**MYKOLAS ROMERIS UNIVERSITY-MIDDLESEX UNIVERSITY**

**INSTITUTE OF COMMUNICATION**

**RENATAS LIBERIS**

(Communication and Creative Technologies program)

# SEMANTIC WEB TECHNOLOGIES IN INTERNET CONTENET SEARCH FOR EDUCATION

**Master Thesis**

Supervisor-
assoc. prof. dr. Gintarė Žemaitaitienė

Vilnius, 2017

# CONTENTS

**LIST OF FIGURES** ..... **4**

**LIST OF TABLES** ..... **5**

**THE LIST OF ABBREVIATIONS** ..... **6**

**INTRODUCTION** ..... **7**

**1 INFORMATION RETRIEVAL USING KWB SEARCH ENGINES AND SWT SEARCH ENGINES** ..... **11**

1.1 Definitions of term *information* by different authors ..... 12

1.2 Definition of term *knowledge* by different authors ..... 13

1.3 Information processing levels by *Data*, *Information* and *Knowledge* ..... 15

1.4 Evolution of information retrieval systems to KWB search engines ..... 16

1.5 Evolution of KWB Search engines ..... 17

1.6 Effectiveness determination for search engines ..... 18

1.7 Limitations of information retrieval from KWB search engines ..... 19

1.8 Semantic Web Technologies for new generation information retrieval ..... 20

**2 EVALUATION OF SWT IN "DIFFUSION OF INNOVATIONS", "CROSSING THE CHASM"AND "MATHEMATICAL THEORY OF COMMUNICATION" THEORIES 22**

2.1 Evaluation of innovativeness dimensions in SWT ..... 22

2.2 Evaluation of SWT on crossing "The Chasm" ..... 23

2.3 Evaluation of "Noise Source" in SWT ..... 24

**3 SCIENTIFIC RESEARCH "APPLICABILITY OF SEMANTIC WEB TECHNOLOGIES IN INTERNET CONTENT SEARCH FOR EDUCATION" METHODOLOGY** ..... **25**

3.1 Qualitative data analysis results ..... 32

*3.1.1 Insight from the interviews with experts from Semantic Web Technologies field ..... 32*

*3.1.2 Insight from the interviews with Mykolas Romeris University librarians ..... 45*

**CONCLUSIONS AND RECOMMENDATIONS** ..... **50**

**REFERENCES** ..... **52**

**SUMMARY** ..... **56**

**SANTRAUKA** ..... **57**

**APPENDICE NO. 1** ..... **58**

**APPENDICE NO. 2 ..... 60**

**APPENDICE NO. 3 ..... 61**

**APPENDICE NO. 4 ..... 62**

# LIST OF FIGURES

**Figure 1.** Comparison of Internet users in the world by the year……11

**Figure 2.** Comparison of search engines in a world market share……..….12

**Figure3.** Levels of information processing……..15

**Figure 4.** Adopter categorization on the basis of innovativeness……22

**Figure 5.** The Revised Technology Adoption Cycle……...……23

**Figure 6.** Communication system model ……...……23

**Figure 7.** Experts and librarians interview method design……..25

# LIST OF TABLES

**Table 1.** Definitions of term *information* by authors……...12

**Table 2.** Definitions of term *knowledge* by authors……..…14

**Table 3.** Criterion sampling for the first part of interviews with experts of SWT……...29

**Table 4.** Criterion sampling for the second part of interviews with MRU librarians……..30

# THE LIST OF ABBREVIATIONS

Data base – DB;

Mykolas Romeris University – MRU;

Semantic Web Technologies – SemWeb or SWT.

# INTRODUCTION

**Relevance of the topic.** As quantity of information is constantly growing, people need to be able to manage information flow and to be able to analyse it with trust of high quality content, reliability of results they been given and short periods of time of retrieving this information and to make this information to knowledge. As such high stakes is needed the department of Information Technology is constantly developing applications to deal with information overflow and accuracy of transmitted information and knowledge through the network of internet.

Researchers defines that information management and knowledge management in business is one of key factors to make a success in production, manufacturing and service industries (Woodside and Quaddus, 2015; Brajer, 2016). Researchers also acclaims that knowledge management became a main term to define the business agility and organisations performance (Patel, 2016). "When data and information are important for the success of an organisation, data and information become assets for the organisation" (Borek, 2014, p. 7). So companies and organisations who deals with loads and loads of data and information and this data and information becomes their assets, it is become crucial factor to manage this data and information so it can be retrieved fast, accurate and fast when there is a need to deal with it. Also when there are a big load of data and information it is advantage to not only retrieve information needed but also to see what lies underneath of it, to see relations between data and information so people don't see it or need a knowledge and time to see it.

In 2001 the idea of Semantic Web Technologies was proposed as a solution to problem of data, information overload and knowledge management arising by Tim Berners-Lee, James Handler and Ora Lassila in science magazine Scientific American on which they said that "The Semantic Web Technologies will enable machines to COMPREHEND semantic documents and data" and "Properly designed, the Semantic Web Technologies can assist the evolution of human knowledge as a whole" (Berners-Lee, Hendler, Lassila, 2001) giving a hope of all this information can be managed to work for people interests through interaction between people and machines so people could share not only information but knowledge too.

**Level of investigation.** Passin (2004) presents concepts and basics of Semantic Web Technologies, Hitzler, Krotzsch and Rudolph (2009) covers the questions on how Semantic Web Technologies is development, about standards should be used or even what languages it should be used to code. Amerland (2014) explains how Semantic Web Technologies can be integrated in certain business structures to increase brand impacts and how Semantic Web Technologies can amplify company online presence. Devedzic (2008) discusses how to use the results and technology of new

fields of research and development, such as Semantic Web Technologies and Web intelligence, to increase effectiveness in online education and make it more appealing to educators, teachers, and authors. The author integrates the examinations of learning-oriented topics such as learner modeling, collaborative learning, learning management, learning communities, ontological engineering of Web-based learning, and related topics with these technical topics.

There is huge number of books, articles and researches done about

Semantic Web Technologies in Information Technology field and number of them is increasing every year. But there is a low number of researches about Semantic Web Technologies done in Education field, so this field is open for explorations on how Semantic Web Technologies could be applied in learning processes.

**Novelty.** Even thought Semantic Web Technologies concept was presented in 2001 but still in 2017 there is a lack of researches done on how Semantic Web Technologies can be integrated to Education. Furthermore, there is lack of information how Semantic Web Technologies could be integrated in content search and information retrieval for students.

Another case is that Mykolas Romeris University librarians are known for as advanced academic librarians in Lithuania who teaches academic society on how to find information and recourses for academic and scientific works, but there were never been any research done on how Semantic Technologies and Semantic Web could effect processes on finding information and recourses in internet and data bases.

**The master thesis problem.** Use of Semantic Web Technologies for educational content search and theoretical assumptions about how it affects the effectiveness of internet content search for studying.

**Research object.** Semantic Web Technologies.

**Research purpose.** To investigate the possibilities to apply Semantic Web Technologies into content search for studying.

**Research objectives:**

1. To present the main aspects of internet content search with KWB search engines and SWT based search engines.
2. To define limitations of internet content search with KWB search engines.
3. To make qualitative interviews with experts of Semantic Web Technologies to get an insight about these technologies on internet content search and

4. To make short analysis of Semantic Web Technologies through "Diffusion of Innovations", "Crossing the Chasm" and "Mathematical Theory of Communication" theories.
5. To make qualitative interviews with librarians and provide an insight on internet content search for studying.

**Research methodology.** According Creswell (2014) participatory knowledge claims, narrative design, and open-ended interviewing is the way to get needed information from individuals by making individual interviews to collect personal experiences on issues we seeking to solve. Also based on Creswell (2014) it was chosen to make a qualitative approach to get insights from two types of experts on this research on their individual experiences, first with experts who works with Semantic Technology and second type of experts - librarians who are working with students who seeks information through internet content search. So in this master thesis two types of interviews where made.

First part of interviews was made with experts of Semantic Technology field to get better knowledge on how semantic technology evolving in different fields such as businesses and education and how it can be applied in these fields, mainly we are focusing on education part and how it can be applied for making better internet content search. So from this part we get a deeper understanding why Semantic Technology is needed and how it can help to concentrate on understanding material, critical assessment and expand knowledge creatively instead of making ineffective and needless content search towards finding information which is needed to reach goals.

Second part of interviews was made with librarians of University who makes presentations for students on how effectively and efficiently find information and recourses using internet content search and data bases of the University for studies and scientific works. Librarians point of view show the insights of problems which students and researchers facing looking through information on internet and databases.

According Creswell (2014) if there is a need of getting more reliable information on research - the qualitative interview is the best approach towards unlocking it especially if there is a little number of researches done in field of investigation.

According Denzin and Linkoln (2000) qualitative approach has more flexible instruments to achieve information which is been researched rather than quantitative approach which is stable from the beginning of research to the end of it. So in order to achieve our purpose of this work to investigate the possible impacts of applying Semantic Web Technologies for internet content search for studies and scientific works purposes the approach of qualitative interviews are the best option to do it.

**The structure of the master thesis.** This master thesis consists of three main pars, conclusions, recommendations, references, summary in English and Lithuanian languages and appendices.

First chapter is about what is the difference between information and knowledge and how data, information and knowledge form information process, what is information retrieval, what is internet content search, why do people need internet search and a brief internet search history through history of the internet.

Second chapter consists of explanation on how theories “Diffusion of Innovation”, “Crossing the Chasm” and “Mathematical Theory of Communications” could be adopted and evaluated by Semantic Web Technologies.

Third chapter consist two parts of qualitative interviews with experts of Semantic Web Technologies about application of Semantic Technologies and Semantic Web Technologies to businesses, government, personal and professional live and education, and second part of interview with Mykolas Romeris University librarians about student abilities to retrieve information from internet and data bases, applicability of information retrieval trainings to other types of businesses and limitations of information retrieval from internet and databases.

In a final chapter there are the conclusions of this master thesis and recommendations for further investigations, summary of this work in English and Lithuanian languages, references and appendices.

# 1 INFORMATION RETRIEVAL USING KWB SEARCH ENGINES AND SWT SEARCH ENGINES

People raises questions that they need to answer them. In order to answer these questions there is a need of information and knowledge. So through the time all types of information are valuable asset for people, businesses, organisations and societies. Information become more and more valuable resources to understand people around us, understand processes happening to us or our environment around us and because of that people become more and more dependant on information. But first of all if people want to get some kind of information they need to retrieve it as written information on paper, visual information in photos etc. For a thousands of years, people known that information is powerful. Evolution of the computers and networks and digital era, people can share, work and store information in computers, smartphones even smart watches, etc. But with such a huge rates of information growth, people encounters with problems such as how to retrieve information you actually need. According The World Bank, in 2015 there was 79,6% people who has internet access in Europe in comparison with 20,6% internet access in 2000 (see Figure 1). According Netmarketshare, Google are the leading search engine in world with 81.56% of market share in 2012 and still leading with 71.41% of market share in 2016 (see Figure 2) and according Google Zeitgeist, people made1.2 trillion searches in 2012.

According data from The World Bank, Netmarktshare, Google Zeitheist we can state that information retrieval is highly important for people, business, government and education, and importance of information retrieval is still growing.

