophy, the breakthroughs which, step by step, lead to the technologies that will be used to build the final search engine, as well as the ultimate cause, which is to combine those technologies and create something new and useful.

In parallel to the above, the initial modelling of the system has begun which will form Chapters 3 and 4, containing the Analysis and Design of the system and its subsystems. It has become visible that a framework should be built, which will glue together the different components of this thesis, thus assisting in future expansions. Additionally, initial prototype applications will be built to test if the requirements of the analysis can be satisfied using specific technologies, both for the searching and the visualisation parts of the thesis.

The completion of the above will later form Chapters 5 and 6, which will show the implementation process of the search engine, as well as the personal milestones that lead to the final system. It will also include testing data and feedback on usability issues concerning the way the new engine works. The testing methods will heavily depend on the form that the new system will take and they have to be decided depending on the results of the analysis.

The last chapter, which will probably be Chapter 7, will contain an analytical evaluation of the system as well as ideas for future work which could be based on what has been created.

Additional chapters might be added, either by splitting existing chapters into more or in the case of identifying new needs. Finally, note that the current bibliography is a work in progress which will change many times while this thesis is written.

## *Architectural design of an efficiently organised 3D search engine*

Based on the aforementioned technological achievements, certain standards have been already decided upon. First of all, the main concern of this thesis is to analyse, design and implement a Web search engine which will efficiently organise the information it contains. In order to reach this objective, the capabilities of Semantic Web will be taken advantage of, which will provide for the needed constraints for successful categorisation.

In addition to this, a 3D interface will be built on top of the underlying mechanisms, to offer ease of use to the users as well as the capability of creating advanced queries, with the help of simple graphical tools, in a matter of seconds. The ideal way to accomplish this is by using X3D graphics, which are XML based, like the Semantic Web, and will provide uniformity and in-browser execution, thus ease of access and an open source approach.

All of the above will be joined together in a framework which will be called “**Metakosmos**”, from the Hellenic words “meta” which means “after” and “kosmos” which means “world” or “universe”. The meaning of the name derives from the purpose of the framework which is to create a graphical virtual world of organised information based on case-specific metadata. Metakosmos is placed in the Computer Science domain as depicted below.

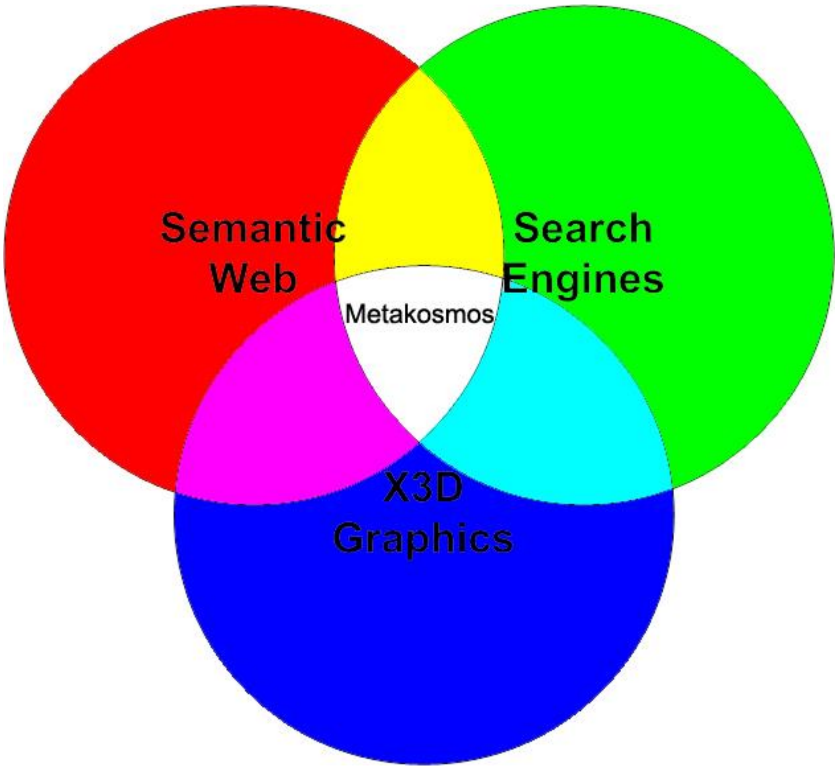



**Figure 1: Metakosmos, an intersection of three Computer Science fields**

It is important to note that a lot of work has been already done for the red (SW), green (SE) and blue (X3D) sections, and a little work has been done for the yellow (SW+SE), violet (SW+X3D) and aqua (SE+X3D) sections. No work, though, has been done for the white (SW+SE+X3D) section in the middle, which combines all three sections and is exactly the gap this thesis aims to fulfil.

This framework will also set certain standards which other developers will be able to follow so as to implement additional capabilities for the search engine. It will also reintroduce the term of "Semantic Synthesis", which is graphically presented below, with the help of a few 2D draft diagrams.

## *Consideration and evaluation of visualisation paradigms*

In order to evaluate this system, which will be offering a new way of searching through information on the Web, the visualisation component has to be decided upon first, since it will play the most drastic and decisive role for the entire thesis.

The Metakosmos framework will aim to provide simplicity to all users, whether they are computer amateurs or experts, as well as offer scalable complexity based on customisable options, while maintaining the same effective search experience despite of the different settings. To show an initial example, Figure 2 presents a 2D draft layout of a possible user interface, set to fulfil the needs of the specific user who customised it.

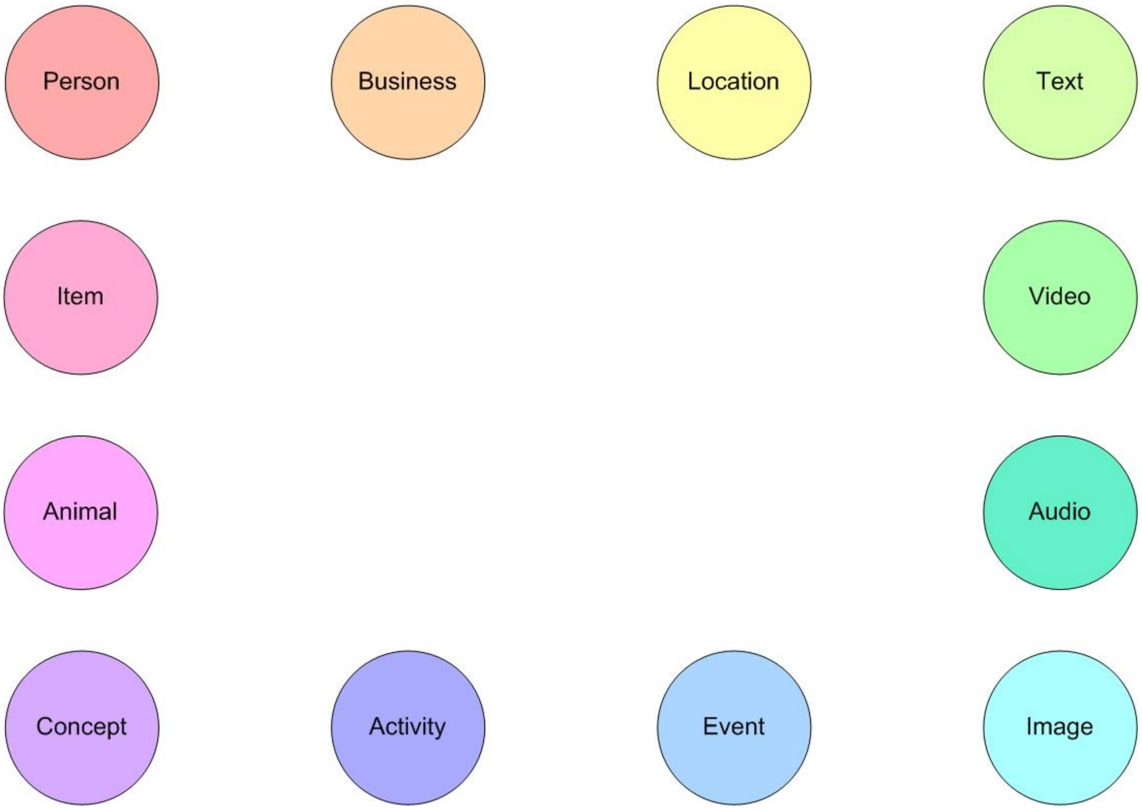



**Figure 2: A draft of a customised layout of the search engine**

This could be the search engine starting screen for a user who usually searches for these types of information and wants to have easy and quick access to them. Additional options will be available (not depicted in these figures), which will offer the addition of new objects on the interface, as well as other options which are to be finalised after the system analysis is completed.

In order to derive new information from the objects, we must first select specific objects and connect them to each other, thus graphically compose what we are looking for. This method will be called "**Semantic Synthesis**", from the Hellenic word "synthesis" which means "composition". An example connection of three objects is presented in Figure 3. More specifically we will be looking for a "Person" who is working for a "Business" in a specific "Location".

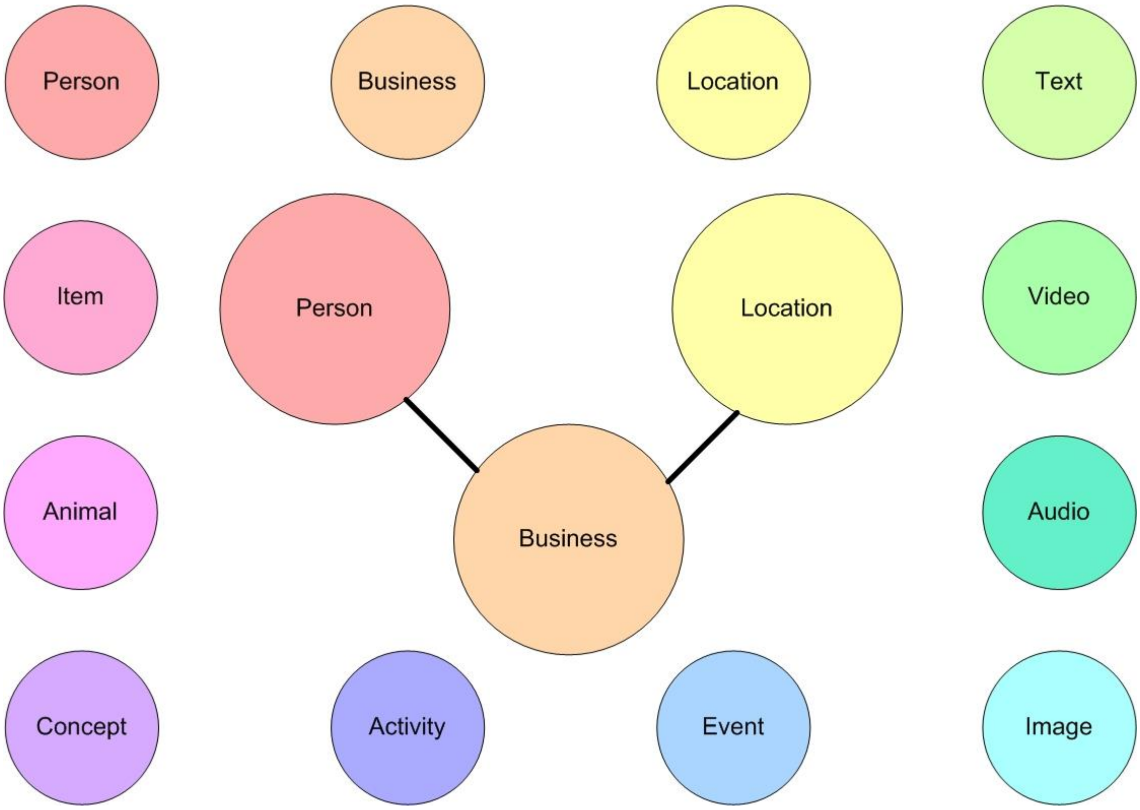



**Figure 3: Three objects selected and connected to form a query**

It is important to note that the term “Semantic Synthesis” was initially coined up by Igor' Mel'cuk in 1965 (Zolkovskij & Mel'cuk, 1965), for his Meaning Text Theory (MTT). It was the suggestion of linking words of different languages together, based on their meaning (semantically).

In this system, the objects will be semantically connected to compose (synthesise) specific queries. The associations between the objects will be assigned roles, based on a list of supported roles for each relation between two specific objects.

In addition to the connection roles, extra information may be provided for each selected object, in a menu offering a choice of specific properties available for each object. These roles and properties are depicted in Figure 4.

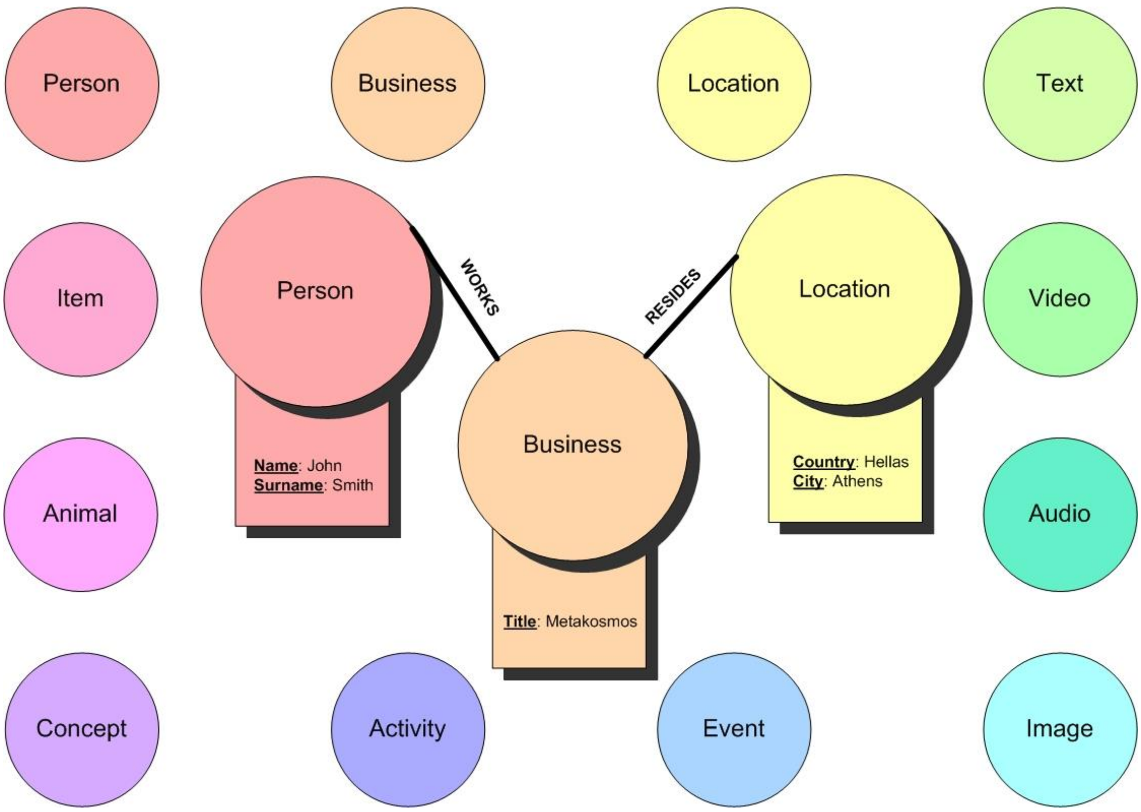



**Figure 4: Adding more details to the query**

It should be noted that with the help of modern technological advancements, users may now take advantage of many different forms of computer input devices, like keyboards, mice, touch-pads and even cyber-gloves, which are electronic gloves with computer control capabilities. In combination with high tech output devices like High Definition monitors and projectors or even 3D goggles, the entire graphical search experience can be drastically enhanced.

