Using Language Technology to Mediate Medical Information on Health Portals
User Studies and Experiments

Andrea Andrenucci

Abstract
The World Wide Web has revolutionized our lifestyle, our economies and services within health care. Health care services are no longer provided only at specialist centers and at scheduled hours, but also through online tools that give health care consumers access to medical information, health records, medical counselling and peer support. Such tools and applications are generally available on larger web sites or gateways called health portals. A large majority of online medical information consumers are laypeople (i.e. non experts) who appreciate the possibility to submit their information needs in their own native language. The information retrieval process where information requests from users and retrieved documents/answers are in different languages is called cross-language information retrieval (CLIR).

Mental health is one of the medical areas where some online applications have been successfully deployed in order to help people by providing in-depth medical information, counseling and advice. Despite the fact that online health portals are considered priority e-health tools for improving mental health, there are no formal knowledge instruments such as knowledge patterns that explicitly support the development of online health portals in the field of psychology/psychotherapy.

The goal of this research is to produce and evaluate a set of knowledge patterns, for the development and implementation of cross-lingual online health portals aimed at information seekers without medical expertise in the domain of psychology and psychotherapy. The knowledge patterns synthesize results of three research foundations: 1) User studies of portal interaction, based on interviews and observations about how users experience health information online and personalized search 2) Knowledge integration of existing language technology approaches, and 3) Experiments with language technology applications, in the field of cross-lingual information retrieval/question-answering. The target groups of this research are developers, researchers and health care providers, i.e. people who are responsible for mediating medical information on online health portals for users without medical expertise.

The chosen research framework is design science, i.e. the science that focuses on the study, development and evaluation of artefacts (objects that help people solve a practical problem). Typical examples of artefacts in IT are algorithms, software solutions and databases, but also objects such as processes or knowledge patterns. The developed and evaluated artefact in this research is a set of knowledge patterns for online health portal development.

The developed artefact contains fourteen knowledge patterns covering the three research foundations. Formative (structured workshops) and summative (online survey) evaluation of the artefact indicate that the knowledge patterns are useful, relevant and adoptable to a large extent, they also provide further directions for development of online mental health portals. Developing portals with multilingual support and tailored interfaces has the potential of helping larger groups of citizens to access relevant medical information.

Keywords: language technology, health portals, cross-language information retrieval, knowledge patterns.
USING LANGUAGE TECHNOLOGY TO MEDIATE MEDICAL INFORMATION ON HEALTH PORTALS

Andrea Andrenucci
Using Language Technology to Mediate Medical Information on Health Portals
User Studies and Experiments
Andrea Andrenucci
A mamma, papà e ai nonni
Abstract

The World Wide Web has revolutionized our lifestyle, our economies and services within health care. Health care services are no longer provided only at specialist centres and at scheduled hours, but also through online tools that give health care consumers access to medical information, health records, medical counselling and peer support. Such tools and applications are generally available on larger websites or gateways called health portals. A vast majority of online medical information consumers are laypeople (i.e. non-experts) who appreciate the possibility to submit their information needs in their own native language. The information retrieval process where information requests from users and retrieved documents/answers are in different languages is called cross-language information retrieval (CLIR).

Mental health is one of the medical areas where online applications have been successfully deployed in order to help people by providing in-depth medical information, counselling and advice. Despite the fact that online health portals are considered priority e-health tools for improving mental health, there are no formal knowledge instruments such as knowledge patterns that explicitly support the development of online health portals in the field of psychology/psychotherapy.

The goal of this research is to produce and evaluate a set of knowledge patterns, for the development and implementation of cross-lingual online health portals aimed at information seekers without medical expertise in the domain of psychology and psychotherapy. The knowledge patterns synthesize results of three research foundations: 1) User studies of portal interaction, based on interviews and observations about how users experience health information online and personalized search 2) Knowledge integration of existing language technology approaches, and 3) Experiments with language technology applications, in the field of cross-lingual information retrieval/question-answering. The target groups of this research are developers, researchers and health care providers, i.e. people who are responsible for mediating medical information on online health portals for users without medical expertise.

The chosen research framework is design science, i.e. the science that focuses on the study, development and evaluation of artefacts (objects that help people...
solve a practical problem). Typical examples of artefacts in IT are algorithms, software solutions and databases, but also objects such as processes or knowledge patterns. The developed and evaluated artefact in this research is a set of knowledge patterns for online health portal development.

The developed artefact contains fourteen knowledge patterns covering the three research foundations. Formative (structured workshops) and summative (online survey) evaluation of the artefact indicates that the knowledge patterns are useful, relevant and adoptable to a large extent, they also provide further directions for development of online mental health portals. Developing portals with multilingual support and tailored interfaces has the potential of helping larger groups of people to access relevant medical information.
Sammanfattning

Webben har revolutionerat vår livsstil, våra ekonomier och till och med tjänster inom vården. Hälsovården tillhandahåller inte längre bara tjänster på traditionella vårdinrättningar på schemalagda tider, men erbjuder även onlineverktyg för att ge hälsokonsumenter tillgång till medicinsk information, patientjournaler och medicinsk rådgivning. Dessa verktyg är generellt tillgängliga på större webbplatser som kallas hälsoportaler. En stor majoritet av konsumenterna av medicinsk information på nätet är icke-expertiser som uppskattar möjligheten att beskriver sina informationsbehov på sitt eget modersmål. Den informationshämtningsprocess där information som begärs och dokumen-ten/svaren som hämtas är på olika språk kallas tvärspråklig informationssökning ("cross-language information retrieval" på engelska).

Mental hälsa är ett av de medicinska områden där onlinetillämpningar har testats framgångsrikt för att hjälpa människor med djupgående medicinsk information, rådgivning och stöd. Trots att onlinehälsoportaler betraktas som prioriterade verktyg för att förbättra mental hälsa, finns det inga formella kunskapsinstrument som formellt stöder utvecklingen av hälsoportaler inom psykologi och psykoterapi.