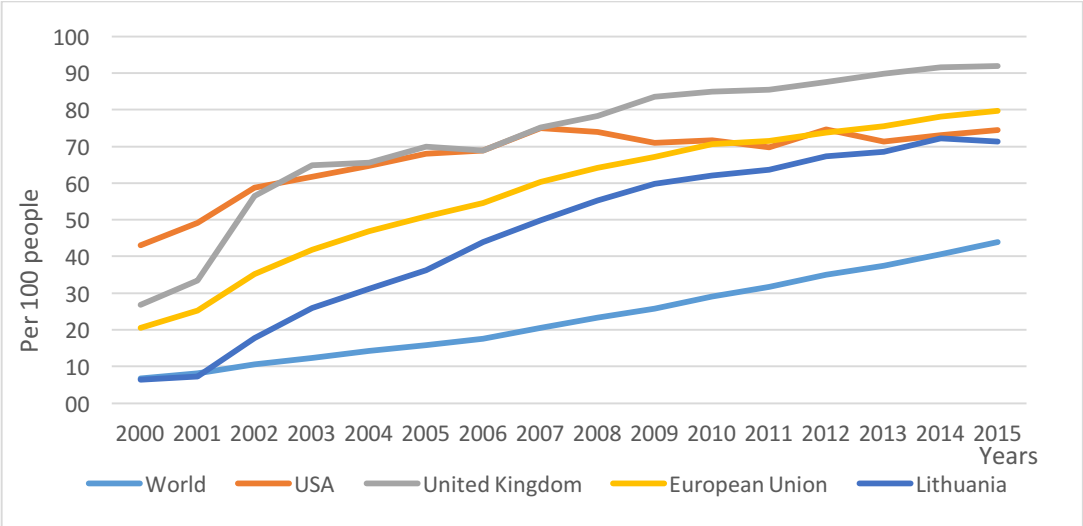



**Figure 1.** Comparison of Internet users in the world by the year. **Source:** prepared by author, according to data form The World Bank

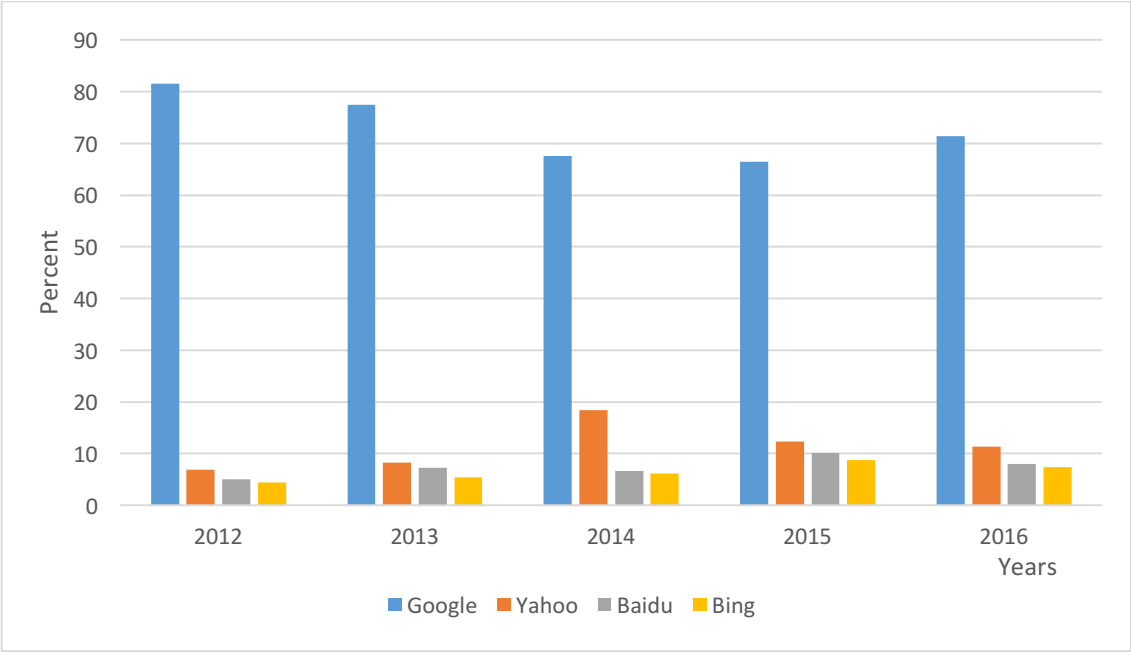



**Figure 2.** Comparison of search engines in a world market share. **Source:** prepared by author according to data from Netmarketshare.

### 1.1 Definitions of term *information* by different authors

To get a deeper understanding about term *information*, we sum up definitions used by different authors in Table 1.

Table 1. Definitions of term information by authors.

| Author | Quote |
|---|---|
| (MacKay, 1955, p. 183) | *"Both mean by information that which promotes or validates representation activity: (activity from which it is possible to infer something about some other state of affairs). Both are entitled to regard the function of information to be selective: to prescribe choice or decision".* |
| (Case, 2002, p. 5) | *"Information can be any difference you perceive, in your environment or within yourself. It is any aspect that you notice in the pattern of reality".* |

| | |
|---|---|
| (Floridi, 2010, p. 2) | *"Thus, following Shannon, Weaver [1949] supported a tripartite analysis of information in terms of*<br>*(1) technical problems concerning the quantification of information and dealt with by Shannon's theory*<br>*(2) semantic problems relating to meaning and truth; and*<br>*(3) what he called "influential" problems concerning the impact and effectiveness of information on human behaviour, which he thought had to play an equally important role".* |
| (Meadow, Boyce and Kraft, 2000, p. 37) | *"An operational definition is that information is data that changes the state of a system that perceives it, whatever a computer or a brain; hence, a stream of data that does not change the state of its receiver is not information. Such a definition was proposed by Shreider (1970)".*<br>*"A related definition has it that information is what is used to affect a decision".* |
| (Yovits, Foulk and Rase, 1981, p. 189) | *"We treat information as data of value in decision-making".* |
| (Hepworth and Walton, 2009, p. 18-19) | *"Information can be any difference you perceive, in your environment or within yourself. It is any aspect that you notice in the pattern of reality".* |

**Source:** prepared by author according to MacKay (1955), Case (2002), Floridi (2010), Meadow, Boyce and Kraft (2000), Yovits, Foulk and Rase (1981), Hepworth and Walton (2009).

There is no particular definition of *information*. Authors defines information as a part of exact sciences which describes information as a data which can be processed by machines and simultaneously it can be a process of a decision-making.

### 1.2 Definition of term *knowledge* by different authors

To get a deeper understanding about term *knowledge*, we sum up definitions of knowledge by different authors they used in their books in Table 2.

Table 2. Definitions of Knowledge by authors.

| Author | Quote |
|---|---|
| (Davenport and Prusak, 1998, p. 17) | *"Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. In originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organisational routines, processes, practices, and norms".* |
| (Blasckburn, 2005, p. 123) | *"Epistemology (Gk., Episteme, knowledge) The theory of knowledge. Its central questions include the origin of knowledge; the place of experience in generating knowledge, and the place of reason in doing so; the relationship between and certainty, and between knowledge and impossibility of error; the possibility of universal scepticism; and the changing forms of knowledge that arise from the new conceptualization of world".* |
| (Oxford dictionary of computing, 2004, p. 265) | *"Information that can be expressed as a set of facts and is known to an agent or program. Knowledge can be distinguished from information or data by its embodiment in an agent; for example, an agent might receive information that increases its knowledge".* |
| (Dictionary of Information & Library Management, 2006, p. 113) | *"<...> abstract information and understanding that somebody has about a subject".* |
| (Dictionary of ICT, 2004, p.137) | *"<...> what is known".* |

**Source:** prepared by author according to Davenport and Prusak (1998), Blasckburn (2005), Oxford dictionary of computing (2004), Dictionary of Information & Library Management (2004), Dictionary of ICT (2004).

According authors Definition of *knowledge* like definition of *information* don't have one exceptional definition, but we can define *knowledge* as information integrity to get understanding of meanings to make a decision".

### 1.3 Information processing levels by *Data*, *Information* and *Knowledge*

According Thierauf (1999) there are five levels of summarization in which three levels are major parts to information processing. *Data* is lowest level in information processing, and this level is represented by all unsorted facts and reports, this level is biggest one in which problem importance is minor to trivial, description approach is individual and nature of problem is structured. Second level is *information* is structured and semistructured data for analysing and resolving critical problems which could have major to minor problem importance and advisors' decision approach. Finally, *knowledge* is the smallest part in information processing which is obtainable form experts with actual experiences, problem impotence at this level is major and nature of problem is semistructured. To understand all levels, see Figure 3.

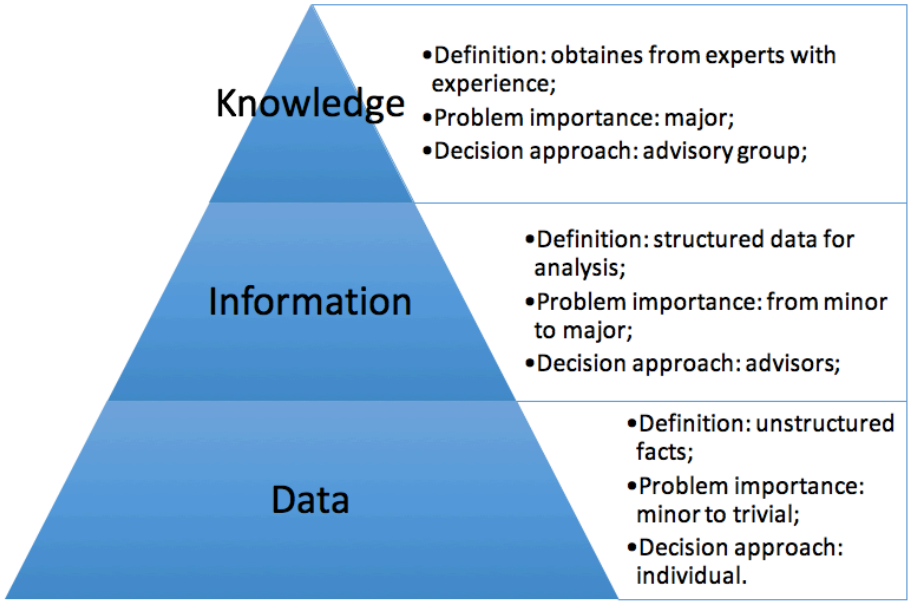



**Figure 3.** Levels of information processing. **Source:** prepared by author according Thierauf (1999)

According Thierauf (1990) levels of information process we can say that in order to reach knowledge of some specific skill or make an intelligent decision, fist there is a need to get data (unstructured facts) process that data to information (structured data for analysis) which leads to knowledge which leads to a skill performance or intelligent decision of some purpose.

## 1.4 Evolution of information retrieval systems to KWB search engines

In previous subchapter we talked about data, information and knowledge. In order to get some kind of knowledge, first we need to retrieve some information. The field of information retrieval has come a long way in the last forty years, and has enabled easier and faster information discovery. (Singhal, 2001).

In early 1950s term Information retrieval was introduced by Mooers as statement "a finding or discovery process with respect to stored information which embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation" (Mooers, 1951).

The term *search* denotes the specific behaviours of people engaged in locating information (Marchionini, 1995, p. 5). In contrast, *information retrieval* is finding material of an unstructured nature that satisfies an information need from within large collections stored on computers (Manning, Raghavan, & Schütze, 2008). Belkin (1993) implies that the central component of retrieving information is person itself, because he is making interaction with text on finding information he is needed and this information retrieval according Belkin (1993) considered as information-seeking behaviour which appears from desires to resolve problems (reading newspapers to get information about world, watching television to entertain ourselves, go to libraries to learn, etc.).

"In the situations which lead to information-seeking behaviours, our interactions with texts often takes place within the context of the systems, social and technical, which are constructed to support us in accomplishing the goals of the behaviours. Such systems include, for instance, libraries, information centres, advisory services, and information retrieval systems" (Belkin, 1993).

According Singhal (2001) information retrieval idea started in 1950s when idea occurred to make a search of the text using computer and H.P. Luhn in 1957 made a proposal that text could be retrieved by simply using words as indexing units and words that overlap could retrieve specific documents from computer memory. Most significant development of information retrieval was in 1960s when the SMART system was created by Gerard Salton which is still used in today Information Retrieval systems. This system allows developers to experiment with the ideas to improve search quality. This system was a breakthrough in text search which give a way for making best methodologies for further systems development. Development of information retrieval continued through 1970s and 1980s but progress of information retrieval using lager scale of documents were not happening until first Text Retrieval Conference were held in 1992 in US. It was a beginning of

modification of used old techniques for retrieving information which were adopted to make it in larger scales. Also the development started in retrieval of spoken information, information retrieval in non-English language, information filtering and the beginning of filtering World Wide Web started on 1996.

In Sanderson and Croft (2012) history analysis of information retrieval systems they note that after the World Wide Web was presented to public information retrieval systems faced issues with a beginning of commercial usage of information retrieval systems and research-oriented systems, but these issues were overcome and there was a beginning new era for commercial searching tools.

Over the years, information retrieval methods and techniques has changed and evolved to meet the demands of speed and relevancy of the results. Web search engines were presented as a substantial solution for internet users to retrieve information more efficient way.