As soon as the query has been synthesised, an option (not depicted in these figures) will begin its process and finally provide the results in the same format as the query, which is interconnected objects based on their meaning. A possible result list of the previous query is presented in Figure 5.

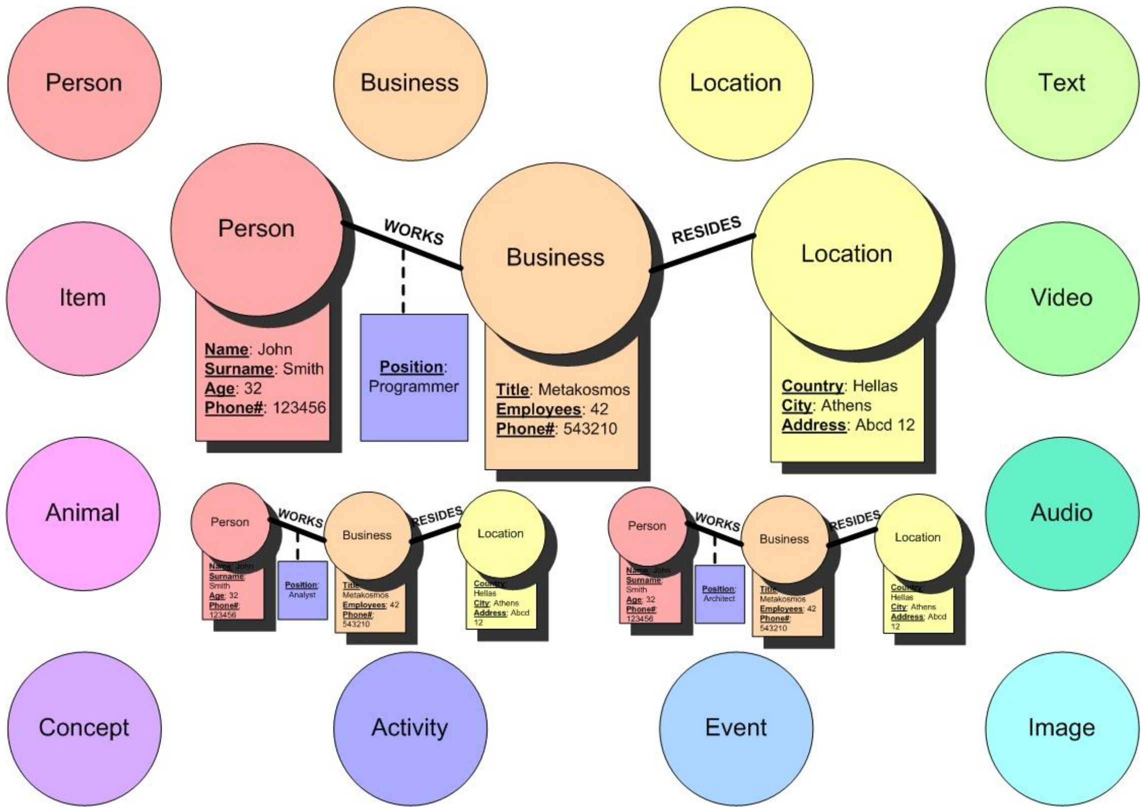



**Figure 5: The search query results, presented as interconnected objects**

The positive factor of showing results this way is that they can be selected and then further explored with additional queries. Specifically designed mechanisms will help the user quickly sift through the results, based on a set of browsing methods to be defined by the system analysis and then evaluated based on their performance for a variety of different queries. The whole search experience must be fast and always relative, which means that total time spent and the quality of results are the most important factors to measure.

Ultimately, the user may read the information provided by the search engine or explore external websites for additional data, like images, videos and so on. There are still a lot of decisions to be made about the end system, but the final aim has been visualised and explored and will be further developed in the continuing of this PhD thesis.

## *Summary*

During the MPhil phase, the need for an efficiently organised 3D search engine has been identified and justified, based on historical developments and breakthroughs of computer science, thus providing the appropriate foundation for further PhD research.

# Appendix B: IV 2011 Conference Publication

## Moving from Folksonomies to Taxonomies: Using the Social Web and 3D to Build an Unlimited Semantic Ontology

Konstantinos Gkoutzis, Vladimir Geroimenko
University of Plymouth
[ konstantinos.gkoutzis, vladimir.geroimenko ] @ plymouth.ac.uk

### Abstract

*The Semantic Web was introduced in 1999 as a method of interrelating information to help computers derive conclusions based on the links between data. With the rising popularity of Social Networks though, unconnected pieces of information have only chaotically increased instead of becoming parts of well-organised Taxonomies.*

*In this paper, we suggest a system which will take advantage of the Social Web and put it to work so that it will operate under the common cause of categorising old and new data into an unlimited Semantic Ontology. This ontology will be created gradually and ever-changing, like a versatile encyclopaedia of information compiled from interconnected data. With the addition of a 3D Web interface on top of the ontology management mechanisms, the entire experience will become more user friendly, providing graphical presentation of all opinions and interpretations in a clear and comparable manner.*

**Keywords**: Semantic Web, Social Networks, 3D, Information Visualisation, Knowledge Management.

### 1. Introduction

We live in the Information Era, where knowledge is everything and everywhere. The proper management of knowledge presupposes the existence of mechanisms for organising information, in order to match the appropriate pieces together. When the Internet came along, it offered users the ability of having access to many different types of electronic data. Unfortunately, these data are still mostly uncategorised and the human mind simply cannot absorb and process the increasingly huge amount of information available. This is the reason why computers are utilised to gather and present information in a more streamlined manner, so that people are able to browse through the findings in a faster and more convenient way.

XML and all the technologies that are based on it, like the Semantic Web [1], provide the facilities for electronic devices, even of different nature, to communicate with each other and exchange data in a commonly acceptable way. This is really useful, especially when it comes to making machines responsible for information gathering and the delivery of coherent results which a human can understand and rely upon. The Semantic Web is already trying to organise information into standardised structures called Ontologies [2]. This will gradually create a common ground for all topics, making information sharing easier and more automated.

Social Networks can benefit from the use of Semantic Ontologies to improve their simple, yet unsophisticated, method of tagging which is based on keywords rather than logical concepts. Adding Semantics may require more time than plain old word tagging, but the long-term gains can prove to be profitable in the quality of search results and in data categorisation.

With the number of 3D-based virtual communities increasing every day, both for Social and Gaming worlds, combining all the aforementioned technologies can result into a powerful online tool which will offer ease of use and meaningful search results [3].

## 2. Background and Related Work

At the original proposal of the World Wide Web in 1989 [4], most could not predict that it would become as widespread as it eventually did. Despite initial expectations, the Web has turned into a killer app of the Internet, gradually taking over the roles of other applications like E-mail, Usenet Newsgroups and IRC.

While the number of websites and webpages kept rising, it became impossible for a human to memorise all the available Web addresses or manually search through online data for specific information. That is the reason why Search Engines were invented, aiming to provide faster searching through a large amount of webpages based on requested keywords. After many attempts, and a variety of algorithms for page ranking, new age Search Engines are beginning to adopt the Semantic Ontology model to categorise their data in order to provide faster and more relevant results.

### 2.1. Search Engines and the Semantic Web

Computers are programmed by people to respond in a predefined way to specific actions. Artificial intelligence scientists have made several attempts to make computers come up with original ideas but, up to this point, it has not been achieved. Given the fact that computer applications can only follow the rules programmed into them (or the rules that those rules produce), they cannot yet cross the line from computing to improvising.

Since Search Engines are computer applications too, their search results are bound to be calculated and precise, based on mathematical formulae for text searching. This way, results may match the given words or phrases, but the meaning is totally lost. A search for the word “apple” will return many links to webpages and pictures of apple trees, as well as Apple computers, among a lot of other things. Researchers suggest [5] that in the circumstance of the results being too many, users usually go through the first few pages and then either accept what they found as the truth or just give up.

The most practical method to deal with this problem is to help the computer “understand” what each piece of information “means”, by using the relations between interconnected objects. The Semantic Web project is trying to address this issue by grouping information into Ontologies, which are logically organised datasets, a formal and well defined version of Taxonomies [6]. This means a user could specifically search for the fruit called apple, retrieving more relevant results. A few Semantic Web based Search Engines already exist [7, 8].

Computer users have grown accustomed to the habit of text search engines where they must type in keywords which they believe are related to their search. Those keywords are then matched to the index of the engine by using proprietary algorithms and a result list is produced, usually ordered by relevance or link popularity.

Semantic Web based search engines are trying to redefine this procedure. They may still be asking the user to type in text but they match those keywords to ontology items thus making results more logically relevant, rather than just plain keyword relevance [9]. This is feasible because ontologies organise items based on logical connections derived from the item meaning.

As time passes, ontology standardisation will play an increasingly significant role in information categorisation and exchange. Scientists will be able to compare their research with others faster and in great detail due to the specific nature of ontology definitions. All individuals can benefit from the advantages of Semantic annotation, for example while searching for music that sounds like their favourite band or for a movie to watch based on which films they have enjoyed so far [10]. It is only a matter of presenting the idea of well-organised data to the Social Web communities in a useful and inviting way.

### 2.2. The power of Social Networks

Soon after the Web became available to the public in the early 1990s, many users became interested into creating their own personal website to present themselves, their businesses or their interests. Certain companies begun giving away size-limited Web space for free on their servers and in exchange they placed advertisements on top of each webpage. This way, Web “citizens” began interacting with each other, linking their websites, exchanging virtual “awards” and comments in electronic “guestbooks”, as well as publishing their personal thoughts, ideas, pictures or even poems and passages. Separated from the commercial websites, the personal websites had a life of their own, either organised in communities or independently posting articles and news of interest to this new “global village”.

Near the end of the 1990s, users started calling their personal websites “blogs”, a contraction of the words “Web” and “log”, leading to the creation of a new trend in the Web culture [11]. This eventually lead to the creation of the first “online social networks” in the early 2000s, a term also popularised as “Web 2.0” by Tim O’Reilly of O’Reilly Media [12]. Social networks reused all the well-known technologies of the Web but, instead of offering free Web space, they prompted users to upload all their information and material inside a standardised personal profile page.

Friendster, MySpace, Facebook, LinkedIn and other well known Social Web giants became increasingly popular because they responded to the need of people to keep in touch with their friends or acquaintances and participate into activities together, like groups or games. Many companies, like YouTube, Flickr, Delicious and others, followed the Social paradigm by adding mechanisms to increase community participation and cooperation.

What all of the above websites have in common is that they categorise their data by using Folksonomies. This term was coined up around 2004 [13] from the words folks and taxonomy and basically defines the method of using keywords to describe the content of a data object. Blog articles, images, video and sound files stored all over the Web have been annotated using the words that uploaders chose to describe their data. Those words might or might not be related to the content, depending on the perception (or mood) of each uploader. It may be an extremely fast way to annotate data but, due to the lack of standards, many tags are vague, misspelled or simply wrong.

When the Social Web eventually merges with the Semantic Web, the users will be the first to benefit because, after the Semantics have been agreed upon, data sharing will be seamless and instant. It may take longer to annotate data than plain old keywords, but visualisation techniques can be used to speed up the process.

## 2.3. 3D environments and virtual worlds

With the rise of the computer age in mid 20th century, every type of business tried to get the best out of the new technological achievements. Even though at first computers only had text command interfaces, they eventually became able to present graphics. These graphics were initially plain, but as computer capabilities increased, so did the complexity of the designs.

The computer game industry has taken advantage of 3D graphics and moved from classic point-and-click games to 3D gaming. From first-person shooters to massively multiplayer online role-playing games, 3D graphics improve the gaming experience by offering a more realistic gameplay to the users.

With the Social Networking niche in mind, companies decided to combine 3D gaming environments with online communities, such as Second Life, Active Worlds and IMVU, among others. These programs, though, require the users to download extra Windows applications, which execute separately from the Web browser in order to access the 3D environment. This fact not only prevents application mobility but also isolates the actual program from the Web realm, leaving only the account management on the company website.

A notable attempt to combine browser based 3D graphics with a Social Network was Google “Lively”, which was launched, popularised and shut down in 2008. Lively used Flash, as well as a proprietary plug-in, in order to execute inside the browser and the final result was a fully in-browser 3D experience. Despite its initial success, Google decided to discontinue “Lively” shortly after its inauguration to focus more on their core search [14]. A similar Flash application called “Smeet” still exists, created by a German company.

As far as 3D search engines are concerned, there have only been a few attempts. In 2005, a company called “INOZON” announced [15] that they would be creating the first 3D based search engine, running inside a browser, but eventually the project fell apart. Another example was “Ergo” by Invu, launched as beta in 2007, running in an application outside the browser, offering a visual environment for searching information [16].