Målet med denna forskning är att producera och utvärdera en uppsättning kunskapsmallar ("knowledge patterns" på engelska) för design och utveckling av onlinehälsoportaler riktade till informationsökande utan medicinsk expertis inom psykologi och psykoterapi. Kunskapsmallarna syntetiserar resultaten från tre forskningsgrunder: 1) Användarstudier av portalinteraktion baserade på intervjuer och observationer om hur användarna upplever hälsorelaterad information och personlig sökning, 2) Kunskapsintegration av befintliga metoder inom språktechnologin, och 3) Experiment med språktechnologiska tillämpningar inom områden för tvärspråklig informationssökning och fråge-bevarande. Målgrupperna för denna forskning är utvecklare, forskare och vårdgivare, dvs personer som är ansvariga för att förmedla medicinsk information från hälsoportaler till informationssökare utan medicinsk expertis.

Den valda forskningsramen är designvetenskap ("design science"), vetenskapen som fokuserar på studier, utveckling och utvärdering av artefakter. I

Den utvecklade artefakten innehåller fjorton kunskapsmallar som syntetiserar de tre forskningsgrunderna. Utvärderingen av artefakten visar att kunskapsmallarna är användbara och relevanta i stor utsträckning, samt att de bidrar till utveckling av detta område. Att utveckla portaler med flerspråkigt stöd och skräddarsydda gränssnitt har potentialen att hjälpa en större grupp medborgare utan att tvinga dem att besöka traditionella vårdinrättningar.
Acknowledgements

This thesis has been a long journey and there are many people who have played a major role in my work in all those years.

First of all, I would like to express my deepest gratitude to my supervisor Prof. Hercules Dalianis and my co-supervisor Dr. Sumithra Velupillai. You have lit my research sparkle once again and you have taught me so much during these years together. Thank you for always supporting me, for always being positive, encouraging and answering to all my questions. This thesis would have not been possible without you! I will never forget you.

Second, I would like to thank my former supervisor Prof. Jacob Palme for all the support during my first years at DSV. It was great working with you in the EU projects and discovering cross-language information retrieval. I would also like to thank Gunborg Palme for all the help with the studies and the experiments. Thanks also to Lars Enderin and Torgny Tholerus for all the help with the programming and for setting up the software for the experiments. I would also like to thank the Director of studies at DSV, Stefan Möller, and the Head of the department Prof. Uno Fors, for granting me time off to complete my research. A big thanks also to the Director of studies at the graduate level, Assoc. Prof. Sirkku Männikkö Barbutiu, for all the help through these years.

I would also like to express my gratitude to the following persons for giving me invaluable feedback and invaluable comments in the creation process of the knowledge patterns: Assoc. Prof. Martin Henkel, Dr. Eriks Sneiders, Prof. Paul Johanesson, Assoc. Prof. Erik Perjons, PhD student Rebecka Weegar and Assoc. Prof. Åsa Smedberg.

Prof. Janis Stirna and Dr. Aron Henriksson, thank you for all your invaluable comments during the first public draft of my thesis, and for the pleasant and informative discussion during the pre-doc seminar. Thank you Prof. Stirna for all your guidance in the world of knowledge patterns!

My thanks go also to Minna Forsell and Markus Näslund for the interesting discussions about cognitive behavioural therapy and dynamic psychology. My
sincere thanks go also to PhD student Andreas Paulsson for all the lively discussions concerning research methodologies and formal knowledge.

I also would like to thank all my colleagues of the Daisy team and the study administration of DSV. Thank you for the support and the nice coffee-breaks together.

A special “thanks” goes to my fellow windsurfing mate Daniel Mauritzell for all our surf trips to the windy island of Öland, and all the magic moments spent on water, windsurfing or wakeboarding. These moments have filled me with joy and positive energy.

Finally, I would like to thank my beloved wife Anna Andrenucci, my beloved daughters Elisa, Alessia and Emilia for all their love and for always being there for me. You are the light of my life.
List of publications

This thesis is a compilation of seven separate articles presented and accepted for publication between 2004 and 2018. The author of this thesis is the sole author in five articles and first author in two of them. Article I, II and III were included as parts of the licentiate thesis of the author of this thesis: Andrenucci A., (2005), Using Web Portals for Medical Information Mediation. Licentiate Thesis. Stockholm University. Below in the list of the articles, it is also included the contribution of the author of this thesis.


Article VII Andrenucci, A., Dalianis, H. and Velupillai, S. (2018) Knowledge patterns for online health portal development. Accepted for publication on Health Informatics Journal, Sage publishing. Published ahead of print. Andrea Andrenucci was responsible for 80% of the article.
Contents

Chapter 1 Introduction and overview .......................................................... 1
  1.1 The knowledge domain – health portals and mental health ......................... 2
  1.2 Problem definition, research goal and research questions ........................... 3
  1.3 Theoretical background ........................................................................... 8
  1.4 Main contributions of this thesis .............................................................. 8
  1.5 Structure of the thesis and publications .................................................... 10

Chapter 2 Research background and related research .............................. 12
  2.1 User modelling in health applications – e-health .................................. 12
  2.2 Cross-language information retrieval .................................................. 14
      2.2.1 Machine translation ......................................................................... 14
      2.2.2 Dictionary-based methods ................................................................. 16
      2.2.3 Parallel or comparable corpora methods ........................................... 18
      2.2.4 Common challenges ........................................................................ 19
  2.3 Knowledge patterns .................................................................................. 20

Chapter 3 Research framework, materials and methods ............................ 23
  3.1 Research framework ................................................................................ 24
  3.2 Design science research framework: design and develop artefact - materials and
      research methodology for foundational aspects .......................................... 30
      3.2.1 Materials .......................................................................................... 30
      3.2.2 Research methodology for the foundational aspects ............................. 38
  3.3 Design science research framework: evaluate artefact - evaluation of the knowledge
      patterns ....................................................................................................... 48
  3.4 Limitations ................................................................................................. 52
  3.5 Ethical issues .............................................................................................. 54