## 1.5 Evolution of KWB Search engines

In order to collect some information, people started using search engines. Search engines are currently the primary information gatekeeper of the Web, holding the precious key needed to unlock the Web both, for users who are seeking information and for the authors of web page wishing to make their voices heard (Levene, 2010, p. 64). According Hock (2001) before search engines were presented, information retrieval on internet were almost impossible, the chaos was prevailed. If you want specific information, you need to know specific web page address or you can try to find information by using so called “Gophers” (no-HTML based, file title or very short description based protocol to find information on internet). In 1994 first successful web search engine “WebCrawler” came to market. So there was a beginning of the search engines era thus the variety of search engines begin to come to users’ attention. After one year competitors like “Lycos”, “Infoseek” and “OpenText” started to crawl the internet too and in a late 1995 “AltaVista” and “Excite” came to market too.

The real competition started in 1998 when “Google” presented their search engine which brought ultra-simple interface and “page-rank” technology which made searching much simpler. Competition were in measuring the size of search engine. In 1999 search engine called “Fast Search” declared that they have over 200 million records on their own databases. And in January 2000 four other competitors declared about crossing 200 million record line.

Over almost two decades some search engines were closed due changing of techniques used to retrieve information and constantly increasing number of web pages to deal with incapability of such a huge search engine demands. Others successfully making searches even today and managing to deal with information overload. Even today "Google" is one of the most usable search engines in the market (see Figure 2).

## 1.6 Effectiveness determination for search engines

It is hard to determine on which search engines are effective or which are less effective. One reason is that there are a variety of quality factors that can be applied to search engines. According (Lewandowski & Höchstötter, 2007) effectiveness of search engine could be determined by four major areas:

- Index quality: this points out the importance of the search engines' databases for retrieving relevant and comprehensive results.
- Quality of the results: this is where derivate of classic retrieval tests are applied.
- Quality of search features: a sufficient set of search features and a sophisticated query language should be offered and work reliably.
- Search engine usability: here it is asked whether it is possible for users to interact with search engines in an efficient and effective way.

According Lewandowski (2015) and his study on evaluating effectiveness of two biggest at the moment web search engines in the world: Google and Bing. In study Lewandowski declares that regarding navigational queries, Google is more efficient than Bing, but in informational queries, there is slight differences which user should not notice.

So to determine effectiveness of search engine we need to take account to its index quality, quality of the results, quality of search features and search engine usability.

## 1.7 Limitations of information retrieval from KWB search engines

Levene (2010) say that people who uses search engines to find relevant information on the internet, they turning to search engines as primary information–seeking strategy. By doing that people sometime faces difficulties so author distinguishes fife types of limitations of information retrieval on search engines:

1. Instant information flux. Information on the web is an open system, so information is constantly changing, new websites are coming and older are disappearing and all content on internet is emergent.
2. Quality in internet content is uncertain. People need to make their own judgement on information quality in internet because even though results were ranked in top positions, but the page ranking does not reflect the quality of the content.
3. Factual knowledge is not objective. At some queries search engine provides the direct answer to the question, but there is no actual proof on this information to be precise.
4. Finding information from the first attempt sometimes is impossible. There's always an uncertainty about information existence on internet people trying to find out, but sometimes one single search is not enough to retrieve this information so people need make a lot of different queries or to combine several strategies to find information or just to give up looking.

According Machill and others (2004) search engines faces some issues which can represent limitations of them:

1. Search engines provides explicit content which can be problematic with the protection with minors (e.g. pornography and violence)
2. Also search engines could lead to extremist sites, which could lead to all kind of terrorism (e.g. right-wing extremism, ISIS, etc.).
3. Search engines provides information with copyrighted material, which could lead to copyright infringement (e.g. MP3 downloads, pirated movie downloads, etc.).
4. Even though search engines provide one box search and it is convenient to people who understand web browsing less, the results of the search may be not relevant to user, and to make relevant search user need to know refinement strategies to properly use search engine.
5. Ranking of lists sometimes is poor due "spam protection team" which could provide invalid results to user's queries.

So even though search engines are great tool to find relevant information on the internet, but it still has its own flaws, such as flux of information, quality of information, poor objectivity to direct questions, uncertainty of result existence and also flaws with results listed which may contain explicit content, copyright infringed material. Although information retrieval looks like simple job, but due simplicity of interface, but user need to know refinement strategies to get relevant information and rankings of list has its own flaws by sometimes making results invalid.

## 1.8 Semantic Web Technologies for new generation information retrieval

Yo (2007) and Passin (2004) summarize ideas of KWB search engines into becoming Semantic Web (by enabling Semantic Web Technologies) into these points:

- Internet is made of wide variety of web pages (or web documents) in only human readable format;
- Machines can't read web pages (documents), they only have instructions on how to find them and present them in most convenient place.
- All web pages to machines looks equally same and they don't have ability to understand meaning of them;
- Because machine can't understand meaning, it can't make intelligent decisions;
- Web developers can't change all web pages on a global scale to make them more traceable;
- There is possibility to slightly change all web pages (documents) giving them some extra information thus enabling machines to understand the meaning.
- To accomplish this task there is a need of tools and agents to be constructed, which can make these changes in web pages (documents) in global scale.
- Agents who can help to retrieve information also can be a servant of humanity, which allows computers to systemize information and documents which can lead to execution of sophisticated tasks for humans.

This Web which consists of slightly changed web pages (documents) enabled by tools and agents is called Semantic Web. So according Yo (2007) and Passin (2010) Semantic Web Technology tools would allow to retrieve information more effectively and reduce some limitation of KWB search engines.

In this chapter we reviewed aspects of internet search such as the importance of information retrieval tools due relentless growth of information over the few decades and growth of internet users. We determined that Google importance in information retrieval is still huge and stable.

Also, reviewed terminology of *information* and *knowledge* presented by different authors and concluded that information is part of exact science which can be described as a data which can be processed by machines and simultaneously it can be a process of a decision-making. Then *knowledge* can be determined as information integrity to get understanding of meanings to make a decision".

We used Thierauf (1999) summarization levels to provide levels of information processing in which *data* takes the biggest part and is in the lowest level, information is in a second level and *knowledge* is the smallest part which is a highest level of information processing.

We determined evolution of information retrieval systems to search engines and pointed out the major events of evolution.

Also there were four areas determined to evaluate search engine effectives and limitations of current KWB search engines were set. So effectiveness of search engine can be determined by four factors like index quality, quality of the results, quality of search features and search engine usability. And limitations of search engines could be determined as flux of information, quality of information must be determined by user, poor objectivity to direct questions, uncertainty of result existence, retrieval of explicit content and copyrighted material, simplistic interface confuses users in result relevancy and rankings of list sometimes makes information retrieval invalid.

Finally, we determined that by enabling Semantic Web Technology tools and agents we could make search engines more efficient with less limitations rather than KWB search engines have and also to enable agents to serve for people by making sophisticated tasks.

# 2 EVALUATION OF SWT IN "DIFFUSION OF INNOVATIONS", "CROSSING THE CHASM"AND "MATHEMATICAL THEORY OF COMMUNICATION" THEORIES

In this chapter we going to review SWT in "Diffusion of Innovations", "Crossing the Chasm" and "Mathematical Theory of Communication" theories to analyze, on how innovativeness dimension (innovation adoption cycle) can be applied to SWT technology, how "The Chasm" (gap) could influence SWT innovation adoption and how "Mathematical (information)" model could be applied to SWT.

## 2.1 Evaluation of innovativeness dimensions in SWT

Joo (2011) made a research about adoption of SWT in companies in which he refers that SWT as technology innovation and these technologies can be examined as it how it affects the adoption and diffusion.

According Rogers (2003) one of technology innovation evaluation criteria can be determination of companies or individuals usage of innovations and these companies or individuals can be classified by innovativeness dimensions (see Figure 4). There are five innovativeness dimensions (or innovation adoption cycles):

- Innovators – companies or individuals who seek to be the first one to adopt innovation, even though it can be inadequate to most. They can be called pioneers of innovation adoption;
- Early Adopters – companies or individuals who adopts innovation after its been explored by innovators;
- Early Majority – the first wave of companies or individuals who adopts innovation because of success of adopting and applying it by Innovators and Early Adaptors;
- Late Majority – the second wave of companies or individuals who follows into footsteps of Early Majority to adopt innovation which give benefit to them.

- Laggards – the last wave, which don't follow any previous companies or individuals and adopts innovation because its widely accepted and has its own benefits or just they have no other option to adopt other suitable innovation.

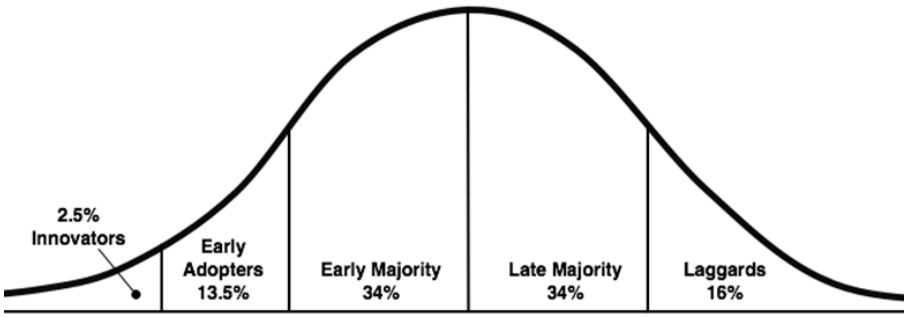



**Figure 4.** Adopter categorization on the basis of innovativeness (innovation adoption cycle), Rogers (2003)

Later in the research this method will be used to determine at which innovativeness dimension companies, businesses and government currently are by adopting SWT as an innovation.

## 2.2 Evaluation of SWT on crossing "The Chasm"

According Moore (2007) the innovation adoption cycle has its own flaws, because there are three gaps between cycles, which can lead to adoption of innovation failure, but there is one particular gap, which is wider than other two, so this gap is called "The Chasm" (see Figure 5). This gap is in between Early Adopters cycle and Early Majority cycle, on which there is a high risk of innovation failure.

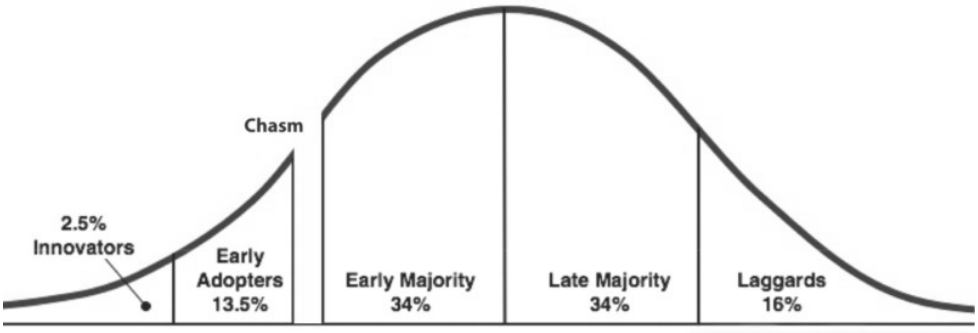



**Figure 5.** The Revised Technology Adoption Life Cycle. **Source:** Moore (2007).

Later in the research this method will be used to determine if SWT already crossed "The Chasm" or its going to do it in the future.

## 2.3 Evaluation of "Noise Source" in SWT

According Shannon and Weaver (1964) to receive a message, it should cross over channels to reach final destination. As we talked earlier, search engine could be defined as a medium to reach information on the internet. But in order to receive this information there could be some obstacles which can interfere with the information that user receives. It can change information slightly or it can make a real damage to it. In Shannon and Weaver (1964) communication system model, there is particular obstacle shown, its called "Noise Source" (see Figure 6).

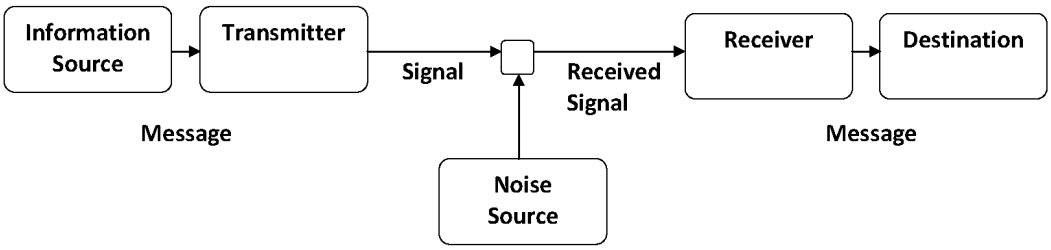



**Figure 6.** Communication system model. **Source:** Shannon and Weaver (1964).

In order to understand this "Noise Source" in SWT later in the research we will determine it according to experts from SWT field.