Although all these attempts are recent, a Web language to express 3D graphics started to be formed as early as 1994 and was called VRML (Virtual Reality Modelling Language). Since that time, it managed to become a standard and reached version 2.0, before being succeeded by X3D.

X3D is an XML based language which aims to popularise a file format to display 3D graphics by using XML syntax [17]. This makes 3D graphics ideal to parse and present inside browsers or to use with any type of XML related API. At this moment, though, X3D is not widely used and its commercial utilisation is limited. Also, Web browsers require extra add-ons to display the X3D environment. Being standardised with HTML5, the latest version of HTML, will help this format to gradually gain more recognition, while the X3DOM [18] project is already trying to bypass the need of a plug-in.

HTML5 will also bring instant embedded support of WebGL, a language based on JavaScript which offers in-browser interactive 3D graphics, as well as CSS3 which will also offer 3D transformations. The basic aim of HTML5 is to be able to run in all types of devices and support the creation of Web applications without the need of extra plug-ins. This will bring the Web to an entirely new level.

## 3. The future Web: Social, Semantic and 3D

What the Web needs now is a new idea, to make users feel that all these advancements can indeed improve the current status quo.

Introducing 3D visualisation techniques to the Social Web would be a start but by itself it would only appear like a “cool” looking improvement of the graphics. There has to be essential change in the way we perceive data annotation and information acquisition. The Social community is a power strong enough to fill websites with a plethora of multimedia files gathered in many different ways. The users only need to become accustomed to a new method of tagging data, in order to make them part of a Semantic Ontology.

This is why we propose the creation of a system that will have the properties of the Social Web, but with a 3D Web interface on top and a Semantic Ontology behind it. It will be a new age Web application, merging the aforementioned technologies into a 3D Social community, which will constantly be improving an underlying, unlimited, Semantic Ontology.

### 3.1. Creating an information driven community

The proposed system will have a search engine which will not only check indexed keywords for similarities to a given input, but will also classify data to specific categories of an unlimited catalogue. This catalogue will be created gradually as users add new multimedia objects or change the already existing ones. Some organisations, such as Wikipedia, Dmoz or Open Source Software communities, have managed to maintain quite a high level of data quality, even though they are not always edited by professionals, but by earnest individuals who dedicate their time to adding and correcting the website material.

The users of this system will enjoy the advantages of a 3D Social community, such as 3D avatars and personalised virtual “lounges”, and at the same time will contribute to the creation of an infinite and global source of knowledge. This way, not only they will be able to participate into Social Networking activities but they will also play an important role into rating the quality of the information available in the system by voting up or down additions and changes based on the correctness of the submission.

This participation will make the Web community more active and raise awareness against spam and phishing attempts because data will be immediately comparable due to the nature of ontology objects. Users will be required to make all changes and additions using their real name which will radically reduce the cases of “trolling” (i.e. uploading inflammatory or off-topic data).

Using an overall reputation based scheme, the level of user access to the system will be altered depending on how their additions and changes are ranked. Thus, material which contains mistakes or is miscategorised will be voted down and gradually replaced by correct ones, while at the same time the overall reputation status of the uploader will decrease, restricting their access to website features.

### 3.2. Making data management user friendly

The most interesting factor of the system is that the users do not have to possess any type of Semantic Web or XML knowledge in order to use it. The 3D Web interface will help them to add, change or search for specific website objects by creating new compositions. In order to manage or look for data, users will select 3D objects and connect them to each other, thus graphically compose what they are adding or seeking.

An example connection of three objects is presented in Figure 1. More specifically it has been composed to look for a “Person” who is working for a “Business” in a specific “Location”.

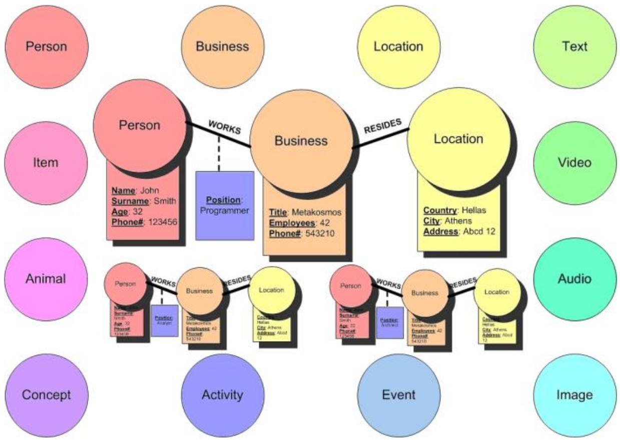



**Figure 1: Draft layout of a query in a custom user interface**

We call this method “Semantic Synthesis”, from the Hellenic word “synthesis” which means “composition”. It is interesting to note that the term “Semantic Synthesis” was initially used by Igor' Mel'cuk in 1965 [19], for his Meaning Text Theory which was the suggestion of linking words of different languages together, based on their meaning (i.e. semantically). In our system the objects will be semantically connected to compose (synthesise) specific queries or additions. The associations between the objects will be assigned roles, based on a list of supported role objects for each relation.

In addition to the connection roles, extra information may be provided for each selected item, in a menu offering a choice of specific properties available for each object. As soon as the query or addition has been synthesised, there will be an option to either add it or search for it. Figure 1 also shows possible search results below the synthesised query.

The positive factor of showing results this way is that they can be selected and then further explored with additional queries. Specifically designed mechanisms will help the user quickly sift through the results. Some are currently being evaluated based on their performance for a variety of different queries. The whole search experience must be fast and always produce relevant results which means that the total time spent and the quality of the results are the most important factors.

To create these mechanisms, information must be viewed not only as words and phrases, but as objects which are all parts of the same unified system. That is where the 3D Web interface will come in handy, to give users a look and feel of actually connecting objects.

Ultimately, the user may search through the information provided by our system and then explore links to external websites for extra references, such as images, videos and additional, not yet stored, material.

### 3.3. Social Semantics drawbacks

Web citizens have been tagging online data sources for many years now. Apart from the obvious powerful virtual workforce of collective intelligence they comprise [20, 21], there are also certain negative aspects to it. The current drawbacks of Social Semantics are summed up by three basic factors of human nature: responsibility, credibility and objectivity.

When people publish data on the Internet, they do not always take the time to label and organise them based on the existing standards for each category, thus making their meaning vague for a computer. This lack of responsibility can be avoided by utilising predefined methods for adding and manipulating objects so that they carry at least some basic annotation, which will be aided by using visual objects and Semantic Synthesis.

Even if this problem is surpassed, no one can guarantee that the categories and relations the user has selected are appropriate for that object, because not everyone is a field expert on everything they post online. That is why users will have the ability of voting additions up or down, which will affect the overall reputation of the uploader appropriately.

Finally, the biggest problem of all is objectivity. Even for an organised consortium of scientists and field experts it would be difficult to agree on a common methodology for characterising every possible piece of information. The only way to prevent the debates from reaching a total deadlock is to try and present all opinions in a very specific and comparable way, so that each researcher can then determine what applies to their specific case.

Philosophy and Sociology will play an important role in the definition of most data categories (e.g. ideas and concepts) which are unsubstantial by nature. Gradually, all views and suggestions will be synthesised and depicted, allowing users and researchers to compare them side by side in a more streamlined manner.

## 4. Metakosmos

All of the above will be joined together in a system called "Metakosmos", from the Hellenic words "meta", which means "after", and "kosmos", which means "world" or "universe". The name reflects the ultimate purpose of the system, which is to create a graphical virtual world of organised information about the entire real world.

### 4.1. Social Knowledge Management

Metakosmos will be a 3D Social Web community that will store its information by taking advantage of Semantic Web Ontologies. The system will have two types of users: unregistered (visitors) and registered (members). Visitors will be only allowed to use the search engine, without being able to adjust the user interface or participate to the addition, categorisation or correction of data. Members will be able to use the website to its full extend.

Additionally, the members will be able to edit their user interface on the website by choosing which items should appear at their home screen, also known as their "lounge". They will select those items by searching through the available categories offered by the website or by adding new ones. They will also be able to add new items to the system, rearrange existing data or even correct mistakes that may exist such as, for example, object miscategorisation or misspelling.

Collecting consistent data is essential for any search engine, especially for one that needs to store them inside an ontology. Members will have to correctly synthesise the data, based on their understanding, in order to add them to the system in the appropriate categories. Other members with better understanding or expertise on a domain will be able to suggest corrections, if needed.

Members that offer a lot to the website, either by adding new material or by correcting existing items, will receive an increased reputation status which will allow them to influence the system faster. Clearly, members that do not cooperate respectfully will receive a decrease of their reputation status, giving them less and less access in influencing the system.

All the user interaction will bring out the social aspect of the system, uniting the members under the common cause of keeping everything organised and as accurate as possible. The additions or changes which are well defined and correctly categorised will be voted up by the members, thus increasing the overall reputation status of the person who synthesised them. In contrast to this, constant mistakes and sloppiness will cause members to vote items down, decreasing the reputation of the person responsible.

Five reputation levels will be used: Ignored (0%), Trainee (25%), Notable (50%), Popular (75%) and Perfect (100%). Newly registered users will start as Trainees and then, depending on how well they adjust, their reputation will increase or drop. If their additions or changes are constantly voted down they will eventually reach the Ignored status which would mean that their new additions will be automatically pre-voted down and they will also be unable to make any changes to the system.

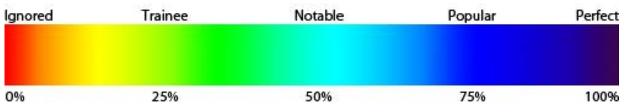



**Figure 2: The five suggested reputation levels**

The Metakosmos system will aim to provide simplicity to all users, whether they are technology amateurs or experts, as well as offer scalable complexity based on customisable options, while maintaining the same effective search experience despite of the settings.

## 4.2. Synthesising, integrating and sharing

By reintroducing the term of "Semantic Synthesis", users will be called to select predefined objects, connect them in a way they consider meaningful, add literal parameter information and finally search for matching results stored in the system. Additionally, registered members will be able to synthesise additions or any needed changes to the system. All of the syntheses will be Semantic Web based, even though the website users will only see the 3D visualisation representation.

Metakosmos will aim to set certain standards which other developers will be able to follow so as to interact externally with the system or to implement additional capabilities for the user interface. Currently set as future work, is creating the mechanisms to integrate Semantic Syntheses from other websites, as well as to share local syntheses with external Web applications by exporting them in an appropriate form.

As time passes, more and more websites will appreciate the advantages of using Semantic Ontologies, making data sharing and integration simpler and instant. Semantic annotation will gradually replace the keyword cloud tagging of Folksonomies, allowing increased access to data of all types and formats.

## 4.3. Adding a 3D User Interface

In order to make Semantic annotation easier and more straightforward, 3D visualisation elements will be used to depict the various Semantic objects that comprise each synthesis. Usability will be an important factor of this new age 3D community, in order for it to get accepted by the general public and integrated into everyday life. It has to combine simplicity with scalable complexity, based on custom options, while maintaining the same effective search experience for all users.

To achieve this, an extra layer of graphics is added on top of text searching, making search queries visually accurate and, as a result, retrieving more relevant search results. With the help of the underlying Semantic Web ontology, as well as with the introduction of 3D graphics, users will be able to query content both textually and visually, depending on their personal preferences and the nature of their search. This way, they will be able to retrieve highly relevant results, which can be presented in many different forms.

With the upcoming official arrival of HTML5 [22] and the constantly increasing improvements of Web 3D graphics like X3D, X3DOM, WebGL and CSS3, the Metakosmos system will obtain its final form and gain its place among the first 3D Social Web communities, while at the same time participate in the storing and categorisation of global knowledge in a large scale.

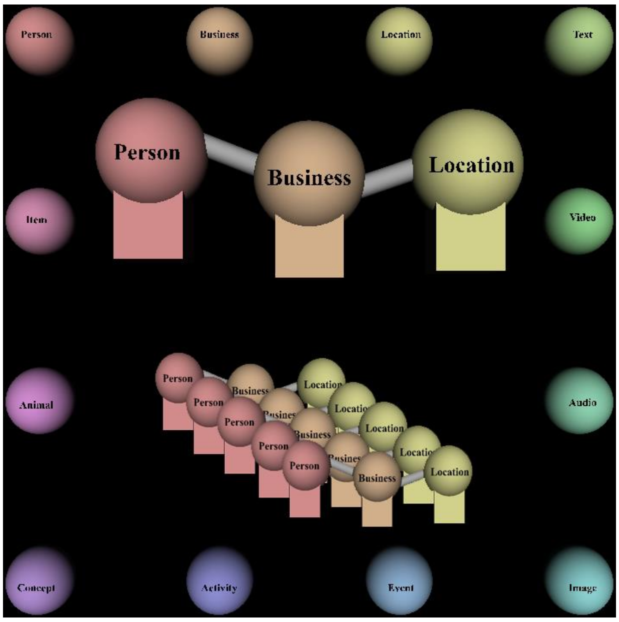



**Figure 3: Mock-up user interface compiled in X3D**

## Conclusions and Future work

When the Metakosmos PhD thesis is complete, the first prototype will be launched online in order to offer users a preliminary beta testing of the system.

The initial version will be a visually searchable and updatable catalogue which will aim to produce useful and relevant results as quickly and as accurately as possible. This means minimising unsolicited advertisements, preventing misinformation and yet presenting all possible views of each subject, both concisely and thoroughly, based on the personal preferences each user sets. Gradually, more and more users and websites will be able to create and exchange their information syntheses and come up with ways to merge and unify new and old data as efficiently as possible.

Eventually, new functionalities will be added to the system which will give members increased access in affecting their user interface themes, thus making searching and editing further adjustable. Additionally, Web Services will be launched to make information integration and sharing more automated for external applications.

The merging of the Social Web with the Semantic Web is an inevitable advancement which will lead to increased information organisation and universally useful data mining. The addition of the 3D Web interface on top of the ontology management mechanisms provides an extra level of simplicity in order to make the experience more user-friendly. Although keyword search will be also available, the proposed visual search environment is only a glimpse of what the future of the Web holds.

## References

[1] T. Berners-Lee, et al., “The Semantic Web,” Scientific American, vol. 279, 2001.

[2] H.P. Alesso and C.F. Smith, Thinking on the Web: Berners-Lee, Gödel & Turing, John Wiley & Sons, 2006.