Chapter 4 Results and discussion .............................................................. 56
  4.1 Research questions 1 and 2 (Article I, Article III) ...................................... 56
      4.1.1 Research question 1: how do information seekers experience interacting with
      a mental health portal? ............................................................................. 56
      4.1.2 Research question 2: which user profiling methods do psychological
      information seekers prefer? ........................................................................ 59
  4.2 Research question 3: which are the main approaches in QA and CLIR and which
      approaches fit online health portals (Article II and section 2.2) .................... 61
4.3 Research question 4: which query translation methods work better for health portals and which are the best practice methods for extraction of bilingual lexicons? (Article IV, Article V, Article VI) ........................................................................................................... 66

4.3.1 Query translation methods and health portals (Article VI) ...................................... 66
4.3.2 Best practice methods for extraction of bilingual lexicons: quality of word relations extracted from different versions of the corpora (Article IV, Article V) ....... 70

4.4 The artefact: knowledge patterns for online health portals development (Article VII) ........................................................................................................................................ 72

Chapter 5 Evaluation of the knowledge patterns ................................................................. 74
5.1 Evaluation results ........................................................................................................... 74
5.2 Comparison with related research .................................................................................. 82
5.3 Lessons learned ............................................................................................................ 83
5.4 Research quality .......................................................................................................... 84

Chapter 6 Conclusions ......................................................................................................... 86
6.1 Summary and contributions .......................................................................................... 86
6.1.1 Theoretical contributions ....................................................................................... 87
6.1.2 Practical contributions and societal consequences ................................................. 89
6.2 Future work .................................................................................................................. 89

References ......................................................................................................................... 91

Appendices ....................................................................................................................... 104
Abbreviations

API Application Programmer Interface
CLIR Cross-language Information Retrieval
DSRF Design Science Research Framework
FAQs Frequently Asked Questions
GUI Graphical User Interface
HCI Human-Computer Interaction
IR Information Retrieval
MT Machine Translation
MWUs Multi-Word Units
NL Natural Language
NLP Natural Language Processing
POS Part of Speech
QA Question-Answering
SMT Statistical Machine Translation
SWUs Single-Word Units
UM User Model
Chapter 1 Introduction and overview

Medical content is one of the most retrieved types of information on the World Wide Web (WWW), both in the United States and in Europe. Previous research has shown that 75% of all American adults online use the Internet to look for health care information (Taylor & Leitman 2001, Fox & Duggan 2013). A study of five European countries (Norway, Denmark, Germany, Greece and Portugal) by Kummervold & Wynn (2012) showed similar results for European information seekers. In Sweden, a country where a vast majority of the population have Internet access, 69% of Internet users search for health information online (Findahl 2013).

Recent research in Sweden (Eklund 2014) has shown that search for health information is following a new trend. From a more generic search utilizing general search engines and the whole web as an information resource, users now tend to utilize specialized websites/portals that they trust or that have been suggested by friends or professionals. Health portals are thus playing a bigger role when it comes to providing medical information to Internet users. The importance of health portals is thus expected to grow as the availability of Internet widens worldwide (Chapman et al. 2010).

A health portal is, according to Bamidis et al. (2005):

“… an interactive service or entry point site to the Web, offering information resources related to health subjects like hospital and doctor information, nutrition, health guide, daily care, health tests, latest published research work, health articles on nearly every subject, health electronic libraries and athletics. Services offered include search engines, links to health portals around the world, e-mail, chatting, news about the pharmaceutical industry and a part with medical information for the people that practice medicine....”

This exhaustive definition captures the essence of what health portals cover and provides a basis for the work presented in this thesis.

Previous research studies have shown that health portal users appreciate the possibility to express their information needs in their native language (Eysenbach & Köhler 2002, Koivunen et al. 2007). The information retrieval process where the user requests information and the documents/answers retrieved are
in different languages is called cross-language information retrieval (CLIR). Since a large majority of consumers of online health information are laypeople (i.e. non-experts) with different profiles and different information needs (Névéol et al. 2006, Wang et al. 2012), the mediated information should be tailored to the user needs and profiles, as well as allow users to express their information needs in their own words and in their native language. This implies that the development of online health portals involves having to consider foundational aspects that address technical solutions such as language technologies and solutions for appropriate graphical user interfaces.

The purpose of this thesis is to:

- Provide indications about how users search for health-related information on health portals and how they experience different graphical user interfaces (GUls) and user-tailored search services.
- Describe and analyse available cross-language information retrieval (CLIR) methods - and question-answering approaches that better fit health portals.
- Develop and evaluate a set of knowledge patterns with design guidelines and implications for future developers of online health portals.

The specific domain of the health portals is in the field of psychology and psychotherapy. This research is limited to Swedish as source language and English as target language. To the best of this author’s knowledge, this is the first comprehensive research that addresses solutions for language technologies as well as user studies with graphical user interfaces, and provides evaluated knowledge patterns for health portals in the field of psychology and psychotherapy.

1.1 The knowledge domain – health portals and mental health

Medical content is among the most retrieved information on the World Wide Web (Hung et al. 2013) and mental health is also one of the areas where health portals are being successfully utilized to help users with in-depth medical information (Webb et al. 2008), psychological counselling (Christensen & Hickie 2010b) and monitoring of health conditions (van der Krieke et al. 2014). Several research projects have shown positive results when it comes to online treatment of panic and anxiety (Richards et al. 2003), post-traumatic stress (Lange et al. 2003), prevention of eating disorders (Winzelberg 1997) and support of citizens with severe mental illnesses (Farrell et al. 2004).
In comparison with treatments by medical professionals, psychological counselling online is more flexible and inexpensive since it is not bound by logistical constraints. Online portals have proved to facilitate the life for citizens that cannot visit mental health experts at scheduled hours, e.g. users living in rural areas (Christensen & Hickie 2010b), or suffering from physical handicaps (Zeng & Parmanto 2004) or mental impairments (Farrell et al. 2004). Some countries with large rural areas (e.g. Australia) consider online portals a priority for improving mental health on a national level (Christensen & Hickie 2010a).