To sum up this chapter we can say that SWT as an innovation can be evaluated by using communication theories like "Diffusion of Innovation", to see at which innovativeness dimension (technology adoption cycle) it stands; "Crossing the Chasm" can show if this innovation already crossed "The Chasm" or it going to cross it over time, or maybe fail it; "Mathematical Theory of Communication" theory can show what is "Noise Source" in communication system model. These question will be determined after the interviews with experts of SWT.

# 3 SCIENTIFIC RESEARCH "APPLICABILITY OF SEMANTIC WEB TECHNOLOGIES IN INTERNET CONTENT SEARCH FOR EDUCATION" METHODOLOGY

This part of work represents methodology applied to research "Applicability of Semantic Web Technologies in internet content search for education". For implementation of research was chosen qualitative research. Chosen strategy was implemented to do generation of the theory rather than just test is, like quantitative strategy does and the data selection in quantitative strategy would not produce the deeper insights of the research rather superficial overview of the topic so quantitative strategy was rejected by the author.

The aim of study is to evaluate applicability of Semantic Web Technologies to education and to evaluate student abilities for information retrieval from internet and databases.

First of all, doing qualitative research will help to get a deeper understanding of what SWT could give as a tool for businesses, government, personal and professional lives etc. and also how it can make improvements in work and life and education related tasks. And second of all, this research will give a deeper understanding about current situation of student's abilities to retrieve information from Internet and DB using content search tools.

**Research methods.** To achieve the purpose of this research, expert interviews were selected as method of qualitative research and factual interview as a form of the interviews (see Figure 7).

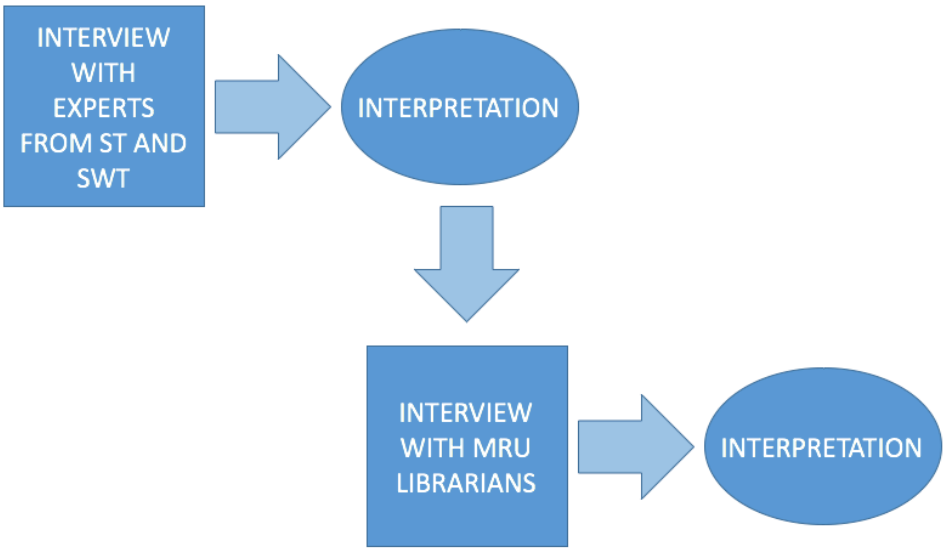



**Figure 7.** Experts and librarians interview method design. **Source:** prepared by author.

According Kvale (2007) sometimes interviews focusing not on interviewees personal perspectives, but on his experience with valid factual information interviewee has. In our case we were interviewing experts of Semantic Technology, so they can provide us information and insights on how Semantic Technologies and Semantic Web could change our perception to information retrieval and its application.

According to Hennink, Hutter and Bailey (2011) to achieve insider's perspective which is a characteristic of qualitative research and perspective, interview should involve these elements:

- Use a semi-structures interview guide to prompt the data collection;
- Establish rapport between the interviewer and interviewee;
- Ask questions in an open, empathic way;
- Motivate the interviewee to tell their story by probing (Hennink, Hutter and Bailey, 2011).

So these methods suggested by Hennik, Hutter and Bailey (2011) were used to get a relevant interview with experts.

Also Kvale (2007) suggests seven stages of an interview inquiry to follow which were followed in our research:

1. *Thematizing.* At this stage purpose of investigation was formulated which is "our research is to get in depth view and understanding how Semantic Web could be integrated in businesses and university studies and what outcomes could be after making this integration".

   Kvale (2007) at this stage suggests to answer these key questions:

   - *why*: to clarify purpose of the research. The purpose is at the beginning of this chapter;
   - *what*: to obtain the pre-knowledge or the field of subject we going to investigate. At the beginning of work, at the preparation stage, the investigation of the topic was made by studying books and reports of Information retrieval, Semantic Technologies and Semantic Web to get a basic knowledge on of situation dealing with and expecting to get deeper on knowledge after interviews were done.
   - *how:* to understand types of the techniques of interviewing and how to use them when the application of intended knowledge.

2. *Designing.* According Kvale (2007) taking all these seven steps of investigation is the designing of the research, so all these steps were taken into account preparing and implementing researches.

3. *Interviewing.* At this stage there were two parts of interviewing made, at first part there were experts of Semantic Technology interviewed. Most of the interviews where made using video conversation software "Skype", so participants of this research were included from different

parts of the world. Also one of the participants requested to make an interview using email correspondence due to differences in world time. Second part of interviews were made with Mykolas Romeris University librarians who are responsible for teaching academic society on how to find information and recourses for studies. One interview was made face-to-face and others using email correspondence. For all these interviews interview guide was used (see appendices) with a reflecting approach and interpersonal relations of the interview situation.

4. *Transcribing.* After every interview was made, the transcript of the interview was prepared for analysis. Interviews which were made face-to-face or using video conversation software "Skype", audio records were made and after the interview oral speeches were made to written text using software for audio and video transcribing. Interviews which were made by email correspondence were collected to one document and prepared for analysis.

5. *Analysing.* Even though Kvale (2007) highlight that there is no particular method exist to analyse gathered data but few modes of analysis there suggested. According to research we making "interview analysis as theoretical reading" were selected due it's description of analysis which indicated that no specific systematic analytic tools to decrypt interviews needed but using this mode we can get new knowledge of the subject we explore through analytical approach of gathered interview data.

6. *Verifying.* In order to verify gathered and analysed information from interviews all data must to meet the criteria of reliability and validity. All data gathered through interviews are reliable and valid on its consistency and trustworthiness and it meets criteria of ordinary language to the truth because no preconceptions and relationship with all participants were formal and questions were formulated nondirective, no "right answers" were suggested or directions of answering were given.

7. *Reporting.* As an author of the book suggest that final stage of the research should be communication the findings on the research so this thesis is the report of findings to the society which lives up all scientific criteria's which are established by University and academic rules.

According to Saldana (2011) method of interview is effective way to get insights, different personal perspectives, attitudes, feelings, opinions to factual information. "Qualitative research design is best approached provisionally since this mode of inquiry is emergent and evolutionary in its process" (Saldana, 2011). This method is chosen to investigate Semantic Web Technologies in internet content search to get insights from experts of Semantic Technology and MRU librarians about its evolution and possible integrations to field in education field.

**Data collection instrument.** Data collection instrument is semi-structured expert interviews. During interviews in both parts open ended questions experts were asked and interview guide was

used to keep structure in all of the interviews (see appendices). The interview guides questions were prepared before the interviews. The second part interview guide questions for the MRU librarians were prepared after making first part interviews with experts of Semantic Technology. All same questions were given for the experts but due specifications of conversations were made some additional questions to get more accurate or complete answers. Similar situation was occurred during second part of making interviews with MRU librarians.

During first part of interviews with experts of Semantic Technology there 21 open ended questions were asked of which 9 questions were about functionality and effectiveness of Semantic Technologies and Semantic Web Technologies, 5 questions were about applicability of Semantic Web Technologies in business area, 4 questions were asked about applicability of Semantic Web Technologies in personal and professional lives, 2 questions about what experts think on Semantic Web Technologies through Everett Rogers communication theory "Diffusion of Innovations" and 1 question about Mathematical (information) theory about "noise" in these technologies.

During second part of interviews with Mykolas Romeris University librarians which are making lectures to students and academic society on how to find information for studying, 6 open ended questions were asked of which 3 questions were about specifics of students and academic society making on information retrieval and recourses findings from Internet and data bases, 2 questions about limitations of information retrieval and recourses findings and 1 question on lectures applicability for other fields of interests.

**The research sample.** For this qualitative research the purposeful (in other literature - purposive) – experts based sampling method was used. According Bogner, Littig and Menz (2009) and Seidman (2006) there is no particular number to define how many participants should be interviewed to get a representative sample of experts involved in the research firstly because there is no definite number of experts known who works in the particular sphere.

For conducting first part of research with experts of Semantic Technology online research was done to find participants for interviews. Approximately 500 contacts were found online including Semantic Technology company's CEOs, personnel, University Professors, creators of Semantic Web Technologies journals, etc. Over 2000 emails there sent with invitations to interviews in period of time 6$^{th}$ of November 2016 – 10$^{th}$ of March 2017 (see letters in appendices). 4 experts agreed to participate in interview from two continents – Europe and North America. All of them are experts in different branches of Semantic Technology with different experiences in this field including many years of experience and positions and responsibilities in companies they working for. One of the participants is data strategist in company which is leading company creating software standards for Internet, another is Semantic Technology company CEO, also one of the participants is CEO and

principal consultant who is currently working with Semantic Technology software development for healthcare company and his also teaching younger generation experts about software development. Last participant is an Assistant Professor in the Computer Science and Engineering Department who is mainly making researches on Semantic Web Technologies topics. All of them are experts in Semantic Technology department but also they have different perspectives on working with it. All interviews which were made using video conversation software "Skype" took approximately 45 minutes. Table 3 present criterion sampling for the first part of interviews.

**Table 3.** Criterion sampling for the first part of interviews with experts of SWT.

| Expert No. | Position/ responsibilities | Experience in Semantic Technology field | Form of interview | Date (duration of interview) | Country |
|---|---|---|---|---|---|
| **1** | Data Strategist, Activity lead | 17 years | "Skype" video conversation | 16th of December, 2016 | United Kingdom |
| **2** | Assistant Professor | 16 years | Email Internet interview | From 5th of March, 2017 to 29th of March, 2017 | United Stats of America |
| **3** | CEO, software developer | 15 years | "Skype" video conversation | 15th of March, 2017 | Poland |
| **4** | CEO, Principal consultant, technology strategist | 16 years | "Skype" video conversation | 15th of March, 2017 | United States of America |

**Source:** prepared by author.

For conducting second part of research with Mykolas Romeris University librarians there was short online research was done. Mykolas Romeris University librarians contacts were found on University web page. All the librarians who are making lectures to academic society about information retrieval and references finding are called Subject Librarians. All of them are experts in

field on how correctly retrieve information from Internet and data bases, they are teaching on where you need to go to retrieve information, how to make better queries for searching, what functions to use to filter your results of many recourses not needed. There were 5 librarians who participated in interviews, 1 interview were done face-to-face and 4 were made by using email. Table 4 present criterion sampling for the second part of interviews.

Table 4. Criterion sampling for the second part of interviews with MRU librarians.

| **Expert No.** | **Position/ responsibilities** | **Experience working in library at MRU** | **Form of interview** | **Date of interview** | **Duration of interview** |
|---|---|---|---|---|---|
| **1** | Chief Librarian, Subject Librarian | 15 years | Face-to-face | 5th of April, 2017 | 30 minutes |
| **2** | Senior Librarian, Subject Librarian | 6 years | Email internet interview | From 5th of April, 2017 to 7th of April, 2017 | - |
| **3** | Senior Librarian, Subject Librarian | 5 years | Email internet interview | From 5th of April, 2017 to 7th of April, 2017 | - |
| **4** | Senior Librarian, Subject Librarian | 5 years | Email internet interview | From 5th of April, 2017 to 7th of April, 2017 | - |
| **5** | Senior Librarian, Subject Librarian | 13 years | Email internet interview | From 5th of April, 2017 to 7th of April, 2017 | - |

**Source:** prepared by author.

**Data collection process.** For first part of interviews all participants were found by making online research. Emails with invitations were sent. With participants who agreed to participate in interview the date and method were mutually agreed personally due participants busy work schedules and different world time. Most of the interviews were made by video conversation software "Skype" so every participant (interviewer and interviewee) choose their own comfortable surroundings for interview which were quite and most comfortable to make suitable conversation.