[3] V. Geroimenko and C. Chen, Visualizing the Semantic Web: XML-based Internet and information visualization, Springer, 2006

[4] World Wide Web CERN Proposal: http://www.w3.org/History/1989/proposal.html

[5] iProspect, “Search Engine User Attitudes,” 2004; http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf.

[6] W. Pidcock and M. Uschold, “What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?,” 2003; http://www.metamodel.com/article.php?story=20030115211223271.

[7] Swoogle: http://swoogle.umbc.edu/

[8] Hakia: http://www.hakia.com/

[9] D. Li, et al., “Swoogle: a search and metadata engine for the semantic web,” ACM, 2004

[10] Jinni: http://www.jinni.com/

[11] J. Wortham, “After 10 Years of Blogs, the Future's Brighter Than Ever,” 2007; http://www.wired.com/entertainment/theweb/news/2007/12/blog_anniversary.

[12] T. O'Reilly, “What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software,” http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html, 2005

[13] Folksonomy: http://vanderwal.net/folksonomy.html

[14] Lively: http://googleblog.blogspot.com/2008/11/lively-no-more.html

[15] INOZON: http://prweb.com/releases/2005/02/prweb211778.htm

[16] Invu: http://www.invu.net/

[17] V. Geroimenko and C. Chen, Visualizing information using SVG and X3D : XML-based technologies for the XML-based Web, Springer, 2005, p. xiv, 298 p.

[18] X3DOM: http://www.x3dom.org/

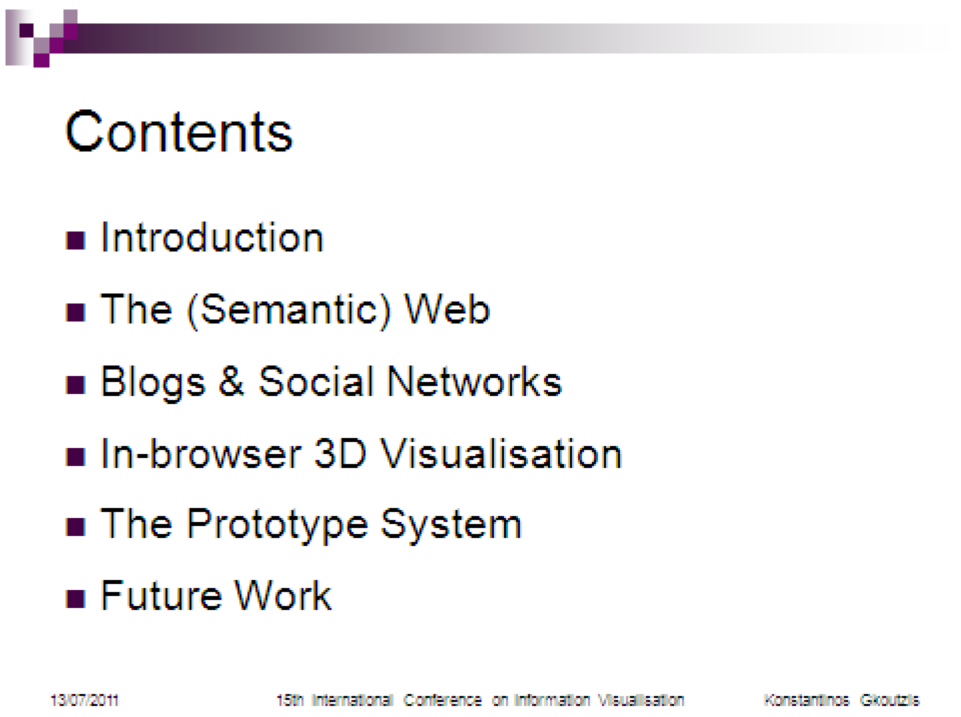
[19] A.K. Zolkovskij and I.A. Mel'cuk, “On a possible method and instruments for semantic synthesis of Texts,” Nauchno-texnicheskaja informacija, vol. 1, no. 6, 1965

[20] H. Halpin, et al., “The complex dynamics of collaborative tagging,” ACM, 2007, pp. 211-220.

[21] V. Robu, et al., “Emergence of consensus and shared vocabularies in collaborative tagging systems,” ACM Trans. Web, vol. 3, no. 4, 2009, pp. 1-34; DOI 10.1145/1594173.1594176.

[22] HTML5: http://dev.w3.org/html5/spec/spec.html

# Appendix C: IV 2011 Conference Presentation

## Introduction

- This is the first publication of my PhD thesis on XML Technologies & the Semantic Web
- The final goal is the creation of a system able to store data in an organised manner
- **Data** with meaning will give us **Information** which can eventually lead to **Knowledge**

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## A Brief History of the Web

- The Web was invented in 1989 by Sir Tim-Berners Lee and soon became a killer app of the **Internet**
- New websites are popping up everyday, containing new (and old) information
- Search Engines are used to find specific pieces of text or other multimedia material, instead of manually checking all websites one by one

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# A Brief History of the Web (2)

- Services like E-mail, Usenet Newsgroups, IRC and FTP are gradually integrated into the Web
- Websites like Wikipedia offer a vast amount of multimedia information, which is freely available (and freely editable)
- Online catalogues like the Open Directory Project (dmoz) are trying to organise websites based on their **content**

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# Adding the "Semantic" Layer

- W3C's Semantic Web project is trying to provide the mechanisms for organising data into **Ontologies**
- Coined up by Sir Tim Berners-Lee in 1999, it describes a Web where all data are interlinked based on their **meaning** and machines can derive better search results

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# The Ontology Concept

- An **ontology** is one type of a **taxonomy**, offering well-classified pieces of data
- Ontologies are standardised representations of knowledge within a domain
- For example, an ontology about wines can describe all existing wines


13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# The "Web 2.0" Trend

- In early 2000s, a certain trend became obvious on the Web, which Tim O'Reilly called "Web 2.0"
- Web users started creating blogs (web + log) to share their thoughts and their work with others
- The users added labels (tags) to their posts, in order to label the content of each entry
- These "clouds" of tags were eventually called **folksonomies** (folks + taxonomy), even though they are not really, by definition, taxonomies

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## The Rise of Social Networks

- Another big chapter of the “Web 2.0” trend concerns the Social Networking communities
- These websites connect people to friends (or strangers) and help them share and exchange data in many different ways
- The tagging paradigm can be found here as well, because users were already accustomed to it from blogs

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## Semantics VS Tagging

- The problem with plain “word tagging” is the lack of actual **meaning**
- A news article about apple, the fruit, will be returned together with an article about Apple, the company, if the search is based on a case-insensitive “apple” word tag
- Categorising data **Semantically** would prevent this from happening because a fruit is not the same as a company

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

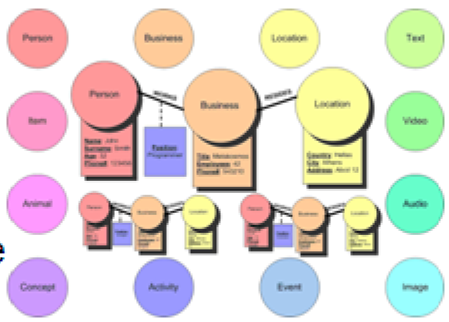
## The Need for a 3D Interface

- Tagging data by just using words is very easy and straightforward, but not meaningful
- In order for users to move from tagging to Semantic annotating, an easy-to-use interface has to be built
- Adding Semantics should also be easy to do since it is based on reason and logic
- Modern users are accustomed to **gaming environments** and **3D worlds** as well as the "realism" provided by those interfaces, simulating our surrounding 3D world

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## A Different Way of Searching

- Using the Semantic method of perceiving search (**Subject – Predicate – Object**), we enter a new way of visualising queries and results




- We call this technique **Semantic Synthesis**

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

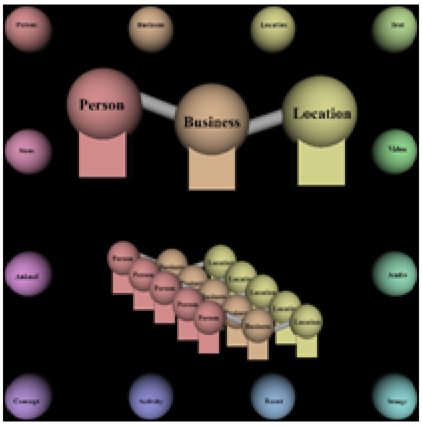
# In-browser 3D Visualisation

- In order to add a 3D environment inside the Web browser, specific languages have been designed
- **X3D** is an XML based standard from W3C, offering 3D graphics described in pure XML
- **WebGL** is a JavaScript based 3D framework which has also gained popularity
- **HTML5**, the upcoming version of HTML, combined with **CSS3**, can also produce in-browser (cross-platform!) 3D results, although both standards are still in draft format

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# Putting it All Together

- A Social, Semantic Web based system with 3D visual query composition and results output
- We call this "**Metakosmos**"




13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

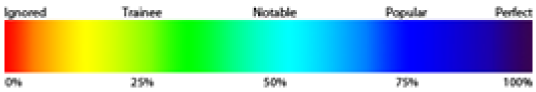
## The Metakosmos System

- Metadata are data about data
- Metakosmos is a “virtual world” about the “real world”
- It will be a 3D Social Web community that will organise its information by taking advantage of Semantic Web Ontologies

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## The Metakosmos System (2)

- Because certain users are not always well-intentioned (*or cannot admit they are not an expert in everything*), a custom rating system will be used, reducing or increasing user access to additions or changes




- The rating system will be handled by the Registered Users and the Administrators will manage special cases of system abuse

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## The Metakosmos System (3)

- Finally, the "ultimate" goal of the system is to organise all information in a **global catalogue**
- The greatest challenge, of course, is agreeing on common standards
- Philosophy and Sociology will play an important role in the definition of most data categories
- Gradually, all views and suggestions will be depicted, allowing users to compare them in a more streamlined manner

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## Future Work

- Within the next year, the initial prototype of "Metakosmos" will be ready to go live for **beta testing**
- A lot of programming still needs to be done (but we are getting there!)
- A really important issue is to give users the ability to fully **customise** their interface
- In the end, the "Semantic Synthesis" method will be evaluated and readjusted based on the results

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## Summary

- Combining the power of Social Networks with the Semantic Web can create a global, ever-changing and ever-expanding, organised catalogue of information
- Using a user friendly 3D interface, inside the Web browser, can make the Semantic annotation concept simpler, more accessible and more realistic

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

## Contact Information

- To share your ideas and suggestions, please contact me at:
  - konstantinos.gkoutzis@plymouth.ac.uk
  - kgk@kgk.gr
  - http://twitter.com/kgkoutzis
- Thank you for your attention :)

13/07/2011 15th International Conference on Information Visualisation Konstantinos Gkoutzis

# Appendix D: Additional Screenshots

This appendix includes additional screenshots from the X3D interface of the system. More screenshots can be found in Chapter 6.