Research studies (e.g. Weisband & Kielser 1996) have also shown that patients tend to suffer less from social anxiety while interacting with computers, thus allowing them to reveal more personal information about themselves. Hence, computer-based interaction online might allow users to reveal more details than a face-to-face session with a medical professional, and therefore provide a complete picture of the personality of the patient (Joinson et al. 2008). Some evaluations of computer use by psychiatric patients have also shown that even users with severe mental impairments interact very successfully with computers, including patients who are unable to communicate with mental health personnel (Farrell et al. 2004).

1.2 Problem definition, research goal and research questions

This thesis focuses on the following practical problem and knowledge gap in current research.

**Practical Problem:**

Health care is moving from a traditional doctor-patient interaction at medical arrangements towards e-health services, in order to stimulate patient empowerment and engagement (Eklund 2014, Ricciardi et al. 2013) or to help people living in rural areas (Christensen & Hickie 2010a). This is particularly evident in mental health, where people suffering from health problems tend more and more to seek help online rather than through traditional arrangements (Webb et al. 2008). Although online health portals are considered priority e-health tools for improving mental health (Christensen & Hickie 2010a), there are no formal knowledge instruments such as guidelines/knowledge patterns that explicitly support the development of health portals in the field of psychology/psychotherapy and consider the implementation of search services in several languages.
A knowledge pattern is a way of formalizing knowledge that describes solutions to design problems (Alexander et al. 1977). Knowledge patterns have been utilized previously in e-health to introduce business models to guide e-health providers and marketers for commercializing e-health services (Mettler & Eurich 2012, Osterwalder & Pigneur 2010) or analysing security requirements for e-health applications (Dritsas et al. 2006), but not for the development of online health portals for mental health.

**Knowledge gap:**
This research is based on *three foundational aspects*: 1) *User studies of health portal interaction*, i.e. studies of how information seekers utilize and experience mental health portals. 2) *Knowledge integration* of language technology approaches within cross-language information retrieval (CLIR), question-answering (QA) systems and their distinctive features; the review resulted in defining three major approaches and their distinctive features. 3) *Experiments* with language technology applications within CLIR on a mental health portal.

This research looks at all these aspects combined, while previous research on health portals have focused on each foundation separately: integration and review of existing knowledge (Luo & Najdawi 2004), studies of portal interaction and user search behaviour (Gurel et al. 2012, Bamidis et al. 2005), experiments and evaluation of techniques and applications available on the portals (Glenton et al. 2005, Moon & Burnstein 2005).

The **goal** of this research is to produce an artefact (knowledge patterns) that synthesize all these three foundations, for the development and implementation of appropriate online health portals aimed at information seekers without medical expertise in the domain of psychology and psychotherapy. To the best of this author’s knowledge, these are the first knowledge patterns that explicitly cover health portals in this field.

The target groups of the results of this research are developers and people who are responsible for mediating health information on health portals for users without medical expertise.

Based on the research goal, we address the following research questions (the relation between research goal, research foundations and research questions is shown in figure 1):

*How should online health portals be designed in order to enable tailored user interaction and cross-language information access?*

In order to answer the main question, the following research questions were individuated:
1) *How do information seekers experience interacting with a mental health portal? (Article I, Article III)*

Mental health is one of the medical areas where online applications have been successfully tested to help people with in-depth medical information (Webb et al. 2008), counselling (Christensen & Hickie 2010b) and advice (van der Krieke et al. 2014). Earlier research in e-health for mental health (e.g. Koivunen et al. 2007) has mainly focused on evaluating portals from the practitioners’ point of view. This research aims to evaluate a mental health portal from the lay users’ point of view, collecting information about how they interact with the portal, and evaluating the acceptance of information mediated by the portal, as well as finding the factors that, according to information seekers, characterize a good and informative health portal.

2) *Which user profiling methods do psychological information seekers prefer? (Article III)*

Mental health information seekers have different interests, knowledge skills and search goals (Borzekowski et al. 2009). Sainfort et al. (2009) pinpoint several specific qualities that GUIs must embrace in order to be optimal for the medical personnel and the patient: they have to be multimodal, i.e. with multiple interaction modalities, personalized, and adaptive, i.e. able to adjust to the knowledge, background and skills of the users. However, health portal technologies are rather limited since they do not allow personalized search facilities (Moon & Burstein, 2005) and as the amount of available health information grows, it is clear that there is a need to tailor the information to the individual needs of health care information seekers (Hawkins et al. 2008). The ranking algorithms of search engines and QA-systems on health portals generally do not consider users’ different backgrounds and profiles, and implement a “one-size fits all” information delivery approach. This thesis tries to provide a multimodal, personalized and adaptive approach that tailors the retrieval of information on a health portal to users’ characteristics and information needs. The approach is used as a tool to investigate which contexts better fit personalization of information retrieval which contexts do not. The aim of the research is not to produce a new adaptation mechanism, but rather to build evidence for the efficacy of adapting the information delivery in the field of mental health.
3) Which are the main approaches in question-answering and cross-language information retrieval and which approaches fit online health portals? (Article II, Section 2.2)

Since natural language interfaces are utilized when searching for information on health portals, a review of the main research approaches within QA and CLIR was performed, outlining the advantages and disadvantages of each method, the context and the applications that best fit each technique in the health domain.

IR systems are traditionally seen as document retrieval systems, i.e. systems that return documents that are relevant to the user’s information need, rather than exact answers. QA systems return specific information passages or text chunks rather than entire text documents.
4) Which query translation methods work better for health portals and which are the best practice methods for the extraction of bilingual lexicons? (Article IV, Article V, Article VI)

Earlier research has demonstrated that users that search for health information online prefer to submit information requests in their native language (Eysenbach & Köhler 2002, Lopes & Ribeiro 2013, Pecina et al. 2014, Cline & Haynes 2001, Henriksson et al. 2014).