For second part of interviews all participants were found by making short online research by finding contacts on University web page. Face-to-face interview was arranged to do it in University classroom so it would be comfortable place to University librarians and researcher.

**Research ethics.** According Yow (2005) release form can give a security for two reasons. First, researcher keeps all rights for making publications and presentations of information he or she conducted in a research process and second release form give protection in lawsuit if interviewee suddenly decide to make a lawsuit for publishing results and conclusions of the research. Also Yow (2005) says that all participants must be free consent and their participation in the research must be voluntarily.

So taking into account that this research at some point its going to be published, terms, conditions of this research were presented and and release forms were signed (see appendices). All participants agreed to make this research voluntarily and with free consent.

Another part of ethics is the identification of the participants of the research that should be highly considered as it can have negative effects from public as the results of research going to be polished publicly. So in this research anonymity and confidentiality of the first part research experts form Semantic Technology companies been taken into account and even few participants agree to be disclosed using their personal information, but anonymity was followed and no participant names or company names they are working for were not mentioned in research due its possibility of secret information disclosure in the research. For the second part of the research anonymity of participating persons was followed to but the work place there disclosed due to type of a research is made for University they are working for.

**Research limitations.** For the first part of the research limitations could occur in using video conversation software "Skype". Due poor internet connection quality of audio and video decreases so the quality of the record decreases too and misheard words could occur. Also limitation could be in the differences in English language due different word pronunciation, accents and levels of English language understanding could cause uncertainties. Limitations could occur in terms of technical vocabulary and terminology due researcher not being expert in Semantic Web Technologies field, specific words or terms could give different meanings to different conditions. To surpass these

limitations of research there been action taken such as asking for expert to repeat unknown words or make a brief explanation on it to reduce obscurity so these limitations would not make severe affect to the outcomes of the results.

For the second part of the research limitations could occur in making transcripts due making interpretation from Lithuanian language to English language. These limitations of interpretations could be determined as a minor or not severe so they would not affect the outcome of the results.

## 3.1 Qualitative data analysis results

According Seidman (2013) making qualitative interviews generates a huge amount of textual information, so in order to get most significant information from transcripts which is in these words and paragraphs must be reduced. Very important thing in reducing these words is the attitude of the researcher, that is researcher must come to transcript reduction with an open mind, to make it with no prejudice, bias or even anger in case of conflicts of the opinions. Seidman (2013) also highlights that reduction of text in the research must be done by the researcher inductively rather than deductively. The interviewer must come to the transcript prepared to let the interview breathe and speak for itself (Seidman, 2013).

Saldana (2011) says that interviews, especially improvised conversations can lead up to new understandings and insights from which there can be new inquires formed and analytical follow-ups may occur.

### 3.1.1 Insight from the interviews with experts from Semantic Web Technologies field

In this subchapter we will make evaluation of Semantic Web Technologies. To make evaluation we classified questions into five sections which contains subsections:

- Functionality and Effectives of SWT;
  - Purpose of SWT;
  - Differences of KBE and SWT based search engines;
  - Advantages of SWT;
  - Disadvantages of SWT;
  - Effectiveness of SWT;
  - Reliability (accuracy) of SWT;
- Development of SWT;

- SWT becoming mainstream technology;
- Challenges of SWT;
- Issues of SWT;

- Application of SWT to business and government;
  - Business interest and use of SWT;
  - Government interest and use of SWT;
  - Benefits of SWT for business and government;
  - Limitations of SWT for business and government;
  - Business and government engagement in development of SWT;
  - Profit for business of using SWT;
- Application of SWT for personal and professional lives and education;
  - Advantages of applying SWT to personal and professional lives;
  - Disadvantages of applying SWT to personal and professional lives;
  - Application of SWT in education;
- SWT through "Diffusion of Innovations" and "Information" communication theories;
  - Technology Adoption Cycles of ST;
  - ST on crossing "The Chasm";
  - "Noise" in ST.

**Purpose of SWT.** According experts purpose of these technologies is to enable machines to "understand" the context of queries and texts as people do it. Also Expert No. 4 mentions that Semantic Web Technologies is only a tool to enable searching rather that a search engine itself.

Segments from interviews with experts:

*"<…> purpose is to organize information, purpose to make sense of the world <…>". "<…> form engineering point of view it's about helping to use the web to organize the world's information"* – Expert No. 1.

*"To allow applications and machines to "understand" and reason about data as humans do (though not necessarily in the same way), such that they can fulfill their purpose more effectively".* – Expert No. 2.

*"<…> to create the way to make knowledge to be executable by machines"."<…> we use this knowledge to communicate and to make some decisions, however, if we are able to express this*

*knowledge in formal way, for example using Semantic Technologies", then computers will be able to execute this knowledge".* – Expert No. 3.

*"The key thing for me is that I'm using Semantic Web Technologies primarily to increase my search relevance". "Semantic Web to me is not a search engine, it is a tool to only enable searching".* – Expert No. 4.

**Differences between Keyword-based search engines and SWT.** According experts differences are in searching same information in two totally different ways: Keyword-based search engines are using keywords form queries created by the user and Semantic Technologies overpasses these keywords by understanding the meaning of query.

Segments from interviews with experts:

*"<...> traditional web search engines are more reliant on basic natural language processing strategies and things like PageRank, while semantic web search engines attempt to make use of implicit knowledge gleaned through ontologies, data structure and links <...> and other background knowledge".* – Expert No. 2.

*<...> differences are visible only in approach to some text processing they use <...> in standard way <...> the document index is keyword based so if you specify set of keywords the search engine provides you with set of results that fits all that results which contains these words or was based on a specified query". "Semantic search is all about trying to deeply understand the context of the text".* – Expert No. 3.

**Advantages of SWT.** Experts on question about advantages of Semantic Web Technologies have different opinions, for example for Expert No. 1 the main advantage is ability to link information one to another, for Expert No. 4 the main advantage is for companies which use this technology gets benefit in optimization of search engines, which leads to economical value for company. Both Expert No. 2 and Expert No. 4 see advantage in higher precision and recall of information which is being searched.

Segments from interviews with experts:

*"The ability to link one thing to another, you get the network effect <...>"* – Expert No. 1.

*"<...> semantic search engines are capable of higher precision and recall"* – Expert No. 2.

*"<...> semantics is something that is very broad category and for search it is turned out to be very important economically because of search engine optimization". <...> it increases the precision and relevancy of my search results."* – Expert No. 4.

**Disadvantages of SWT.** There are some disadvantages of technologies and experts identify them as lack of semantically annotated documents which affects the effectiveness of them and another disadvantage is that Semantic Web Technologies are not for general users, who enables searching, at least now, so people who don't know basics of these technologies, they can't work and develop it as tool for public users, who are creating queries to get information.

Segments from interviews with experts:

*"<...> there is not yet enough semantically annotated data available on the web to achieve the "critical mass" necessary to make semantic search engines truly effective"* – Expert No. 2.

*"<...> complexity and it's not general purpose tool at all right now <...>"* – Expert No. 4.

**Effectiveness of SWT.** Evaluating effectives of Semantic Technologies and Semantic Web Technologies all experts agree on the fact that time is reduced by searching process itself, because user don't need to spend time on creating queries to match the keywords to find information that they are looking for. Expert No. 2 adds up by saying that user don't need to use search tools in particular situations, search would be conducted by agents who would interpret all possible results and the best direct answer would be presented according to user's personal interests and needs.

Segments from interviews with experts:

*"<...> adding "schama.org" to the web page <...> enforce you to make your web page in particular way <...> but once you done that process, it works. So they kind of nagging you to organize your data in more machine friendly way, and that is certainly having an effect, so the web is more intelligible by machines and more processable by machines than it was <...>* - Expert No. 1.

*"<...> talking about semantic web search, I imagine that we really talking about "Schema.org" that it doesn't make it faster, it does it more accurate"* – Expert No. 1.

*"<...> by greatly improving the precision of search results and therefore limiting the time the human must spend sifting through the results". "<...> potential of semantic web technologies to eliminate the need for user-directed search in some instances, by recognizing the need to conduct a search, doing so, interpreting the results, and taking the appropriate action independently <...>"* – Expert No. 2.

*"However, they make search faster from the perspective of the search engine user, because he or she can access the most interesting knowledge quicker without using so many queries".*

*"Context is very crucial here and right now there is deep research in this field going on to how we can connect Semantic Technologies with context discovery of the user <...>"* – Expert No. 3.

*"That result what you looking for usually comes as a first page, so you never have to pagination. So in the sense that it speeds up because you don't have go looking through and rerunning your course, but it wont make specific search faster, it will just change relevancy rankings".* – Expert No. 4.

**Reliability (accuracy) of SWT.** Finally, all experts agree on question about results produced by searching with Semantic Web Technologies that these results are reliable. Expert No. 4 claims that people who uses Semantic Technologies is not tied to keyword searching.

Segments from interviews with experts:

*"<...> most uses graph technology <...> and "schema.org" is a massive addition to the knowledge graph <...> it allows distributer and anyone to add that information at scale <...> it's on billion web pages <...> and by just embedding little bit of data, you massively increase the accuracy of these results.". "It's an accuracy rather than reliability"* – Expert No. 1.

*"Currently I see the difference as relatively slight, but the potential for greatly improved performance is definitely there"* – Expert No. 2.

*"It always does make search relevancy better if it's implemented correctly, because you not tied to keyword matching. You where doing two things: you're classifying all documents using a taxonomy so every document has a list of concepts related to this document based on all the words in it and then you have keyword you typing these keywords, you expand those keywords to concepts. So instead of matching low level concepts on low level keywords you're enriching both the query and the documents with concepts and then you're matching concepts"* – Expert No. 4.

**Development of SWT.** Evaluating development of Semantic Web Technologies all experts agree on that this technology is not a mainstream, but it has a potential to become one of them, Expert No. 2 determine that these technologies should become mainstream technologies in period 5-10 years.

Segments from interviews with experts:

*"<...>when a technology is still in the phase in which people debate fundamental things like "what is the meaning of this term?", as in, "what is the meaning of the term 'semantic technology'?" then it cannot yet be considered mainstream". "I think we are relatively close to achieving this (say, in the next 5-10 years)"* – Expert No. 2.

*"<...> they are still considered to be mainstream technology but in 10 years"* – Expert No. 3.

*"<...> it could become "a mainstream technology, but not today"* – Expert No. 4.

**Challenges of SWT.** Experts explains that the biggest challenge that they facing working with Semantic Web Technologies is in people training to work with front-end of these technologies. Expert No. 1 expressing his feelings that people has bad prejudices about these technologies. Expert No. 2 and Expert No. 4 say that another challenge is in tools, which are not suited to work with real world applications and Expert No. 3 and that there are some theoretical problems to deal with. Also, Expert No. 2 discern the issue in Semantic Web Technologies that there is a lack of published data and links between this data to make this technology become mainstream.

Segments from interviews with experts:

*"Prejudice. People say "oh that's complicated Semantic Web stuff, I'm not touching that"* – Expert No. 1.

*"Many existing tools are still at the research prototype level, and are therefore not robust enough to use in real-world applications". "I think one issue is that volume of published data and links between different data sets has not reached the "critical mass" necessary for the utility of semantic technologies to become obvious".* – Expert No. 2.

*"There are several theoretical problems <...> one set of problems – computational problems. In my opinion second problem is related to tough threshold of accessibility. People need to be learned how to use formal language before they can use it"* – Expert No. 3.

*"I think the biggest challenge I have is the infrastructure. I wish more tools supported RDF and SKOS, the interfaces between those <…>". "The training is a huge concern, especially than people don't have the ability to get light small tools they can build and test themselves <…>". "We don't have enough standards like that, we need more and those need to be widely adopted and supported by all these different vendors". "Everybody want to use everybody else data but nobody wants to pay them to put that data there. So we have a lot of problems with building enough bridges between ontologies, we don't have enough of the ontology builders and there's no economic models for that to get"* – Expert No. 4.