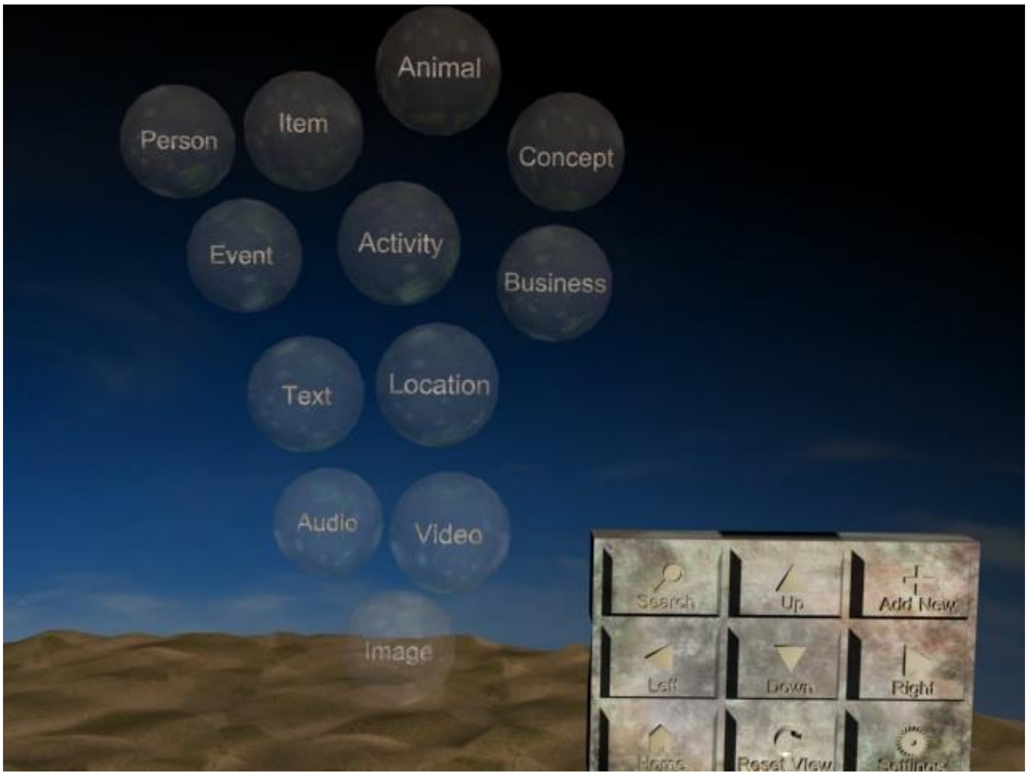



**Figure 62: The interface items rearranged by the user**

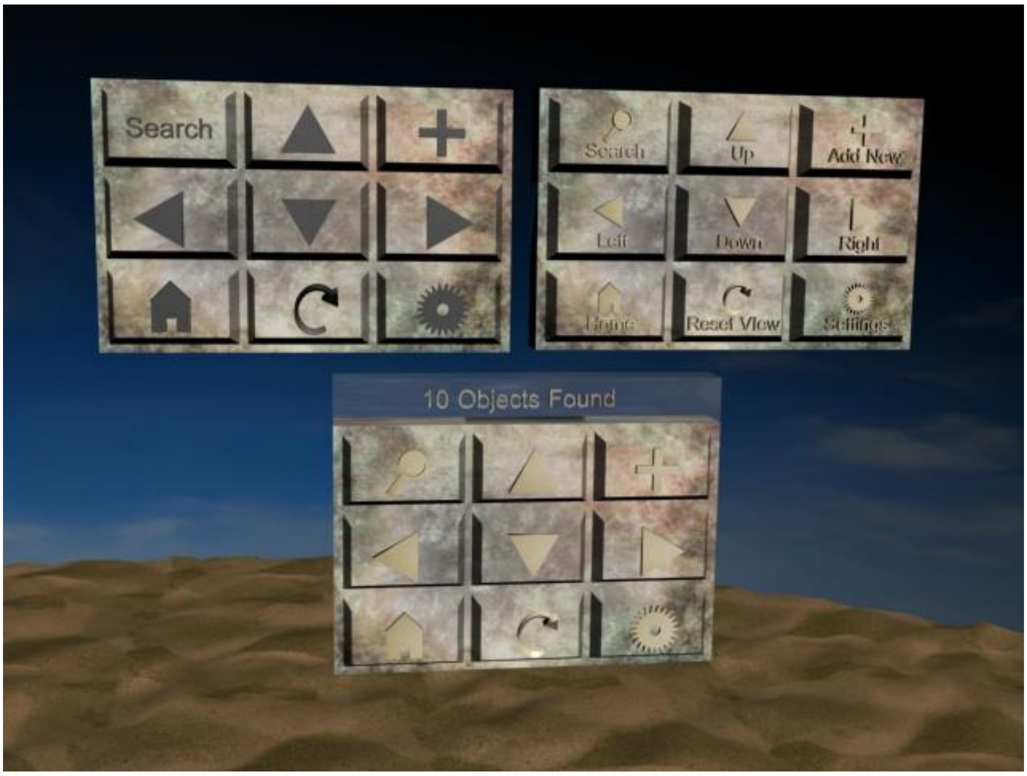



**Figure 63: Different versions of the companion menu**

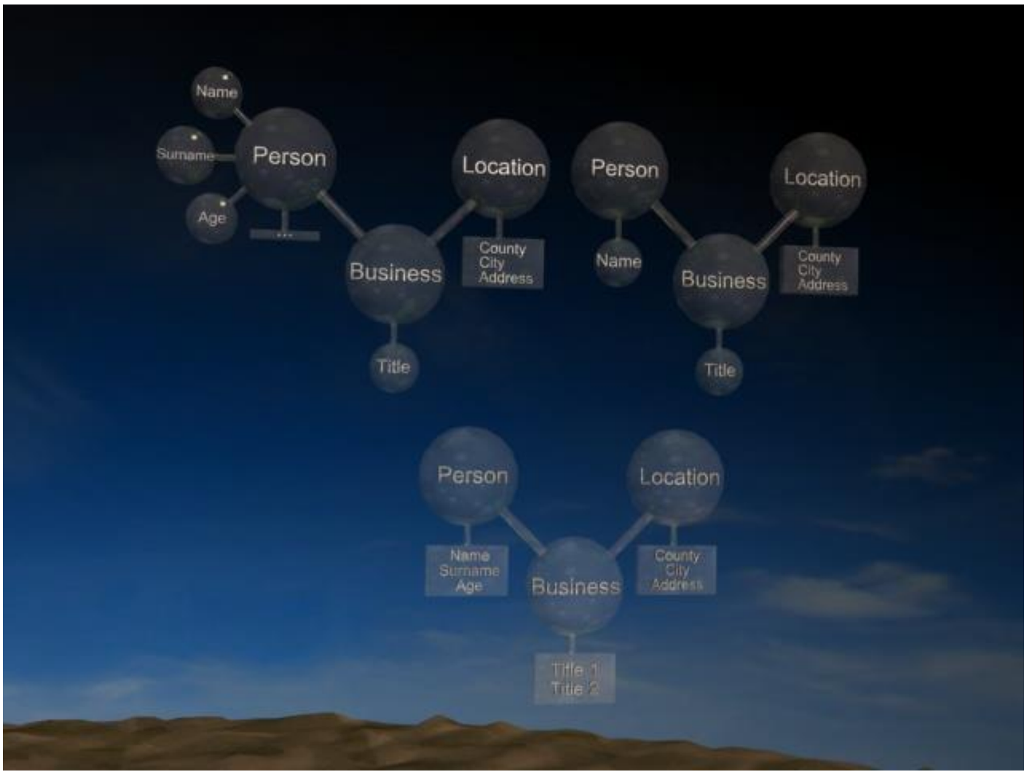



**Figure 64: Different versions of syntheses**


**Figure 65: Designing a login screen for the future Metakosmos system**

# Appendix E: Semantic Synthesis User Study

## *The User Study*

“Semantic Synthesis”

A User Study

by Konstantinos Gkoutzis

## Statistical Questions 1/4

- Gender?
  - ❑Male
  - ❑Female

2

## Statistical Questions 2/4

- Age?
  - ❑18-24
  - ❑25-30
  - ❑31-40
  - ❑41-50
  - ❑51-60
  - ❑61-80
  - ❑81+

3

## Statistical Questions 3/4

- Educational Background
  - ❑No education
  - ❑Primary education
  - ❑Secondary education
  - ❑Tertiary education
    - ❑Further/Continuing education
    - ❑Undergraduate education
    - ❑Postgraduate education
  - ❑Other: ________________________________________

4

## Statistical Questions 4/4

- IT Background
  - ❑No IT background
  - ❑Basic computer skills
  - ❑Basic computer/Internet skills
  - ❑Advanced computer/Internet skills
  - ❑Certified computer/Internet skills
    - ❑ECDL/Microsoft/Other certificate
    - ❑Relevant tertiary education degree

5

Ready?

6

**Welcome to Metakosmos**

7

**Would you like to view a short guide on how to use the system?**

**Yes / No**

8

**Guide 1/10**

**Welcome to the Metakosmos guide**

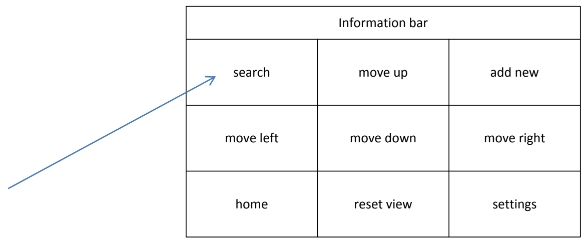
9

# Guide 2/10

**In order to search for items, you will "synthesise" (compose) your queries visually**

10

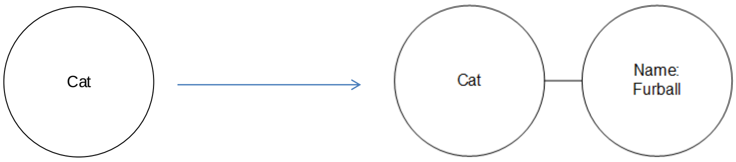
# Guide 3/10




**You can use the "Search" option of your Companion Menu to add more search items onto your interface**

11

## Guide 4/10




**You can start with a general concept, for example, of a <u>Cat</u>, and then make it more specific by adding extra information (e.g. a <u>Cat</u> with the <u>Name</u>: *"Furball"*)**

12

## Guide 5/10

**After you have synthesised your question, you can use the "Search" button in your Companion Menu to submit the query, and then all the relevant results – that are known to the system – will appear**

13

## Guide 6/10

**Let's say, for example, that you want to look for a person, first name** *John*, **last name** *Smith*, **who works at a business named** *Metakosmos*, **which resides in the location of the city named** *Athens* **in the country named** *Hellas*

14

## Guide 7/10

**1) a person,**
- **1.1) first name** *John*
- **1.2) last name** *Smith*

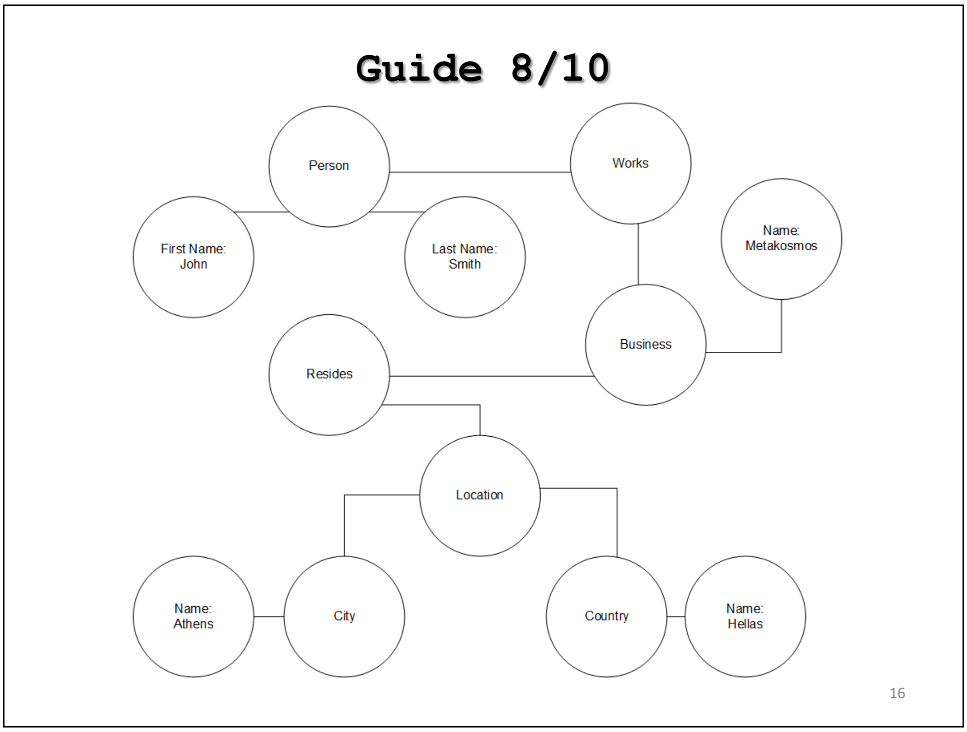
**2) who works**

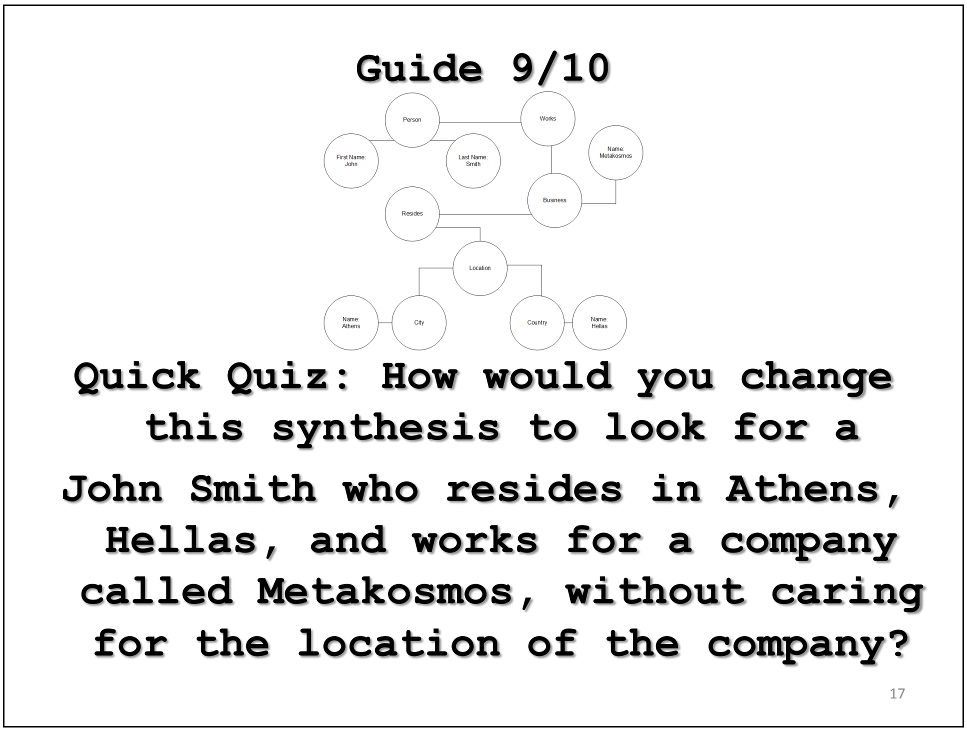
**3) at a business**
- **3.1) named** *Metakosmos*