Several available CLIR techniques can be implemented to support input from users who have different native languages. These techniques support translation of information requests or documents from source to target language, or translating both into an intermediate language (pivot language) or a semantic representation (interlingua translation).

The information request translation is the most common translation approach since it is less computationally costly and easier to maintain (Kishida 2005, Rosemblat et al. 2003). There are three main types of methods for translating information requests or user queries: approaches based on a) bilingual dictionary search, b) machine translation and c) parallel corpora (Gey et al. 2005, Kishida 2005). Earlier research in this field (e.g. Pecina et al. 2014, Daumke et al. 2007) has not focused specifically on the domain of psychology and psychotherapy, nor considered both Single-Word Units (SWUs) and Multi-Word Units (MWUs). We try to fill this gap with the help of applications and evaluations of available query translation techniques.

Since the health domain is very specialized, health information seekers’ vocabulary might not be sufficiently advanced to match medical terminologies of health portals and specialized websites (Zhang 2010, Néveol et al. 2006), so bilingual lexical resources for lay people might be needed in the field of CLIR.

With the increase of online health information and health gateways, the availability of parallel and comparable medical corpora is growing on the web (e.g. the Kreshmoi project1, webhealth.info). The extraction of bilingual lexicons from parallel corpora is a way to produce bilingual lexical resources (Kishida 2005) for query translation in CLIR. Parallel corpora can be pre-processed and annotated in many different ways: e.g. they can be lemmatized2, part-of-

---

2 Finding the base form of a word
speech tagged\textsuperscript{3}, syntactically parsed\textsuperscript{4} and chunked. If health portals and medical resources online have access to parallel or comparable corpora, how much pre-processing of the texts is necessary if we need to extract a domain-specific bilingual lexicon from the corpora? We have studied this in two articles, focusing on the quality of word relations extracted from differently pre-processed versions of a corpus.

1.3 Theoretical background

The theory behind our research was provided by scientific articles and books about the following subject areas: information retrieval (IR), cross-language information retrieval (CLIR), natural language processing (NLP), e-health, knowledge management (KM) and user modelling (UM). This research is based on the design science framework (Gregor & Hevner 2014, see chapter 3), i.e. the science that focuses on study, development and evaluation of artefacts. Basically, artefacts are objects that help people to solve a practical problem. They could take the form of physical objects such as hammers, computers or chairs, or non-physical objects such as regulations and guidelines. Typical examples of artefacts in IT are algorithms, software solutions and databases. The artefact produced and evaluated in this research is the set of knowledge patterns for health portals development.

1.4 Main contributions of this thesis

This thesis investigated the mediation of health information on a portal for psychological counselling and psychotherapy.

Earlier research in the field of health portals have focused on usability studies when information is searched (Gurel & Cagiltay 2012, Bamidis et al. 2005), analysis of information seekers’ search behaviour while interacting with health portals (Eklund, 2014, Andreassen et al. 2007) or on search keyword effectivity for lay users and medical practitioners (Friberg Heppin 2010). Other studies have evaluated the information or the functionalities available on portals (Glenton et al. 2005, Moon & Burstein 2005,) and the portals’ trustworthiness according to existing literature (Luo & Najdawi 2004) or according to users’ opinions (Eysenbach & Köhler 2002).
To the best of this author’s knowledge, this is the first thesis that covers health portals and knowledge patterns based on three foundations combined: 1) user studies based on interviews and observations about how users experience health information online and personalized search; 2) literary surveys and knowledge integration; 3) experiments with applications in the field of cross-language information retrieval/question-answering.

Based on these foundations, the major contributions of this thesis are:

User Studies
- A study of how psychological information seekers experience and utilize an online portal for mental health care.
- A comparative study that elicits the differences between retrieval results with different personalized search services and with or without profiling of the user.
- Indications about how users search for health-related information on health portals and how they use the different interfaces that support request formulation.

Knowledge integration
- Literary reviews and crystallization of the main approaches within CLIR and QA, and their distinctive features.

Experiments
- Two versions of a Swedish-English parallel corpus from the web4health portal: a version with inflected forms and a version with word lemmas. For each version, the texts were annotated with POS tagging, parsed syntactically, and aligned at the sentence and word level.
- A pilot study and a follow-up study for best practice findings of methods for extraction of bilingual Swedish-English lexicons for laypeople in the domain of psychology and psychotherapy.
- A bilingual Swedish-English domain-specific lexicon, consisting of approximately 14,000 entries.
- A study that assesses the quality of available CLIR query translation methods (bilingual dictionary search and MT), with Swedish as a source language and English as target language.
- Indications about the CLIR methods that work best for health portals.

A joint contribution for all three parts is a set of knowledge patterns for developers of health information portals. The patterns cover GUI, QA, CLIR, bilingual corpora, search patterns of health information seekers and how users experience psychological information online.
All articles included in this thesis have been presented at conferences or accepted for publication in journals for e-health, information systems and health informatics.

1.5 Structure of the thesis and publications

The first chapter is an introduction that presents the research problem and the research questions. The second chapter reviews related research and the research settings in which this research is positioned. The third chapter defines the research framework, materials, methodology, limitations and ethics. The fourth chapter discusses the research foundations results, and answers the research questions. The fifth chapter contains the results of the evaluation of the knowledge patterns, a comparison with related research, a discussion of the lessons learned and a discussion of the validity of the research; the summary, conclusions and future work are found in chapter six.

**Article I** Andrenucci, A. and Forsell M. (2004). Computer-based Psychological Counseling on the Web: an Empirical Study of Web4health. Andrea Andrenucci’s contribution to this paper is approximately 50%. The contribution corresponds to writing the article based on the qualitative and quantitative results of the study conducted by the co-author.