**Business interest and use of SWT.** Evaluating application of Semantic Technologies and Semantic Web Technologies to business expert agree on that businesses are generally interested in adopting these technologies to their production. Also all experts suggest that these technologies are for companies who deals with huge number of data and documents, managing this data and documents to get a meaning of them, which regular person can't manage do this. Expert No. 1 mentions that these technologies were used by researches to find information and to connect dots in leaked documents of financial frauds called "Panama papers". Expert No. 3 say an example that these technologies would give advantages for manufacturers and telecommunication companies.

Segments from interviews with experts:

*"Panama papers is an example there was using it via 4G <…>"* – Expert No. 1.

*"I feel that business, led by initiatives like IBM's Watson and Google's Knowledge Graph, are in many ways expressing interest in Semantic Technologies. These companies have been hiring top researches in the field for several years now". "A very wide variety of businesses and other organizations are interested in Semantic Web Technologies, from retailers who are trying to use semantic technologies to more effectively recommended additional products to their costumers, to scientific organizations who are trying to merge scientific findings from different fields, to law enforcement agencies that are trying to track the behavior of individuals suspected of violence".*

*"<…>professions that require decision making based on a large amount of data that may be spread across different locations "*– Expert No. 2.

*"The main interest is from medical and pharmaceutical domain because information about medical stuff is gargantuanly heterogeneous <…>". "<…> mostly there were you have some set of documents and you need to process them. For example, manufacturing and telecommunications"*- Expert No. 3.

*"Businesses and government have business problems. For example, in the healthcare project that I'm working, their problem is that they get many calls in the call center". "Any company and organizations with many numbers of documents they need semantics to help find meaning of those documents"* – Expert No. 4.

**Government interest and use of SWT.** Expert No. 4 says that "Businesses and government have business problems", so these technologies could be applicable for government too. Experts mentions that Semantic Web Technologies could be applied in museums, galleries, libraries, life sciences, media and gives some examples where it is currently applied, like British document legislation web page, Eurostat, European Data Portal, "OpenPHACTS, BBC and others. Expert No. 2 propose that government in United States in America not interested in these technologies anymore, while Expert No. 3 says that they participated in fraud detection system development with governmental company form United States.

Segments from interviews with experts:

*"So, come across "legislation.gov.uk". This is a puzzle that give you an access to every bit of legislation pass in Britain <…>". "<… uses of SemWeb Technology include cultural heritage so all the big museums use it, library's <…> and galleries who are simply using it to link creators to works". "On governmental site "Eurostat" it's just doing more and more <…>". "<…> European Commission are big supporters of linked data. European Data Portal gathers information together from lots of different portals and makes it available for using it". "In life science you may want to look to "OpenPHACTS". "And another big user of linked data is the BBC"* – Expert No. 1.

*"The interest from United States government has not kept pace with that of industry. It does not seem to be key focus area for them <…>". "I have the impression that the European Union government has more interest in that field, however"* – Expert No. 2.

*"<…> form government perspective it is true as we developed system together with company form United States for fraud detection <…>"* – Expert No. 3.

**Benefits of SWT for Business and Government.** Talking about benefits of Semantic Web Technologies experts say that it allows to people work more effectively and produce knowledge in shorter periods of times and it can lower costs by reducing human recourses in production.

Segments from interviews with experts:

*"Semantic Web Technologies allow people and organizations to make more effective use of data in a shorter amount of time. Additionally, they can allow people to "connect dots" or see relationships between data that may have been difficult to notice <...>"* – Expert No. 2.

*"Major benefit is to have executable knowledge so you can allow computers to do some cognitive analytical tasks by machines <...>"* – Expert No. 3.

*"<...> users will be able to find what they are looking for in websites and call center <...>". "<...> they can lower the cost, increase their precision of getting the results that users need them <...>".* – Expert No. 4.

**Limitations of SWT for Business and Government.** Limitations according Expert No. 2 and Expert No. 4 to these technologies are that expectations of companies applying these technologies are to high and companies who provides these technologies over-sold the idea of it. Other limitations expressed by experts are related to infrastructure and integration of these technologies to current environment and tools to achieve it.

Segments from interviews with experts:

*"<...> for Semantic Web Technologies to be useful, data must be published in a standardized format and linked to other data, and the tools to achieve this are currently somewhat limited". "I think semantic technologies may have been "over-sold" as a panacea for many problems at first, and when the early promises did not result in quick results, the field fell someone out of favor with funding agencies"* – Expert No. 2.

*"<...> there is a problem that you need to have an infrastructure, you need to build system, you need to integrate these products with your current environment"*- Expert No. 3.

*"<...> many companies promise more than they can deliver. <...> They don't hire people with enough training to manage those taxonomies <...> and then the searches don't work". "<...> second risk is feedback cycles. <...> they didn't look at the search logs and every week they're getting this information about searches that failed <...> lack of feedback in these things often stopping progress of projects from getting better". "<...> we often stopped if a project is to expensive <...>"* – Expert No. 4.

**Business and Government engagement in development of SWT.** According experts, businesses and government are engaged in development of Semantic Web Technologies and Expert No. 2 notes that investments are increasing.

Segments from interviews with experts:

*"<…> in the commercial world they tend to not talk about how they doing stuff. They simply offer you off-the-shelf solutions and so be wary of assuming that because it isn't in the public eyes, they are not using it"* – Expert No. 1.

*"Businesses and government are key sources of funding for researchers working in the field, and they are the source of applications and "challenge problems" that drive important advances in the field. Additionally, the Semantic Web inherently involves information from many different sources. Government projects frequently bring individuals from different fields together to try to tackle the challenges of this type of multi-disciplinary effort". "Investment on the part of these organizations in Semantic Web Technologies seems to be increasing"* – Expert No. 2.

*"<…> government agencies, US National Information Center on Health Services Research and Heath Care, National Health Care databases, Center od Medicare and Medical Services, Snomed CD, big one – RxNorm for drugs and diseases, they do publish ontologies and taxonomies"* – Expert No. 4.

**Profit for businesses of using SWT.** All Experts agree on terms of making profit by companies who uses these technologies.

Segments from interviews with experts:

*"There are companies that they are providing commercial services that use Semantic Web Technologies <…> they using it because it happens to be good for what they want to do"*- Expert No. 1.

*"My feeling is that some companies, like Amazon for example, even small improvements in Semantic search could result in a noticeable impact in the company's bottom line"* – Expert No. 2.

*"After you invest in implementation of these technologies at the end of the day you can limit your costs because the part of tasks can be used directly by computer system. They can compute it in the similar way that people can do this"* – Expert No. 3.

**Advantages in applying of SWT for personal and professional lives.** Evaluating application of Semantic Technologies and Semantic Web Technologies to personal and professional lives advantages of it according experts would be in time trying to find information, it would be reduced and people could focus on decision-making.

Segments from interviews with experts:

*"<...> will allow humans to spend less time on data search, acquisition, and integration and more time on high-level knowledge synthesis and decision-making* – Expert No. 2.

*"We will have an access to specific knowledge in faster way in general and we will benefit form it that computer systems will be able to attract more complex problems that we can't. In my opinion advantage of Semantic Technologies is intelligence amplification, it is sometimes also called Cognitive Enhancement of human being"* – Expert No. 3.

*"People will depend more and more on chat bots, now people are using personal assistants like Siri from Apple or Alexa from Amazon. People control them by voice and I think that things controlled by voice become a thing <...>"* – Expert No. 4.

**Disadvantages in applying of SWT for personal and professional lives.** Expert No. 4 notes that people will be dependent on chat bots and voice controlled assistants for example, Siri or Alexa, but also he says that these assistants are not Artificial Intelligence, at least for now as a disadvantage of these technologies. Also experts highlight that this dependency on computer systems could become big issue because people could stop thinking for themselves and all decisions could be made by computers.

Segments from interviews with experts:

*"<...> over-reliance on the semantic web technologies to provide "all" of the relevant information and accurately assess the trustworthiness of the underlying data"* – Expert No. 2.

*"People will become dependent on computer systems. <...> they will stop thinking on their own <...>"* – Expert No. 3.

*"But I can't say that these agents are applications of Artificial Intelligence, at lest not yet".* (talking about agents like Siri and Alexa – remark of the author). – Expert No. 4.

**Application of SWT for education.** Experts say that Semantic Web Technologies could help in education process by fastening learning process by giving you reliable information in short periods of time with large amount of data. Expert No. 2 say that these technologies would enable scientists to share huge amount of information to share over the world.

Segments from interviews with experts:

*"Semantic web technologies have the potential to allow experts from two different domains <…> to share information in a meaningful way <…>". "They can enable students to work on multi-disciplinary projects involving large amounts of data"* – Expert No. 2.

*"Applications can appear that can be used in form of a fast feedback loop. In other words, we can create systems that during the learning process will return to you with the correct answer and it will boost your learning processing"* – Expert No. 3.

*"I think that students at class can get faster knowledge. They can find information they been asked faster and with bigger reliability"* – Expert No. 4.

**ST in "Diffusion of Innovations" theory. Technology Adoption Cycles of ST.** Evaluating Semantic Web Technologies through "Diffusion of Innovations" communication theory experts have different opinions about this question, experts find these technologies on Early Adopters and Early Majority Cycles, because only a minority of companies using these applications and there is lack of proof on usefulness of these technologies.

Segments from interviews with experts:

*"<…>business and government are in the Early Adopters phase". "There are a few companies, such as IBM and Google that are investing in semantic technologies and have products or services that use them, but these companies are still in the minority, and semantic technologies are not central to the income stream of many companies"* – Expert No. 2.

*"Semantic Technologies are in Early Majority mostly because of lack real life cases that can be proven to be beneficial from using it."* – Expert No. 3.

*"I think that in medical and financial businesses Semantic Technology is definitely in Early Majority Cycle, because these businesses integrating Semantic Technology in their companies, but still it's not a majority of them doing it. Other businesses are on Early Adopters Cycle"* – Expert No. 4.

**ST on crossing "The Chasm".** Opinion of experts about ST crossing "The Chasm" differs on ether Semantic Web Technologies crossed area of so called "the Chasm" gap in Adoption Cycles or not. Expert No. 2 and Expert No. 4 claims that "the Chasm" is not crossed, but they believe that it can be done, and Expert No. 3 says that "the Chasm" is recently crossed.

Segments from interviews with experts:

*"I think this is another aspect of the lack of a "critical mass" of data usable by Semantic Technologies being readily available, which can be overcome by better tool support for data publishers"* – Expert No. 2.

*"I think it crossed recently and the reason is the BigData. The size of archives is not accessible for "standard" technologies but we need to understand them somehow. That's why, even if no many people understand semantic technologies they willing to give them a try"* - Expert No. 3.

*""The Chasm" could be crossed, but not yet. I think that medical and financial businesses crossed it, but others still not and we need to wait to see what's going to happen"* – Expert No. 4.

**ST in "Information" theory, "Noise" in ST.** For question about the "Noise" in "Information" communication theory, experts identify different possibilities about appearance of "Noise". Expert No. 2 thinks that "Noise" could occur from errors made by developers adopting technologies, Expert No. 3 thinks that it can appear in barriers of language, that machines could not process common language that people use and Expert No. 4 thinks that errors could occur in information transmission from old systems to new systems.

Segments from interviews with experts:

*"<…> might include inaccurate data, data presented without context or links between data that are valid only in one context, inconsistent data, syntax errors such as spelling mistakes, inconsistent use of ontologies, etc."* - Expert No. 2.

*"The noise can be easily interpreted as in terms of so-called "common language". We know how to process formal text to some extend but if we want to process "common language" we deal with sarcasm, context dependence"* – Expert No. 3.

*"I think that noise could be then there is processes of Semantic Technology adoption form old systems. It's called "loss of fidelity in data transportation" or "loss of lineage"* - Expert No. 4.

Overall, making conclusions from interviews with experts about SWT, we can confirm that functionality and effectiveness of SWT is absolute and it is clear that these technologies have definite preeminence over the Keyword-based search engines.

Also we can conclude that there is need of further development of SWT tools, a need to overcome challenges in teaching back-end users to work with these technologies and a need of more published data which would be linked one to another which would lead these technologies to become mainstream.

Also, we can propose that business and government are generally interested in investing into development of SWT and applying them to their companies even though there are some limitations in lack of infrastructure and high expectations to resolve all company problems by using them, but making benefits in production time, work effectiveness and knowledge production in a way that people can't do overcomes the limitations by making profit using these technologies.