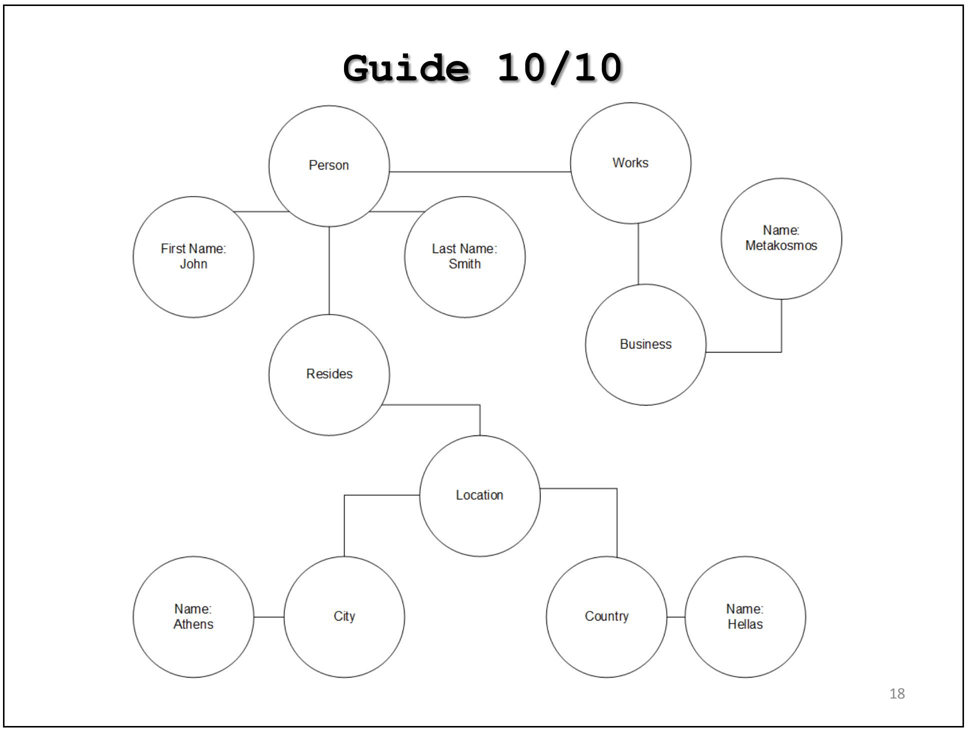
**4) which resides**

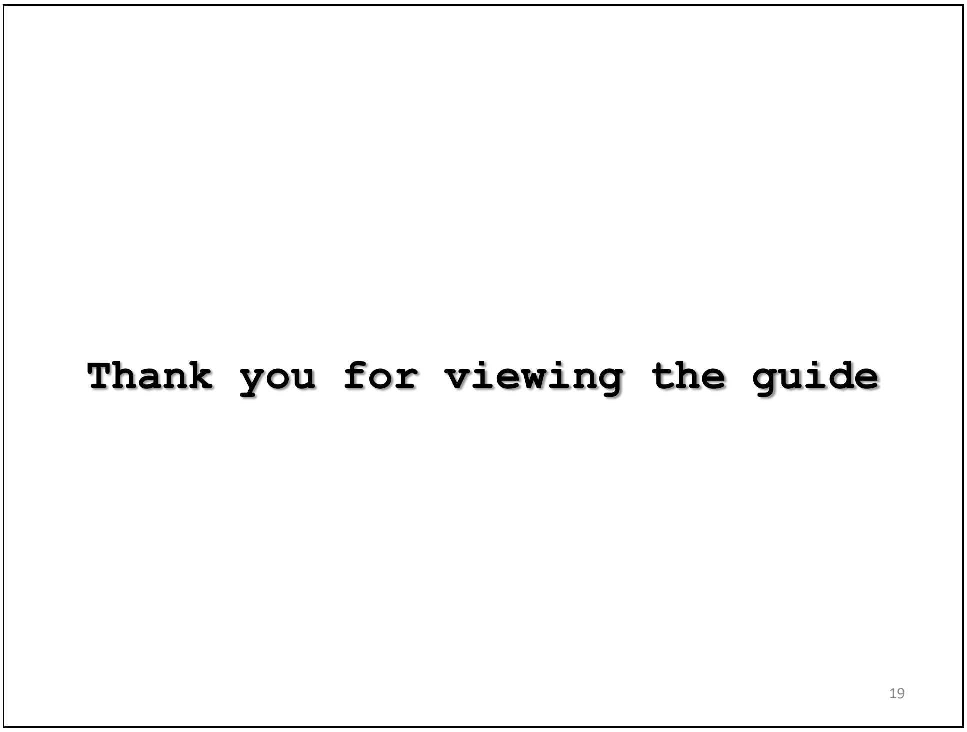
**5) in the location**
- **5.1) of the city**
  - **5.1.1) named** *Athens*
- **5.2) in the country**
  - **5.2.1) named** *Hellas*

15


Guide 8/10
Person
Works
First Name: John
Last Name: Smith
Name: Metakosmos
Business
Resides
Location
Name: Athens
City
Country
Name: Hellas
16


Guide 9/10
Quick Quiz: How would you change this synthesis to look for a
John Smith who resides in Athens, Hellas, and works for a company called Metakosmos, without caring for the location of the company?
17


Guide 10/10
Person
Works
First Name: John
Last Name: Smith
Name: Metakosmos
Business
Resides
Location
Name: Athens
City
Country
Name: Hellas
18


Thank you for viewing the guide
19

**Logging in...**

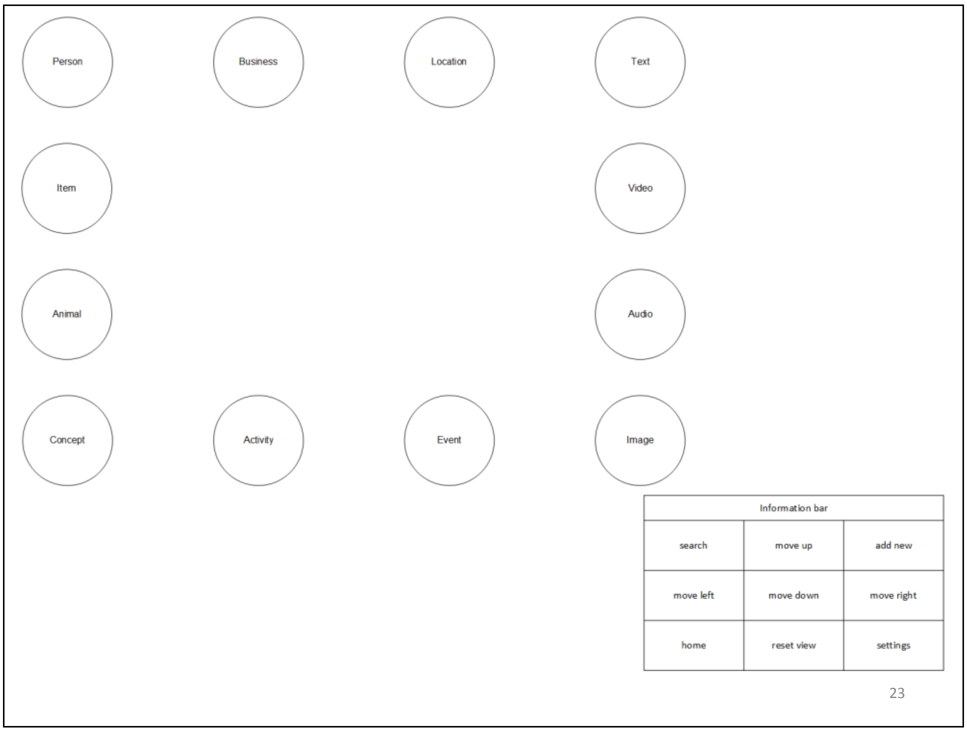
20

**Welcome, Test User**

21

**Loading your interface...**

22



Question #1: Find all the people named Jean Doe who are female.

*(you may use a separate answer sheet)*

**Note**: you will **not** be timed for this question

24

Question #2: Decide to look for something and make up your own Semantic Synthesis query.

*(you may use a separate answer sheet)*

**Note**: you will **not** be timed for this question

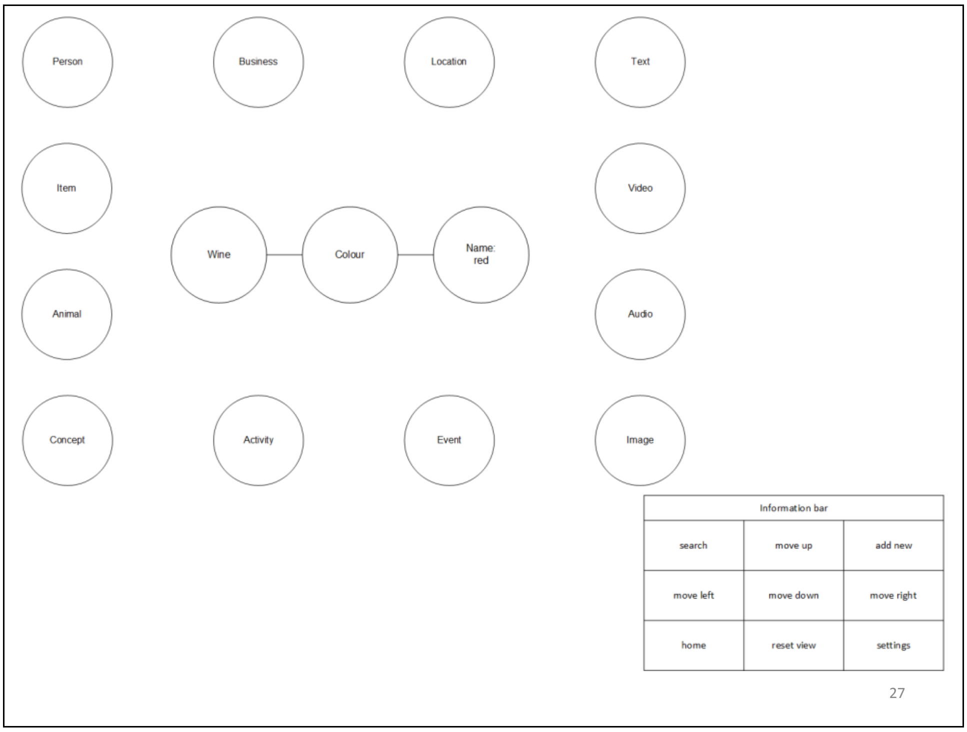
25

Question #3: Find all the wines of red colour.

*(you may use a separate answer sheet)*

**Note**: you will be timed for this question

26

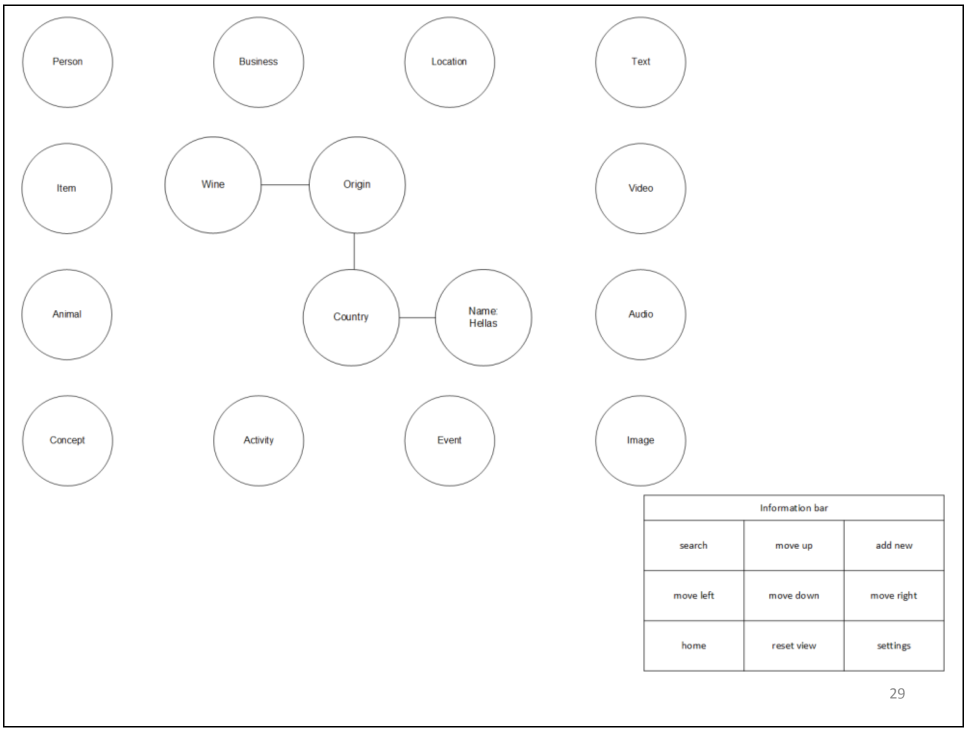


Question #4: Find all the wines with Hellas as their origin.

*(you may use a separate answer sheet)*

**Note**: you will be timed for this question

28

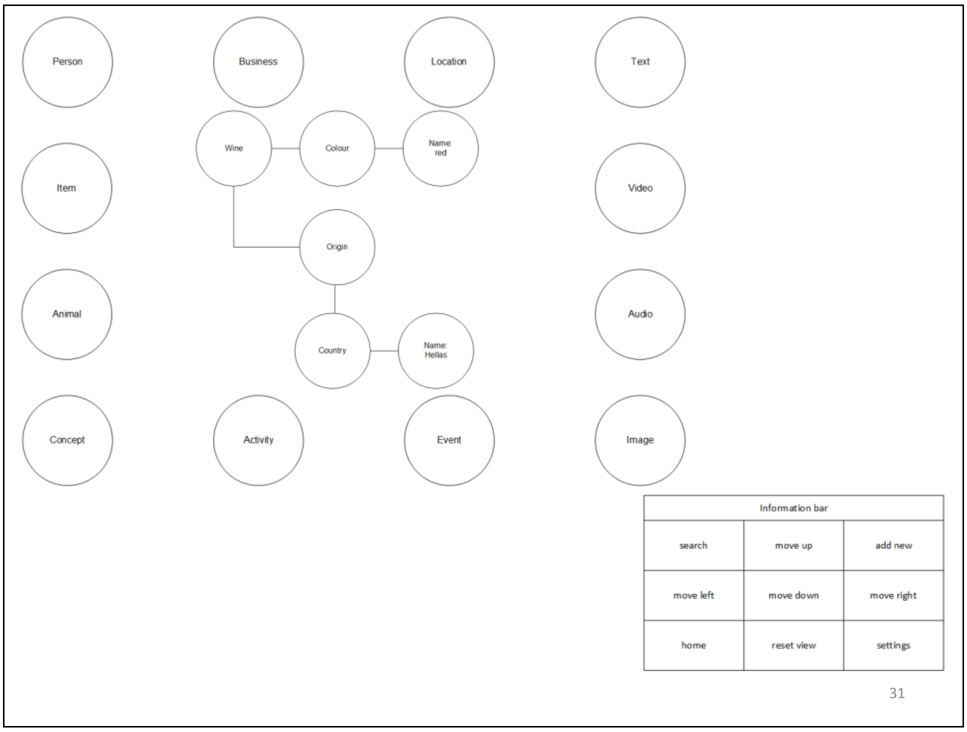


Question #5: Find all the wines of red colour with Hellas as their origin.

*(you may use a separate answer sheet)*

**Note**: you will be timed for this question

30



Question #6: Wikipedia has many information about Grape Varieties (https://en.wikipedia.org/wiki/Category:Grape_varieties).