**Article II** Andrenucci, A and Sneiders, E. (2005). Automated Question Answering: Review of the Main Approaches. Andrea Andrenucci was responsible for 80% of the article. He is the first author of the article and contributed to all parts of it, and was responsible for collecting and reviewing the research material. Eriks Sneiders contributed with valuable domain feedback and feedback about the paper structure which improved the paper overall.

**Article III** Andrenucci, A. (2006). Medical Information Portals: An Empirical Study of Personalized Search Mechanisms and Search Interfaces. Andrea Andrenucci is the sole author of this article. He developed the UM-software, performed the study, collected the results and wrote the article.

**Article IV** Andrenucci, A. (2007). Creating a Bilingual Psychology Lexicon for Cross Lingual Question Answering, A Pilot Study. Andrea Andrenucci is the sole author of this article. He created the bilingual lexicon, compiled the reference data, performed the pilot study, analysed the data and wrote the article.
**Article V** Andrenucci, A. (2010). Creating a Bilingual Psychology Lexicon for Cross Lingual Question Answering, A Follow-up Study. Andrea Andrenucci is the sole author of this article. He created the bilingual lexicon, compiled the reference data, performed the study, analysed the data and wrote the article.

**Article VI** Andrenucci, A. (2016) Experiments with Cross-language Information Retrieval on a health portal for psychology and psychotherapy. Andrea Andrenucci is the sole author of this article. He created one of the bilingual lexicons utilized in the experiments, designed and planned the evaluation study, wrote the requirements specification for the evaluation software, analysed the data and wrote the article.

**Article VII** Andrenucci, A., Dalianis H. and Velupillai S. (2018) Knowledge patterns for online health portal development. Andrea Andrenucci is the first author and was responsible for 80% of the article. He developed and evaluated the knowledge patterns and contributed to all parts of the article. Hercules Dalianis and Sumithra Velupillai contributed with continuous feedback on the studies and the evaluation, improving the writing and the quality of the article overall.
Chapter 2 Research background and related research

This chapter describes the larger research settings and the related research in which this research is positioned.

2.1 User modelling in health applications – e-health

According to Australia’s national e-health strategy summary,6 e-health is “the means of ensuring that the right health information is provided to the right person at the right place and time in a secure, electronic form for the purpose of optimizing the quality and efficiency of health care delivery”. Personalization of the retrieved information plays an important role when it comes to securing that the right health information is provided to the right person (Grasso & Paris 2011). One way to perform personalization is to utilize a user model (UM), i.e. “the knowledge about the user, either explicitly or implicitly encoded, that is used by a system to improve the interaction between the user and the system” (Kass & Finin 1998). Explicit knowledge about the user is gathered by letting the user directly state her preferences or goals for example with a form. Implicit knowledge is acquired indirectly by monitoring the interaction between the user and the system. Both approaches have advantages and disadvantages. In the explicit approach, the user is in control of the information contained in the user model and can easily update it. However, some users might consider this a tedious and time-consuming step, which might prevent them from submitting important information (Waern 2004). Implicit knowledge overcomes this problem since the user model is acquired without bothering the user; however, misunderstandings might arise or wrong information might be presented if the model is wrong or incomplete.

Two successful examples of explicit UM are described in Colineau and Paris (2011) and Camerini et al. (2011). Colineau and Paris (2011) utilized UM for

---

text generation on a family-centred health portal in the field of nutrition and healthy lifestyle. The UM encapsulates knowledge both on an individual level (i.e. concerning every single user) and a group level (i.e. concerning the whole family as a group). This information is provided explicitly by the user and covers health improvement goals, health condition including smoking/drinking habits, diet and physical activities. The UM also encapsulates group information about how the families rated their level of confidence for achieving the health improvement goals. The portal contains the following tools: 1) a lifestyle idea tool where families can monitor their current lifestyle and submit ideas about how to improve it; 2) a diary tool where families can keep track of their progress towards a healthier lifestyle and emotional wellbeing; 3) a message board tool where families can send and receive messages to help and encourage each other, and where the portal sends recommendations 4) a feedback tool with summaries of the families’ activities and encouragement messages. Camerini et al. (2011) utilizes UM to present tailored exercise videos in order to cure patients afflicted by fibromyalgia, i.e. a chronic pain condition that affects muscles, ligaments and tendons. The UM encapsulates information about the user age, gender, training experience, pain level, pain localization, available training tools and available time. The system then presents the exercises that best fit the UM, and users have the opportunity to rate them and select their favourite exercises.

One example of implicit UM is the one implemented by De Rosis et al. (2006), also in the field of dietary counselling, where a health promotion system tries to improve the eating habits of its users. The system monitors the dialogue between the user and the system, in order to detect the user’s mental state and emotions, adapting the conversation according to these parameters. The dialogue utilizes different persuasion techniques in order to convince the user to improve his/her eating habits. The “medical knowledge” is partly based on a corpus built from published transcripts of dialogues between human dieticians and patients.

The user model utilized in this research (see section 3.2.1, Article III) is both explicit and implicit. Previous research studies (e.g. Kobsa et al. 2001) have shown that it is preferable to utilize the explicit approach, however, as mentioned above, users may be reluctant to reveal personal information due to lack of trust or lack of time (Schwab et al. 2000), so it was decided to implement both approaches. Similarly to other health information retrieval systems (e.g. Santos et al. 2003), our UM also keeps track of the user objectives with the search.
2.2 Cross-language information retrieval

The information retrieval process where the query submitted by users and the documents/answers retrieved by the search engine are in different languages is called cross-language information retrieval (CLIR). According to research literature (Kishida 2005) translation methodologies within CLIR can be classified in three main approaches: 1) query translation, i.e. user queries in a source language are translated into a target language in order to match the document collection; 2) document translation, i.e. the documents are translated into the user query’s language; 3) interlingua translation, i.e. both user queries and the documents are converted into an intermediate language or semantic representation for matching purposes. The query translation method is the most common translation approach since it is less computationally costly and easier to maintain (Kishida 2005, Rosemblat et al. 2003). There are three main types of query translation methods: approaches based on a) bilingual dictionary search, b) machine translation and c) parallel corpora (Gey et al. 2005, Kishida 2005).