And finally SWT can help to fasten process of learning and making researches for science by enabling faster information retrieval.

### 3.1.2 Insight from the interviews with Mykolas Romeris University librarians

In this subchapter we will make evaluation of student abilities to retrieve information from Internet and DB, training processes of information retrieval. To make evaluation we classified questions into three sections:

- Student problems and individual abilities of retrieving information form Internet and DB;
- Training processes and applicability of training into information retrieval from Internet and DB;
- Limitations in information retrieval from Internet and DB.

**Student problems of retrieving information form Internet and DB.** Evaluating the problems which students encounters with information retrieval from Internet and databases according

MRU librarians are knowledge about existence of curtain databases, poor knowledge on how to make proper keywords and queries as a consequence lost in abundance of results are inevitable. Also Librarian No. 1 notes that one of the problems are that students often try to use shortcuts to access information faster which leads to failure of finding information they needed. Librarian No. 2 says that if subject of interest is new and unexplored, that capabilities of finding information reduces, even though student knows how to deal with Internet and databases. Other factor according Librarian No. 2 is that poor English language skills can also reduce of retrieving reliable information.

Segments from interviews with librarians:

*"Main problem – fastest route <...> students try to retrieve information on Google Search <...> try to get into all systems by using Google Search".*

*"<...> input of the terms is too broad <...> It is quite complicated to select information needed in abundance <...>"* – Librarian No. 1.

*<...> they don't know where to find science literature". <...> it's quite hard to find science literature in Google Search, and even if they found it, sometimes it's not full text". "Only few of them knows about databases, and even if they do know, they usually get lost in abundance, they don't know how to select databases for they need, etc.". "<...> poor selection of keywords and translation of them into foreign language". "<...> there's always are exceptions when subject is new and unexplored, so even then if they select good keywords, there's no opportunity to find information enough". "<...> sometimes occur problems with students which English language is poor and most science literature is only in this language"* – Librarian No. 2.

*"Often students don't know about existence of certain information retrieval systems. They fail to evaluate reliability of search engine and eligibility of information they find"* – Librarian No. 3.

*"Sometimes it's hard for students to select most valuable results from large amounts of results given by databases. Also it's hard to make proper keywords"* – Librarian No. 4.

*"Insufficient knowledge about sources of scientific information, poor knowledge about making keywords and how to create queries"* – Librarian No. 5.

**Individual abilities of retrieving information from Internet and DB.** All librarians agree that often students are incapable on finding reliable information on their own, but after trainings capability on retrieving reliable information is sufficient.

Segments from interviews with librarians:

*"Results are not bad <...> students in groups are able to retrieve information fast <...> individually – everything happens, but mostly issue is that they select bad search engines"* – Librarian No. 1.

*"Mostly, without additional training or consultations, there were problems with data bases <...> they simply don't know where to find these databases". "Using Internet <...> not everybody able to know, what is science literature"* – Librarian No. 2.

*"It depends on information they try to find. More specific information retrieval needs more experience and knowledge. Also it depends on database or search engine, you need to know how to make a search in certain databases <...>"* – Librarian No. 3.

*"Students start to retrieve information independently by using Google Search or Google Scholar. After attending trainings in library on "General searching for the scientific information" or individual consultations with Subject Librarians students know how to use databases and know how to use keywords correctly"* – Librarian No. 4.

*"According to number of students attending to consultations and trainings we can assume that assistance is needed"* – Librarian No. 5.

**Training processes in information retrieval from Internet and DB.** Evaluating training processes in information retrieval from Internet and databases all librarians say that this processes includes from making group and individual trainings at the library on how to use library and external recourses to coming to lectures to make presentations on how to deal with information retrieval.

Segments from interviews with librarians:

*"There are opportunity to come to library trainings. We are trying to teach them on how to find books and other literature in library by them selves". "Library have seminars in which we are teaching them practical stuff". "<...> there are support system in which there are from 150 to 210 different trainings <...> trainings on "General searching for scientific information" <...>". "<...> they can order individual consultation <...> in library or on "Skype"". "<...> lecturers can invite Subject Librarian to the lectures to present information sources and how to do searching on it"* – Expert No. 1.

*"Theoretical lectures and practical exercises, also consultations on how to find information for specific topics"* – Expert No. 2.

*"It depends on the situation, but usually training begins from learning on how to use library catalog, data bases, <...>"* – Expert No. 3.

*"<...> presented with learning material on how to create queries, searching information sources, main principles of searching scientific information and systems. Also there are video prepared on "General searching for scientific information". Also students can come to individual consultations <...>"* – Expert No. 4.

*<...> students were presented with University library <...> and the searching of the published literature in it". "Trainings on how to make search to retrieve information <...>". "<...> librarian was invited to lecture to present information retrieval recourses <...>"* – Librarian No. 5.

**Applicability of trainings of information retrieval from Internet and DB.** About applicability of these trainings all librarians agree that this could be very helpful for all professions, starting teachers, scientists, lawyers to industry workers. Librarian No. 1 notes that teaching of information retrieval should begin in school, so that shows that knowing how to retrieve information is a huge advantage in people lives.

Segments from interviews with librarians:

*"<...> teaching should be done in schools already". "<...> for specific groups like, teachers, school librarians or even housewives <...>"* – Librarian No. 1.

*"<...> from law to exact sciences. Trainings would be useful in all fields; differences would be in search engine specifications <...>"* – Librarian No. 2.

*"In modern world, information literacy is needed in all fields"* – Librarian No. 3.

*"In all fields of science and industry"* – Librarian No. 4.

*"<...> in scientific work, in corporate management or completing daily tasks assigned by executives"* – Librarian No. 5.

**Limitations in information retrieval from Internet and DB.** Evaluating limitations in information retrieval from Internet and databases, librarians say that common problem of all databases have different specifications of using them to retrieve results and if people want to retrieve information needed, they need to spend time on learning how to use it. Librarian No. 5 notes that almost all scientific databases are commercial and access to information is limited.

Segments from interviews with librarians:

*"<...> there are huge amount of search engines and all of them has specifications of it's own <...> if you don't search for information every day <...> you can't remember all refinements <...>". "It's hard to select information you need from abundance of results". "<...> by default some search engines give basic search". <...> abundance of functions in search engines <...> if you need retrieve specific information you need to know where to push specific buttons"* – Librarian No. 1.

*"It's hard to evaluate reliability in Internet. Search is more complicated in databases and you need you make separate search in different databases"* – Librarian No. 2.

*"Sometimes information is separated into sections and sometimes when you try to find information by using advanced search, the search is made only in one section <...>"* – Librarian No. 3.

*"In some databases there's no basic search, so it can slow down information retrieval time, because you need to focus on aspects of making search"* – Librarian No. 4.

*"<...> there's no access free information <...> most of scientific information is only on commercial databases <...> so user can access only databases subscribed by library <...>"* – Librarian No. 5.

Overall, making conclusions from interviews with Mykolas Romeris University librarians, we can say that students often without proper training on how to use Internet and databases are somewhat incapable of finding reliable information and it sources for their studies due broad formulations of terms and queries and lack of knowledge about existence of these sources from Internet and databases.

Also, these trainings could be advantageous not even for University students, but also for any profession, which requires high demand of information literacy.

Finally, we can conclude, that information sources like internet or data bases have their own issues like they are very different, not unified and time consuming on learning how to use them properly to retrieve reliable information.

# CONCLUSIONS AND RECOMMENDATIONS

The purpose of this master thesis is to make theoretical and empirical research to investigate the possibilities to apply Semantic Web Technologies into internet and databases content search for students to retrieve information from internet and databases.

After theoretical analysis these final conclusions can be determined as:

1. Theoretical analysis of internet content search with Keyword-based search engines and Semantic Technology and Semantic Web Technologies based search engines shows that there are three levels of information processing and in order to get knowledge (the highest level of information processing) to perform some kind of skill or to make an intelligent decision, first of all people need to retrieve data (lowest level of information processing), then process it to make information (structured data for analysis, middle level of information) and process information to get knowledge. Data, information and knowledge interaction leads to specific skill performance or intelligent decision making.
2. Theoretical analysis shows that since the beginning of information retrieval by using computers and internet, all systems needed improvements to perform relevant information search and due to constantly growing numbers of web pages and documents on the internet many search engines tried to find the best solutions to their customers. Solution of Semantic Web Technologies was proposed to increase information retrieval relevancy by enabling machines to understand the meaning of all documents on the internet and databases and reduce limitations of keyword-based search engines like content quality assessment, lack of objectivity in factual questions, uncertainties about information existence on internet and poor internet page rankings.

Summarizing two parts of qualitative research final conclusions can be determined as:

3. After interviews with experts of Semantic Web Technologies we can determine that overall Semantic Web Technologies based search engines are more effective and has more functionality rather than Keyword-based search engines.
4. Business and government are generally interested in investing into development of Semantic Web Technology and applying them to their companies even though there are some limitations in high expectations to this technology to resolve all company problems by using them, but making benefits in production time, work effectiveness and knowledge production in a way that people can’t do, overcomes the limitations.
5. There is need of further development of Semantic Web Technologies to overcome a lack of infrastructure and tools, overcome challenges in teaching back-end users to work with these technologies, a need for more published data which would be linked one to another.

6. Semantic Web Technologies can help to fasten processes of learning and completing researches for science projects by enabling faster information retrieval and information adoption.

7. Semantic Web technologies according Rogers theory "Diffusion of Innovations" are in between Early Adaptors and Early Majority innovation adoption cycles and according Moore theory "Crossing The Chasm", "The Chasm" is still not crossed yet, but it should cross it then these technologies become mainstream.

8. Also according Shannon & Weaver "Mathematical Theory of Communication" theory "Noise Source" in Semantic Web Technologies could be determined as errors made by developers adopting technologies, also barriers of common language issues and information transmission errors adopting these technologies from old systems to new.

9. After interviews with librarians we can determine that students often are somewhat incapable of finding reliable information without proper training.

10. These trainings could be advantageous not even for University students, but also for any profession, which requires high demand of information literacy.

11. Information sources like internet or databases have issues with major usability differences which leads to time consumption on learning how to use them and lack of unified search engines to make search faster.

After making conclusions there are few **recommendations** for further investigations:

1. To make deeper investigation of Semantic Web Technologies in Rogers theory "Diffusion of Innovations" and Moore "Crossing the Chasm" to determine innovation adoption cycles and determine either these technologies going to cross "The Chasm" in certain areas (e.g. health care, e-government, industries, etc.).

2. To make comparative investigation on effectivity and limitations of information retrieval on booth Keyword-based search engines and Semantic Web Technologies based search engines.

# REFERENCES


1. Amerland, D. (2013). *Google Semantic Search: Search Engine Optimization (SEO) Techniques That Get Your Company More Traffic, Increase Brand Impact, and Amplify Your Online Presence*. Que Publishing.
2. Belkin, N. J. (1993). Interaction with texts: Information retrieval as information seeking behavior. *Information retrieval*, *93*, 55-66.
3. Berners-Lee, T., Hendler, J., Lassila, O., CHABOYER, B. C., NICOLAOU, K., BODDY, C. N., ... & GARB, H. N. (2001). 34 The Semantic Web. Retrieved from: http://oscar.ayoun.free.fr/www/ebook/Scientific%20American/2001/Scientific%20American%20-%202001%20-%2005%20-%20Mai.pdf
4. Blackburn, S. (2005). *The Oxford dictionary of philosophy*. OUP Oxford.
5. Bogner, A., Littig, B., & Menz, W. (eds.) (2009). Interviewing Experts.
6. Borek, A. (2014). *Total Information Risk Management: Maximizing the Value of Data and Information Assets*. Amsterdam: Morgan Kaufmann.
7. Brajer-Marczak, R. (2016). Elements of knowledge management in the improvement of business processes. *Management (1429-9321)*, *20*(2), 242-260. doi:10.1515/manment-2015-0063
8. Case, D. O. (2002). *Looking for information: A survey of research on information seeking, needs, and behavior*. Academic Press.
9. Cole, C. (2012). Information Need: A Theory Connecting Information Search to Knowledge Formation. Information Toady, Inc.
10. Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.
11. Deved, V. (2006). *Semantic web and education* (Vol. 12). Springer Science & Business Media.
12. *Dictionary of ICT.* (2004). Italy: Legoprint.
13. *Dictionary of Information & Library Management*. (2006). Italy: Legoprint.
14. Devedzic, V. 2006. Semantic Web and Education. USA: Springer Science-Business Media, LLC.
15. Floridi L. (2010). *Information - A Very Short Introduction*. Oxford University Press.
16. Hennink, M., Hutter, I., & Bailey, A. (2010). *Qualitative research methods*. London: Thousand Oaks (Calif.) : SAGE Publications.
17. Hepworth, M., & Walton, G. (2009). *Teaching information literacy for inquiry-based learning*. Elsevier.