It has all the Red Wine Grape Varieties here:
https://en.wikipedia.org/wiki/Category:Red_wine_grape_varieties

and all the Grape Varieties of Greece here:
https://en.wikipedia.org/wiki/Category:Grape_varieties_of_Greece

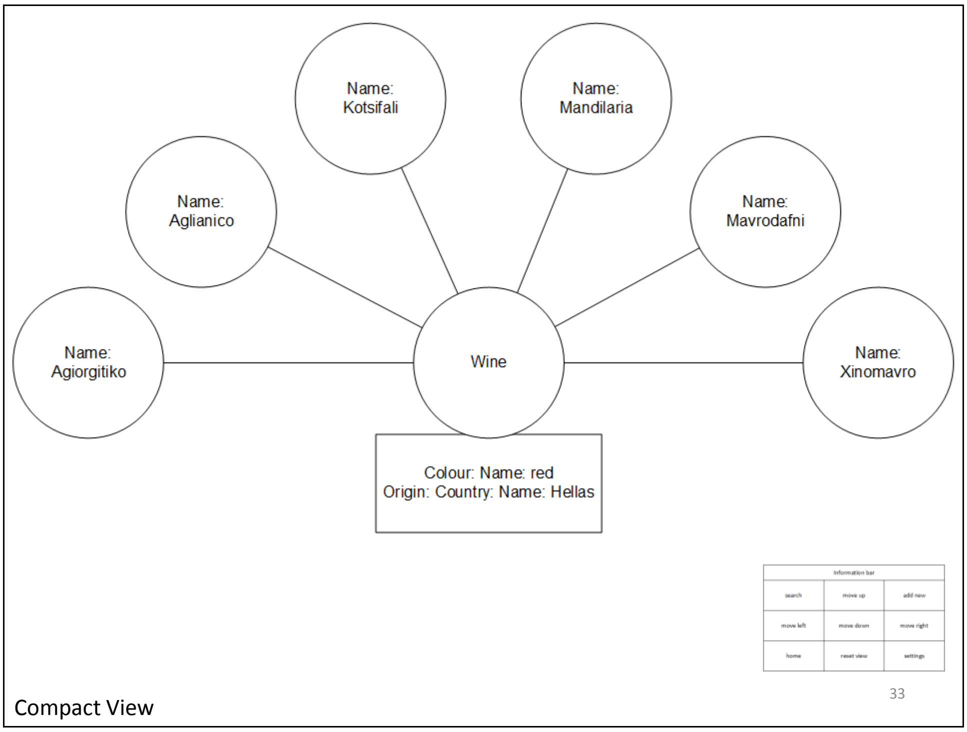
Find all the Red Wine Grape Varieties of Greece.

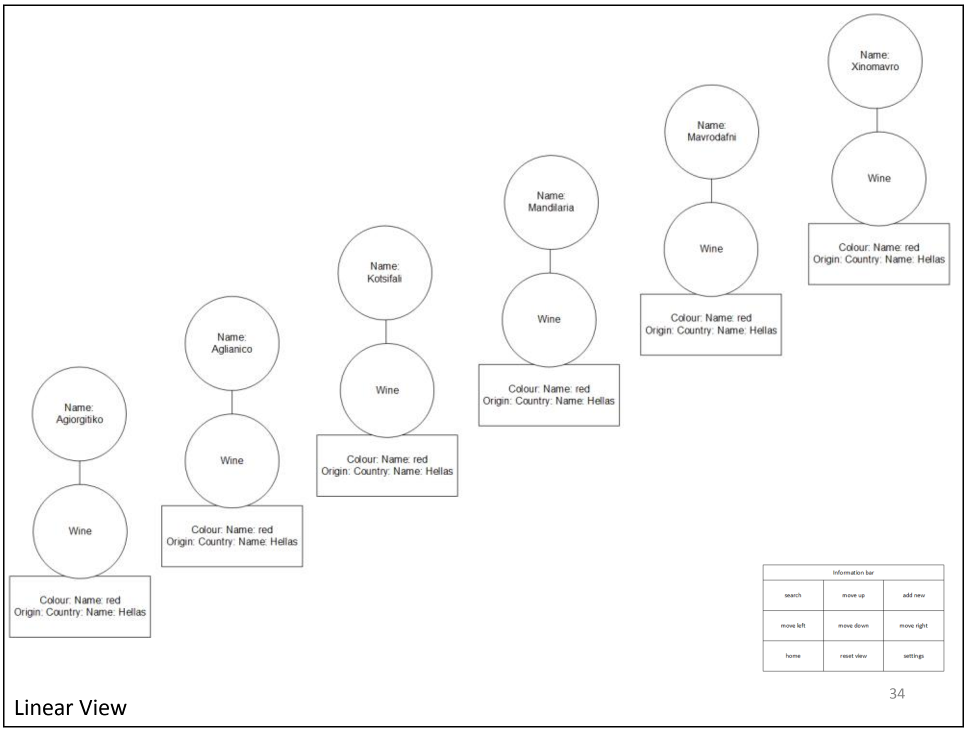
*(you may use a separate answer sheet*

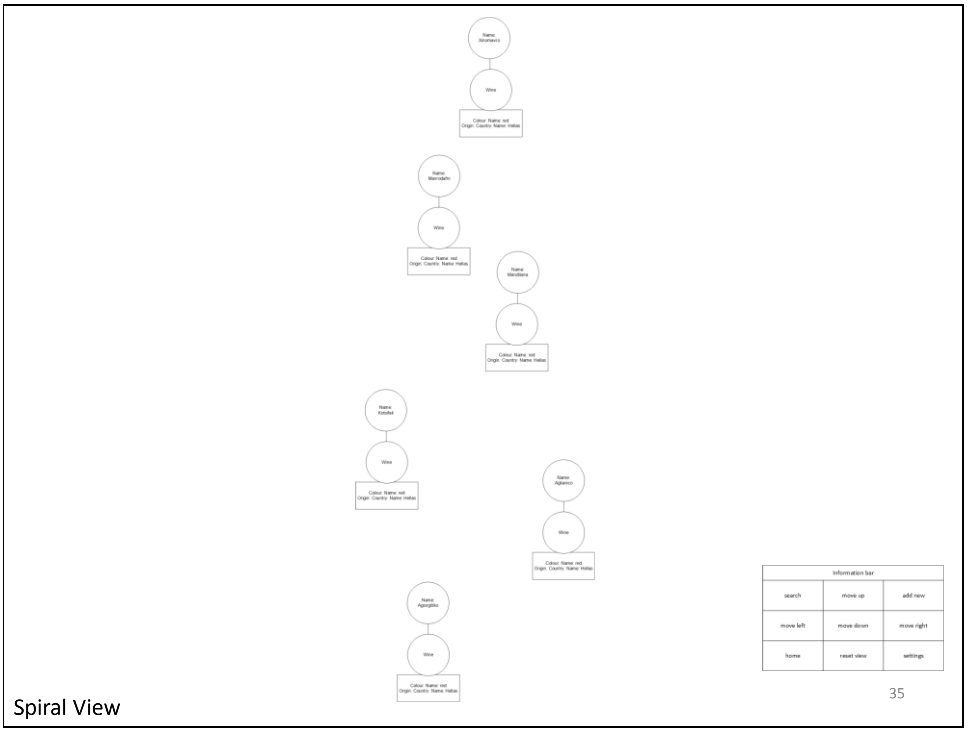
*as well as any other means available)*

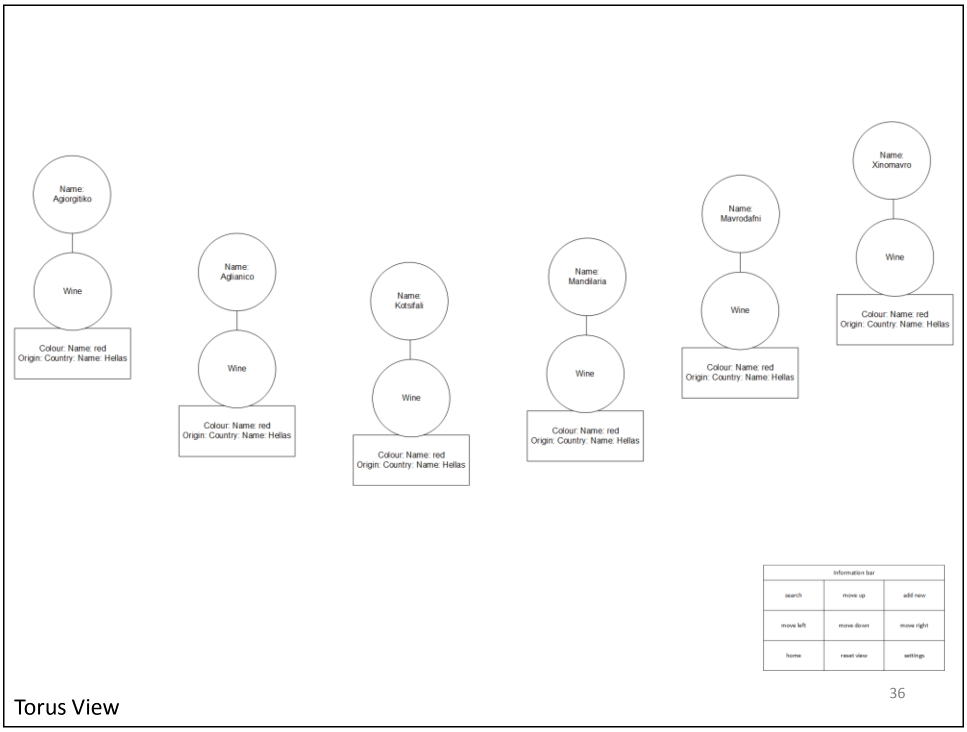
**Note**: you will be timed for this question

32




Name: Agiorgitiko
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Name: Aglianico
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Name: Kotsifali
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Name: Mandilaria
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Name: Mavrodafni
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Name: Xinomavro
Wine
Colour: Name: red
Origin: Country: Name: Hellas
Information bar
search
move up
add new
move left
move down
move right
home
reset view
settings
34
Linear View


Information bar
search
move up
add new
move left
move down
move right
home
reset view
settings
35
Spiral View




Question #7: Which <u>one</u> of the above result-set layouts did you prefer/like <u>most</u>?

Answer:

**1**

**2**

**3**

**4**

**None/Other**

37

**Logging out...**

38

**Thank you for using Metakosmos**

39

**Scale: 1 (= no, not at all) ➔ 5 (= yes, very much)**

Question #8:
Did you find "Semantic Synthesis" easy to understand?
Answer:

Question #9:
Did you find "Semantic Synthesis" easy to use?
Answer:

Question #10:
Would you use a system like Metakosmos with "Semantic Synthesis"?
Answer:

40

Thank you for participating in this user study ☺

K.

41

Please remember that you may always opt out at any given point.

The hard copies of your answer sheets (if any) will be destroyed after the results have been processed/aggregated.

You may contact the researcher at:

**Konstantinos.Gkoutzis@plymouth.ac.uk**

42

## The Results

| User Study | Participant #1 | Participant #2 | Participant #3 | Participant #4 | Participant #5 | Participant #6 | Participant #7 | Participant #8 | Participant #9 |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Female | Female | Female | Male | Female | Male | Female | Male | Male |
| Age | 18-24 | 25-30 | 18-24 | 18-24 | 31-40 | 41-50 | 31-40 | 18-24 | 18-24 |
| Educ. Background | Postgraduate | Undergraduate | Postgraduate | Undergraduate | Postgraduate | Postgraduate | Postgraduate | Further | Further |
| IT Background | Rel. Certificate | Rel. Certificate | Basic Comp/Int. | Rel. Tertiary | Rel. Certificate | Basic Comp/Int. | Rel. Certificate | Basic Comp/Int. | Basic Comp/Int. |
| View Guide | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Quiz | Correct | Wrong | Wrong | Correct | Correct | Correct | Correct | Correct | Wrong |
| Question #1 | Sem. Correct | Wrong | Sem. Correct | Sem. Correct | Wrong | Wrong | Wrong | Wrong | Wrong |
| Question #2 | Correct | Sem. Correct | Wrong | Correct | Wrong | Wrong | Wrong | Wrong | Sem. Correct |
| Question #3 | Sem. Correct | Sem. Correct | Sem. Correct | Wrong | Sem. Correct | Sem. Correct | Sem. Correct | Sem. Correct | Sem. Correct |
| Question #3 Duration | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| Question #4 | Wrong | Wrong | Wrong | Wrong | Correct | Wrong | Sem. Correct | Sem. Correct | Correct |
| Question #4 Duration | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 |
| Question #5 | Wrong | Wrong | Wrong | Correct | Correct | Wrong | Correct | Correct | Wrong |
| Question #5 Duration | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 2 | 2 |
| Question #6 | Correct | Correct | Wrong | Wrong | Wrong | Correct | Correct | Wrong | Wrong |
| Question #6 Duration | 4 | 4 | 5 | 2 | 4 | 5 | 4 | 6 | 6 |
| Question #7 | Compact | Compact | Compact | Compact | Compact | Compact | Torus | Compact | Torus |
| Question #7 Alt | Compact | Compact | Linear | Spiral | Compact | Compact | Torus | Linear | Linear |
| Question #8 | 5 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 3 |
| Question #9 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Question #10 | 5 | 5 | 5 | 5 | 5 | 3 | 3 | 4 | 4 |

**Figure 66: The results of the user study**

# Bibliography

Ackoff, R. L. (1989) 'From Data to Wisdom'. *Journal of Applied Systems Analysis*, 16 (1). pp 3-9.

Aitchison, J., Gilchrist, A. & Bawden, D. (2000) *Thesaurus construction and use: a practical manual.* Europa Publications.

AK-Encyclopaedia (2012) 'Color'. [Online]. Available at: http://academickids.com/encyclopedia/c/co/color.html.

Alesso, H. P. & Smith, C. F. (2006) *Thinking on the Web : Berners-Lee, Gödel, and Turing.* Hoboken, NJ: John Wiley & Sons.

Andrei, B. (2002) 'A taxonomy of web search'. *SIGIR Forum*, 36 (2). pp 3-10.

Antoniou, G. & Van Harmelen, F. (2008) *A semantic Web primer.* Cooperative information systems. 2nd edn. Cambridge, Mass.: MIT Press.

Barnet, B. (2005) 'The Magical Place of Literary Memory™: Xanadu'. [Online]. Available at: http://www.latrobe.edu.au/screeningthepast/firstrelease/fr_18/BBfr18a.html (Accessed: 01/09/2012).

Barrett, N. (1996) *The state of the cybernation: cultural, political and economic implications of the Internet.* Kogan Page Ltd.

Baudrillard, J. & Evans, A. B. (1991) 'Simulacra and Science Fiction (Simulacres et science-fiction)'. *Science Fiction Studies*, pp 309-313.