2.2.1 Machine translation

CLIR based on machine translation (MT) translates the source language query into a target language query with the help of software for machine translation (Zhu & Wang 2006). Linguistic and semantic analyses are applied on user queries in order to improve translation quality. In some cases, MT techniques have performed worse compared to simpler dictionary-based translations (Ballesteros & Croft 1998). Short queries based on words can have different meanings in different contexts, so ambiguity is a particular problem. The information contained in one-word queries can be too limited to find out the “right” context of the word. MT Systems also tend to provide only one translation and thus do not expand the queries with synonyms or related words (Nie et al. 1999). Quality MT systems are also difficult and expensive to build and are thus limited to specific language pairs (Zhu & Wang 2006). Other drawbacks of MT systems are that they tend to lack adaptation to specific domains (like the medical domain) and have economic or usage constraints (Pecina et al. 2014).

In recent years machine translation approaches based on statistics (SMT) and training data have become more and more popular. The strength of this approach is that it utilizes advanced machine learning techniques that train the translation models on huge monolingual and bilingual resources in specific domains. They can also utilize external dictionaries and glossaries to improve the quality of the translations. Moses (Pecina et al. 2014) is a system of this kind: it provides multilingual information retrieval in the medical domain with
good translation and retrieval results (Precision<sup>6</sup> at a cut-off of 10 documents - P@10 - slightly around 38% for French-English translations), but it did not manage to outperform Google Translate in query translations (with P@10 over 40%). It utilizes a phrase translation model that statistically predict relations between source phrases and target phrases. The model is trained on parallel sentences from bilingual resources with the help of probabilistic word alignment (Och & Nej 2003).

Wu et al. (2011) proposed a SMT system that translates (for patients) PubMed titles<sup>7</sup> in six different languages (Hungarian, French, Turkish, Polish, Spanish and German) and compared the results with Google Translate’s output translations, providing higher quality translations in this domain. As training data, it utilizes the English titles in PubMed articles and their translation in the other languages. Statistical models such as the IBM Models 1-5 (Brown et al. 1993) and N-grams (i.e. queries and documents are analysed as a sequence of n-gram units)<sup>9</sup> are applied for word alignment between sentence pairs.

Machine translation techniques based on neural networks (neural machine translation NMT, Bahdanau et al. 2015, Sutskever et al. 2014) have recently become the more popular paradigm in MT (Etchegoyhen et al. 2018). This is the MT technique that is also utilized in Google Translate (Wu et al. 2016). Unlike SMT it does not rely on phrase translation models but translates each word separately with the help of neural networks (Sutskever et al., 2014). The source language words are stored into a fixed-length vector (word embedding) with the help of a neural network (encoder), where the states of layers in the neural network are utilized to enrich the content of the vectors with context information. The vectors are then decoded with the help of another neural network (decoder), which utilizes the context information to predict the word translations. This approach has several advantages compared to SMT (Jean et al. 2014): 1) it reads the input sentences as bag-of-words/sequences of words and does not need linguistic properties or language models to perform the translations; 2) NMT systems are trained/tuned as a whole unit and do not need separate tuning for each individual part, unlike SMT systems that require tuning of the language models, the tables with the sentence pairs etc.; 3) they require more processor power but less memory capacity than SMT systems, which usually process large tables of source-target sentence pairs. However, NMT has shown some difficulties in coping with the translation of longer sentences (Toral & Sanchez-Cartagena, 2017), especially if they tend to be longer.

---

<sup>6</sup> Precision measures the ability of IR systems to recover relevant documents, see section 3.2.2.4


<sup>8</sup> https://translate.google.se/, Accessed September 10, 2017

<sup>9</sup> N-gram is a contiguous sequence of “n” number of items (e.g. words or letters) from a text or speech
than the sentences utilized as training data (Cho et al. 2014). They also have difficulties in handling rare words (Speerstra 2018) and larger vocabularies (Jean et al. 2014), since the complexity of training NMT systems increases with larger vocabularies. They also need more training time than SMT models (Speerstra 2018). Another drawback is that NMT has shown dropping performance when applied to multi-domains (Farajian et al. 2017) or when moved to domains that differed from the areas of the training data (Koehn & Knowles 2017). However, this is quite a common drawback for all MT systems in general.

2.2.2 Dictionary-based methods

Dictionary-based methods (also called dictionary look-up) implement word by word translations of queries by using bilingual dictionaries. According to previous literature (e.g. Pecina et al. 2014, Zhou et al. 2012, Volk et al. 2002), this approach has the following drawbacks:

The “out-of-vocabulary words” problem: If the word is not in the dictionary it cannot be translated. In some cases, this problem can be solved with stemming, i.e. conflating words by removing their inflectional or derivational suffixes without morphological analysis of the word, and lemmatization, which aims at finding the base form of a word and implies a morphological analysis of the word. For example "argues", "arguing", and "argued" share the stem "argu" and the lemma “argue”. Grouping different words into a common stem or lemma increases matching possibilities as long as the stem or lemma of the query word is in the dictionary. Another possible solution to this problem is query expansion, i.e. adding synonyms and related words to the query words.

Many translation candidates – ambiguity: If the dictionaries contain many different translation candidates for a source word (ambiguity), then finding the right candidate is a complex task. Sometimes backward translation has been utilized to solve this problem. It implies finding first the candidate translations for each query term in a bilingual dictionary, then submitting the candidate translations in the target language to the dictionary again, comparing the results with the original query in the source language (Boughanem et al. 2002).