18. Hitzler, P., Krotzsch, M., & Rudolph, S. (2009). *Foundations of semantic web technologies*. CRC Press.
19. Hock, R., & Price, G. (2004). The Extreme Searcher's Internet Handbook: A Guide for the Serious Searcher.
20. Joo, J. (2011). Adoption of Semantic Web from the perspective of technology innovation: A grounded theory approach. *International journal of human-computer studies*, *69*(3), 139-154. Retrieved from: https://pdfs.semanticscholar.org/216a/50b2d741f03cc63bc7475f9d54281f2801df.pdf
21. Kvale, S. (2007). *Doing interviews*. Los Angeles (Calif.) : Sage Publications
22. Ledford, J. L. (2009). *Search Engine Optimization Bible*. Indianapolis, Ind: Wiley.
23. Levene, M. (2010). *An introduction to search engines and web navigation*. John Wiley & Sons.
24. Lewandowski, D. & Höchstötter, N. (2007). Web Searching: A Quality Measurement Perspective. In Web Searching: Interdisciplinary Perspectives (in press). Dordrecht: Springer.
25. Lewandowski, D. (2015). Evaluating the retrieval effectiveness of Web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*, *66*(9), 1763-1775. EBSCO.
26. MacKay, D. M. (1955). Operational aspects of some fundamental concepts of human communication. *Synthese*, *9*(1), 182-198.
27. Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge, United Kingdom: Cambridge University Press.
28. Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge, United Kingdom: Cambridge University Press.
29. Meadow, C. T., Boyce, B. R., Kraft, D. H., & Barry, C. (2000). *Text Information Retrieval Systems*. Academic Press.
30. Methodology and Practice. Basingstoke England : Palgrave Macmillan. Retrieved from: www.researchgate.net%2Ffile.PostFileLoader.html%3Fid%3D569f637e614325aafc8b4574%26assetKey%3DAS%253A319919911047176%2540145328626265 7&usg=AFQjCNHjU6qLPSbrW0KdkvSP847y-bArQQ&sig2=OMlrTPnC-bHvctNTlU6RlQ
31. Mooers, C. (1951). Zatocoding applied to mechanical organization of knowledge. American Documentation, 2(1), 20–32.
32. Moore, G. A. (2007). *Crossing the chasm: marketing and selling disruptive products to mainstream customers.* New York: HarperBusiness.
33. Patel, K., & Parrish, J. (2016). The Impact of Knowledge Management is Critical in Aligning IT- Business for Effective Business Performance: From A Maturity

Perspective. *Proceedings for The Northeast Region Decision Sciences Institute (NEDSI)*, 1-10.

34. Passin, T. B. (2004). *Explorer's guide to the semantic web* (pp. 141-169). Greenwich: Manning.
35. Rogers, E. M. (2003). *Diffusion of Innovations*. Fifth Edition. New York (N.Y.): Free Press.
36. Saldana J. (2011). *Fundamentals of Qualitative Research.* New York : Oxford University Press, Inc.
37. Sanderson, M., & Croft, W. B. (2012). The history of information retrieval research. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1444-1451.
38. Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers college press. Retrieved from: https://www.google.lt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiX3_qFoprTAhXLjCwKHYUXAjMQFggYMAA&url=https%3A%2F%2Fwww.researchgate.net%2Ffile.PostFileLoader.html%3Fid%3D563ce2da6225ff3cae8b4590%26assetKey%3DAS%253A292843798188032%2540144630810198&usg=AFQjCNFmg1HP_ML6gcKo-KtgNBmiZqLvQQ&sig2=JK3g9A3uLVaQixfwHlfFJg
39. Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, *24*(4), 35-43.
40. *The Oxford dictionary of computing*. (2004). OUP Oxford.
41. Thierauf, R. J. (1999). *Knowledge Management Systems for Business*. Westport, Conn: Greenwood Publishing Group. EBSCO.
42. *Visuotinė Lietuvių enciklopedija*. (2014). Vilnius: UAB "BALTO print".
43. Shannon, C. W., Weaver, W. (1964). *The Mathematical Theory of Communication.* USA*:* Press UoI.
44. Woodside, A. G., & Quaddus, M. A. (2015). *Sustaining Competitive Advantage Via Business Intelligence, Knowledge Management, and System Dynamics*. Bingley, UK: Emerald Group Publishing Limited.
45. Yovits, M. C., Foulk, C. R., and Rose, L. L. (1981). Information flow and analysis: Theory, simulations and experiment. Part 1. Basic theoretical and conceptual development. Journal of the American Society for Information Science 32(3), 187-210. EBSCO.
46. Yow, V. R. (2005). *Recording oral history: A Guide for the Humanities and Social Sciences.* USA: AltaMira Press. Retrieved from: http://tlnistorie.granturi.ubbcluj.ro/pdf/Recording%20Oral%20History.pdf

47. Yu, L. (2007). *Introduction to the semantic web and semantic web services*. CRC Press.

## SUMMARY

The purpose of this master thesis is to investigate the possibilities to apply Semantic Web Technologies into internet and databases content search for studying. Following objectives were set 1) to present aspects of internet content search, 2) to define effectiveness and limitations of internet content search, 3) to make qualitative interviews with experts to get an insight about Semantic Web Technologies on internet content search, 4) to make brief analysis of Semantic Web Technologies through "Diffusion of Innovations", "Crossing the Chasm" and "Mathematical Theory of Communication" theories; 5) to make qualitative interviews with librarians and provide an insight on internet content search for studying.

First part consists of theoretical analysis on aspects about information retrieval with information retrieval systems and search engines. Second part is about evaluation of Semantic Web Technologies through "Diffusion of Innovations", "Crossing the Chasm" and "Mathematical Theory of Communication" theories. Third part is the empirical research which consists of two types of expert interviews. First part of interviews was with the experts of Semantic Web Technologies to determine the functionality and effectiveness, applicability to business, government, personal and professional lives and education. Ant second part of interviews was with Mykolas Romeris University librarians to determine students' capabilities of information retrieval, applicability of trainings to other businesses and limitations of current information retrieval methods from internet and databases. The study is finalized with conclusions of this master thesis and recommendations for further researches.



**Keywords:** Semantic Web Technologies, Keyword-based search engines, Information retrieval.

# SANTRAUKA

Magistro darbo tikslas yra ištirti Semantinių technologijų ir Semantinio tinklo technologijų pritaikymo galimybes internetinėms ir duomenų bazių turinio paieškoms. Suformuoti uždaviniai 1) pristatyti interneto turinio paieškų aspektus, 2) apibrėžti interneto paieškų efektyvumą ir trūkumus, 3) atlikti kokybinius interviu su ekspertais ir pateiktį įžvalgas apie paieškų turinį naudojant Semantines technologijas ir Semantines tinklo technologijas, 4) atlikti trumpą semantinio tinklo analizę pritaikant “Diffusion of Innovations” teoriją, “Crossing the Chasm” teoriją ir “Mathematical Theory of Communication” teoriją; 5) atlikti kokybinius interviu su bibliotekininkais ir pateikti jų įžvalgas apie interneto paieškas mokymosi tikslais.

Pirmoji darbo dalis yra internetinės paieškos internetinių paieškų sistemose ir paieškos varikliuose teorinė analinė. Antroji darbo dalis yra empirinis tyrimas, susidedantis iš dviejų dalių. Pirmoji interviu dalis yra su Semantinių technologijų ir Semantinių tinklo technologijų ekspertais apibrėžti šių technologijų funkcionalumą, efektyvumą ir pritaikomumą versle, valdžioje, asmeniniame ir profesiniame gyvenime, bei moksle. Antroji interviu dalis yra su Mykolo Romerio Universiteto bibliotekininkais apibrėžti studentų gebėjimus informacijos paieškoje, informacijos paieškos mokymų taikymą verslo srityse ir informacijos paieškos trūkumus internete ir duomenų bazėse. Darbas baigiamas išvadomis ir rekomendacijomis tolesniems tyrimams.

# APPENDICE NO. 1

## Questionnaire for expert interviews

1. What is the purpose of semantic web technologies?
2. What is the main differences between nowadays web search engines and semantic web search engines?
3. What are the main advantages and disadvantages of semantic technologies against other search engines?
4. What is the similarities and differences between semantic technologies and artificial intelligence?
5. How do you think what measures should be taken to reach “critical mass” necessary to make semantic search engines truly effective? What do you think how much time it’s going to take to reach this “critical mass”?
6. What are the biggest challenges you are facing working with semantic web technologies?
7. How semantic web technologies make content search faster?
8. Do results of the search with semantic web technologies are more reliable than results from keyword based search engines?
9. Do semantic technologies get interest from business, government?
10. What are benefits and downsides of semantic web technologies for people, business and government?
11. How business, government are involved in process of semantic web development?
12. What kind of businesses are interested in semantic web technologies? Are these companies investing in production of semantic technologies?
13. Are semantic technologies makes profit for companies providing it?
14. For what kind of professions semantic web technologies would be / is advantage?
15. What do you think, how semantic web technologies will change our professional and personal lives? Please provide what would be advantages and disadvantages of semantic web technologies participating in our daily life.
16. In your opinion how semantic web technologies can improve and impair communication between people?
17. How semantic web technologies can be used in education?
18. Does Semantic Technologies are “a mainstream” technology today?
19. According Everett Rogers Communication theory “Diffusion of Innovations” there are five Technology Adoption Cycles (Innovators, Early Adopters, Early Majority, Late Majority

and Laggards). What do you think at which Technology Adoption Cycle Semantic Technology users (businesses, government) are?

20. According same theory from last question in Technology Adoption Cycles there are area called “The Chasm”. What do you think does semantic technologies crossed “The Chasm”? What was the causes of crossing (not crossing) it?
21. Mathematical (Information) Theory of Communication by Shannon & Weaver provides Transmission Model of Communication, we can adjust this model to Semantic Web Technologies. What can be named as source of “Noise” in Semantic Web Technologies? How it can effect the message that user of Semantic Web Technologies gets?

## APPENDICE NO. 2

Questionnaire for librarians

1. What are the main problems for students with information retrieval from internet and data bases?
2. How enough students individually are capable of information retrieval from internet and databases?
3. What is the training process for students to learn information retrieval form internet and databases?
4. What are limitations of information retrieval in internet and databases?
5. What should be done to improve information retrieval from internet and data bases?
6. In your opinion where could be training of information retrieval could be applied?

# APPENDICE NO. 3

## Consent for Participation in Interview Research

1. My participation in this project is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time.
2. I have the right to decline to answer any question or to end the interview.
3. After this interview there will be transcript made of our conversation otherwise requested by the participant. There may be additional follow-up/clarification through email, unless otherwise requested by participant.
4. I understand that the researcher will not identify me by name in any reports using information obtained from this interview, and that my confidentiality as a participant in this study will remain secure. If participant wishes for the use of his/her full name in the study, this request will be adhered to as well.
5. I will be one of approximately 5-10 people being interviewed for this research.
6. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.
7. I have been given a copy of this consent form*.

   *Form is send over email to participant.

# APPENDICE NO. 4

**Email letter to with the invitation to interviews for experts of Semantic Web Technologies**

Hello,

My name is Renatas Liberis, I am master degree student from Mykolas Romeris University, Vilnius, Lithuania, and my field of study is Communication and Creative Technologies.

I want to invite you to participate in research about Semantic Web Technologies in Internet Content Search. Research type is Interview with experts of Semantic Web Technologies.

Interview is anonymous, so there will be no names of participants mentioned. Interview can be done in few different ways. It can be done in the most convenient way to you through Skype (video or audio) conversation, correspondence by emails or phone call. Duration of interview is up to 1 hour.

This research is a qualitative part of my final master degree thesis so all participants give a huge value for this research, so I would like to kindly ask you to spare some time to talk with me.

Kindly regards,

Renatas Liberis