Behr, J., Eschler, P., Jung, Y. & Zöllner, M. (2009) 'X3DOM: a DOM-based HTML5/X3D integration model'. ACM, pp. 127-135.

Bellinger, G., Castro, D. & Mills, A. (2004) 'Data, Information, Knowledge, and Wisdom'. *Available at: http://www.systems-thinking.org/dikw/dikw.htm*,

Berners-Lee, T. (2006) 'Linked data-design issues'. *W3C*, 2009 (09/20).

Berners-Lee, T., Hendler, J. & Lassila, O. (2001) 'The Semantic Web'. *Scientific American*, 279

Bizer, C., Heath, T. & Berners-Lee, T. (2009) 'Linked data-the story so far'. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5 (3). pp 1-22.

Bratt, S. (2007) 'Semantic Web, and Other Technologies to Watch'. [Online]. Available at: http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb (Accessed: 01/09/2012).

Brutzman, D. & Daly, L. (2007) *X3D : extensible 3D graphics for Web authors.* Amsterdam ; Boston, MA: Elsevier.

Budd, T. (1991) *An introduction to object-oriented programming.* Addison-Wesley Reading, Massachusetts.

Bush, V. (1945) 'As We May Think'. [Online]. Available at: http://www.theatlantic.com/doc/print/194507/bush (Accessed: 01/09/2012).

Cantor, G. (1891) 'Uber eine elementare Frage der Mannigfaltigkeitslehre'. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 1 pp 75-78.

Chebotko, A. & Lu, S. (2009) *Querying the Semantic Web: An Efficient Approach Using Relational Databases.* LAP Lambert Academic Publishing.

Chen, S. Y., Magoulas, G. D. & Macredie, R. D. (2004) 'Cognitive styles and users' responses to structured information representation'. *International Journal on Digital Libraries*, 4 (2). pp 93-107.

Cornell University (1996) 'In Memoriam'. *Computer Science Department Annual Report*. [Online]. Available at: http://www.cs.cornell.edu/Info/Department/Annual96/Beginning/salton.html (Accessed: 01/09/2012).

Curé, O., Lamolle, M. & Le Duc, C. (2011) 'Ontology Based Data Integration Over Document and Column Family Oriented NOSQL stores'. pp. 122.

'D2R Server'. (2012). [Online]. Available at: http://d2rq.org/d2r-server.

Engelbart, D. (1962) *Augmenting human intellect: a conceptual framework.* Stanford Research Institute. Available at: http://www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0289565.

'Folksonomy Source'. [Online]. Available at: http://vanderwal.net/folksonomy.html (Accessed: 01/09/2012).

Geroimenko, V. & Chen, C. (2005) *Visualizing information using SVG and X3D : XML-based technologies for the XML-based Web.* London: Springer.

Geroimenko, V. & Chen, C. (2006) *Visualizing the Semantic Web : XML-based Internet and information visualization.* 2nd edn. London: Springer.

Gibson, W. (1984) *Neuromancer.* Ace.

Gkoutzis, K. & Geroimenko, V. (2011) 'Moving from Folksonomies to Taxonomies: Using the Social Web and 3D to Build an Unlimited Semantic Ontology', *Information Visualisation (IV), 2011 15th International Conference on*. 13-15 July 2011. pp. 205-210.

Goldfarb, C. F. (1996) 'The Roots of SGML -- A Personal Recollection'. [Online]. Available at: http://www.sgmlsource.com/history/roots.htm (Accessed: 01/09/2012).

Goodman, A. (2002) 'An End to Metatags (Enough Already, Part 1)'. Traffick.com. [Online]. Available at: http://www.traffick.com/article.asp?aID=102 (Accessed: 01/09/2012).

Google (2009) 'Google Milestones'. [Online]. Available at: http://www.google.com/intl/en/corporate/history.html (Accessed: 01/09/2012).

Google (2012) 'The Knowledge Graph'. [Online]. Available at: http://www.google.com/insidesearch/features/search/knowledge.html (Accessed: 01/09/2012).

'Hakia'. (2012). [Online]. Available at: http://www.hakia.com/ (Accessed: 01/09/2012).

Halpin, H., Robu, V. & Shepherd, H. (2007) 'The complex dynamics of collaborative tagging'. *Proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Canada: ACM, pp 211-220.

Helft, M. (2007) 'In Silicon Valley, the Race Is On to Trump Google'. *The New York Times*. [Online]. Available at: http://www.nytimes.com/2007/01/01/technology/01search.html (Accessed: 01/09/2012).

Henderson, M. (2005) 'INOZON Launches the First Search Engine in 3D'. *Press Release Newswire*. [Online]. Available at: http://www.prweb.com/releases/2005/02/prweb211778.htm (Accessed: 01/09/2012).

Hogan, A., Harth, A., Umrich, J. & Decker, S. (2007) 'Towards a scalable search and query engine for the web'. ACM Press New York, NY, USA, pp. 1301-1302.

Hogben, G. (2007) 'Security issues and recommendations for online social networks'. *ENISA position paper*, (1).

'HTML5'. (2012). [Online]. Available at: http://dev.w3.org/html5/spec/spec.html (Accessed: 01/09/2012).

'HTML5 Fullscreen API'. (2012). [Online]. Available at: http://dvcs.w3.org/hg/fullscreen/raw-file/tip/Overview.html (Accessed: 01/09/2012).

Intel (2012) 'RiverTrail'. [Online]. Available at: https://github.com/RiverTrail/RiverTrail (Accessed: 01/09/2012).

'Invu'. (2012). [Online]. Available at: http://www.invu.net/ (Accessed: 01/09/2012).

Invu (2008) 'Ergo - Frequently Asked Questions'. [Online]. Available at: http://www.yourergo.com/help.php (Accessed: 01/01/2009).

iProspect (2004) 'Search Engine User Attitudes'. [Online]. Available at: http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf (Accessed: 01/01/2009).

J Stasko, J. D., MH Brown, B Price (1998) *Software visualization: Programming as a multimedia experience.* The MIT press.

'Jinni'. (2012). [Online]. Available at: http://www.jinni.com/ (Accessed: 01/09/2012).

Johnston, C. (2012) 'Cloud storage: a pricing and feature guide for consumers'. [Online]. Available at: http://arstechnica.com/gadgets/2012/04/cloud-storage-a-pricing-and-feature-guide-for-consumers/ (Accessed: 01/09/2012).

Kittur, A. & Kraut, R. E. (2008) 'Harnessing the wisdom of crowds in wikipedia: quality through coordination', *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, pp. 37-46.

Klein, F., Sons, K., John, S., Rubinstein, D., Slusallek, P. & Byelozyorov, S. (2012) 'Xflow: declarative data processing for the web'. ACM, pp. 37-45.

Knublauch, H., Oberle, D., Tetlow, P. & Wallace, E. (2006) 'A semantic web primer for object-oriented software developers'. *W3c working group note, W3C*,

Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D. & Shepherd, M. (1990) 'Cyc: toward programs with common sense'. *Communications of the ACM*, 33 (8). pp 30-49.

Li, D., Tim, F., Anupam, J., Rong, P., Cost, R. S., Yun, P., Pavan, R., Vishal, D. & Joel, S. (2004) 'Swoogle: a search and metadata engine for the semantic web'. *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. Washington, D.C., USA: ACM.

Longman (2009) 'Dictionary of Contemporary English - A'. [Online]. Available at: http://www.ldoceonline.com/dictionary/A_1 (Accessed: 01/09/2012).

Lopez, X. D., Souripriya (2010) *Semantic Technologies in Oracle Database 11g Release 2: Capabilities, Interfaces, Performance.* Available at: http://download.oracle.com/otndocs/tech/semantic_web/pdf/2010_ora_semtech_capintper.pdf.

Markoff, J. (2007) 'Start-Up Aims for Database to Automate Web Searching'. *The New York Times*. [Online]. Available at: http://www.nytimes.com/2007/03/09/technology/09data.html (Accessed: 01/09/2012).

Mizzaro, S. (1998) 'How many relevances in information retrieval?'. *Interacting with computers*, 10 (3). pp 303-320.

Netscape Communications Corporation (2004) 'Open Directory RDF Dump'. [Online]. Available at: http://rdf.dmoz.org/ (Accessed: 01/09/2012).

Novak, J. D. & Cañas, A. J. (2008) 'The theory underlying concept maps and how to construct and use them'. *Florida Institute for Human and Machine Cognition Pensacola Fl, www. ihmc. us.[http://cmap. ihmc. us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps. htm]*,

O'Reilly, T. (2005) 'What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software'. [Online]. Available at: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (Accessed: 01/09/2012).

OrbLog (2009) 'Intel Core Series Sidebar Gadget'. [Online]. Available at: http://blog.orbmu2k.de/sidebar-gadgets/intel-core-series-sidebar-gadget (Accessed: 01/09/2012).

Pidcock, W. & Uschold, M. (2003) 'What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?'. *Q&A with an expert*. [Online]. Available at: http://www.metamodel.com/article.php?story=20030115211223271 (Accessed: 01/01/2009).

PU (2012a) 'Plymouth University - Research Ethics Policy'.

PU (2012b) 'Plymouth University - Guidelines to Application for Ethical Approval of Research Involving Human Participants'.

Putnam, R. D. (2000) *Bowling alone : the collapse and revival of American community.* New York: Simon & Schuster.

Quine, W. V. (1976) *The ways of paradox, and other essays.* Rev. and enl. edn. Cambridge, Mass.: Harvard University Press.

Rappa, M. (2000) 'Business Models on the Web'. [Online]. Available at: http://digitalenterprise.org/models/models.html (Accessed: 01/09/2012).

Raskin, V., Buck, B., Keen, A., Hempelmann, C. F. & Triezenberg, K. E. (2008) 'Accessing and manipulating meaning of textual and data information for information assurance and security and intelligence information'. ACM New York, NY, USA.

Rheingold, H. (1994) *The virtual community : homesteading on the electronic frontier.* New York, NY: HarperPerennial.

Robu, V., Halpin, H. & Shepherd, H. (2009) 'Emergence of consensus and shared vocabularies in collaborative tagging systems'. *ACM Trans. Web*, 3 (4). pp 1-34.

Rowley, J. E. (1992) *Organizing knowledge : an introduction to information retrieval.* 2nd edn. Aldershot, Hants, England ; Brookfield, Vt., USA: Ashgate.

Russell, B. (1903) 'On the meaning and denotation of phrases'. *The Collected Papers of Bertrand Russell*, 4 pp 283-296.

Russell, B. (1996) *The principles of mathematics.* WW Norton & Company.

Saracevic, T. (1996) 'Relevance reconsidered'. pp. 201-218.

Schwartz, M. F., Emtage, A., Kahle, B. & Neuman, B. C. (1992) 'A Comparison of Internet Resource Discovery Approaches'. *Computing Systems*, 5 (4). pp 461-493.

Shneiderman, B. (2002) *Leonardo's laptop : human needs and the new computing technologies.* Cambridge, Mass.: MIT Press.

Shneiderman, B. & Plaisant, C. (2004) *Designing the user interface : strategies for effective human-computer interaction.* 4th edn. Boston: Pearson/Addison Wesley.

Sonnenreich, W. (1997) 'A History of Search Engines'. [Online]. Available at: http://www.wiley.com/legacy/compbooks/sonnenreich/history.html (Accessed: 01/09/2012).

Stone, N. J. & English, A. J. (1998) 'Task type, posters, and workspace color on mood, satisfaction, and performance'. *Journal of Environmental Psychology*, 18 (2). pp 175-185.

'Swoogle'. (2007). [Online]. Available at: http://swoogle.umbc.edu/ (Accessed: 01/09/2012).

The Lively Team (2008) 'Lively no more'. *The Official Google Blog*. [Online]. Available at: http://googleblog.blogspot.com/2008/11/lively-no-more.html (Accessed: 01/09/2012).

Thuraisingham, B. M. (2002) *XML databases and the semantic Web.* Boca Raton, FL: CRC Press.

Vinson, N. G. & Singer, J. (2008) 'A practical guide to ethical research involving humans'. *Guide to Advanced Empirical Software Engineering*,

Watson, M. (2010) 'Practical Semantic Web and Linked Data Applications'.[in USA. (Accessed:Watson, M.

Wikimedia (2012) 'Wikimedia Fundraiser Concludes with Record Breaking Donations'. [Online]. Available at: http://blog.wikimedia.org/2012/01/02/wikimedia-fundraiser-concludes-with-record-breaking-donations/.

Wikipedia (2011) 'SOPA initiative'. [Online]. Available at: https://en.wikipedia.org/wiki/Wikipedia:SOPA_initiative (Accessed: 01/09/2012).

Wikipedia (2012) 'List of grape varieties'. [Online]. Available at: https://en.wikipedia.org/wiki/List_of_grape_varieties (Accessed: 01/09/2012).

'World Wide Web CERN Proposal'. [Online]. Available at: http://www.w3.org/History/1989/proposal.html (Accessed: 01/09/2012).

Wortham, J. (2007) 'After 10 Years of Blogs, the Future's Brighter Than Ever'. *WIRED Magazine*. [Online]. Available at: http://www.wired.com/entertainment/theweb/news/2007/12/blog_anniversary (Accessed: 01/09/2012).

'X3DOM'. (2012). [Online]. Available at: http://www.x3dom.org/ (Accessed: 01/09/2012).

Yu, L. (2007) *Introduction to Semantic Web and Semantic Web services.* Boca Raton: Chapman & Hall/CRC.

Zolkovskij, A. K. & Mel'cuk, I. A. (1965) 'O vozmoznom metode i instrumentax semanticeskogo sinteza [On a possible method and instruments for semantic synthesis (of Texts)]'. *Nauchno-texnicheskaja informacija*, 1 (6). pp 23-28.

## Colours

I only like the colours I do not see
those I have not associated with any bad image
all that exist to make life beautiful
and have beautified mine with their absence

These are the colours that bring me joy
because they were not there in the times of sorrow
not partaking in any depressing sight
they were always discreet and innocent

Even if I never get to see them
or they do not appear in some pleasant memory
I will know that they were always there
reflecting on my thoughts

This is the world of my dreams
strange sequences of unconnected memories
filled with light from the beyond…
I only like the colours I do not see