In cases where bilingual resources are not available for the source and/or the target languages, linguistic resources in so-called “pivot languages” might be utilized. This means that the source language query is translated into an intermediate language (generally English) and then the query in the intermediate language is finally translated into the target language (where a bilingual lexicon is available). A drawback of this approach is that word sense disambiguation might be needed in both translation steps (from the source to the intermediate and from the intermediate to the target).
Morphosaurus (Marko et al. 2005) is a cross-lingual document retrieval engine for the medical domain that utilizes bilingual lexicons/thesauri. The lexicon/thesaurus entries consist of minimal morphemes (morphologically meaningful word fragments), their attributes, synonym classes and semantic relations between them. The lexicon is partly automatically acquired: first parsing texts from different language resources and then converting them to interlingua representations according to rules and limitations specific for each language. Already existing lexicons in several languages are also utilized in the translation process, as well as tailored string substitution rules between languages (for instance ce-s and c-k between English and Swedish – Iceland/Iceland, Cramp/Kramp). The system also utilizes UMLS (Unified Medical Language System), i.e. a repository of medical knowledge resources that contain a metathesaurus (a database of medical terms and their relationships in several languages), a specialist lexicon and a semantic network of the concepts of the metathesaurus. UMLS is utilized for finding additional lexicon entries and for removal of false friends in different languages, i.e. removal of words with similar strings but with different meanings in each language (for example the English word “blanket” and the Swedish word “blankett”, which means “form”).

Kotsonis et al. (2008) created a cross-language retrieval system to help Greek users to retrieve medical information on the web. The system utilizes a Greek-English bilingual dictionary, created by merging together several Greek-English dictionaries and glossaries available on the Internet. In the translation process, the system considers translations of larger units as more accurate, so it extracts the largest word units/phrases found in the dictionary that matches the query phrases and then provides translations in the target language. The system also utilizes stop words removal (i.e. removal of common words that are not relevant for search purposes such as articles, conjunctions etc.) in order to speed up the translation process, and stemming in order to ease the out-of-vocabulary problem. Pirkola et al. (1998) created a system that applied dictionary look-up translation based on both a generic and a Finnish-English medical dictionary with over 60 thousand entries. The system was tested in CLIR and proved to be as effective as monolingual systems; however, the input queries utilized in the assessment were limited in their structure according to given templates. In order to overcome the “out-of-vocabulary words” problem, the system applies morphological analysis of source language words in Finnish, finding the base form of the words (lemma).

2.2.3 Parallel or comparable corpora methods

A parallel corpus contains documents that are direct translations of each other. The documents can be aligned at the sentence level and the word level. Text pairs that are not exact translations of each other, but are similar or share similar content, are called comparable corpora. The following drawbacks characterize this approach: multilingual domain-specific resources are not available in many languages, and it is time-consuming and costly to build multilingual corpora, aligning them at the sentence and word level.

The easiest way to implement corpora based CLIR (provided access to a parallel or comparable corpus) is to process a query in the source language, find the relevant documents in the source language and then retrieve their exact counterparts in the target language (Kishida 2005). No translation of the original query into the target language is required.

Statistical methods that find similarity likelihood among query terms in source and target texts are often used in the translation process. For instance, methods that compute the co-occurrence of terms in the corpus documents (Gaussier et al. 2000) or calculate translation probabilities of terms from sets of sentence alignments in the parallel corpus (IBM algorithm, Xu et al. 2001). Déjean et al. (2005) utilize a model for extracting bilingual lexicons from both parallel and comparable corpora. Their model first builds context vectors of the source and the target words (i.e. considering all words occurring in a proximity window over several sentences in the corpus). The target context vectors are translated with a general bilingual dictionary and compared to the source vectors with a cosine measure, i.e. comparing their similarity in a vector space (Vector Space Model, Salton & Lesk 1968). They even expand the context vectors with semantic conceptual information from UMLS and MESH\(^\text{11}\) (Medical Subject Headings, i.e. medical terms utilized to index medical journal papers), computing similarities between the corpus words and the terms in the concepts in UMLS and MESH. Volk et al. (2002) annotated user queries and documents in a parallel corpus with part of speech tagging, morphological analysis (even compound analysis for German) and phrase recognition. Their system also identifies UMLS concepts and their semantic relations both in user queries and in the corpus documents, which consist of English-German scientific abstracts from medical journals published on the Springer’s website\(^\text{12}\).

The Cross-Language Evaluation Forum (CLEF)\(^\text{13}\), is a yearly held conference with evaluation labs and workshops that cover a broad range of issues in the fields of monolingual and cross-language information retrieval. Since 2014,

\(^{11}\) https://www.nlm.nih.gov/mesh/, Accessed September 10, 2018
\(^{13}\) http://clef2017.clef-initiative.eu/, Accessed June 4, 2018
CLEF has a multilingual information retrieval track for e-health (Kelly et al. 2014). The aim is to help laypeople, i.e. people without medical expertise, to search and understand health information, which is quite a new approach since most systems in medical CLIR earlier focused on retrieving information for medical experts. The corpus consists of monolingual and multilingual documents originating from medical websites that have been certified by the Health on the Net (HON) foundation. The systems participating in the evaluation campaign submitted queries translated in different languages as test data. Different techniques have been implemented among the systems participating in CLEF: from standard vector space model as baseline (Dramé et al. 2014, Ksentini et al. 2014), to language models (Verberne 2014, Choi & Choi 2014) and concept-based retrieval techniques (e.g. Shen et al. 2014), where medical concepts are identified both in user queries and in documents.

2.2.4 Common challenges

One common problem related to all query translation methods is word sense ambiguity. Ambiguity can refer to both source and target language. It can be solved by comparing several query translation candidates or query expansion (Zhou et al. 2012). In the first case, the query is enriched with several translation options and the best candidate translations are calculated (for example Pirkola 1998; Darwish & Oard 2003). Query expansion implies enhancing the query translation with synonyms, related terms or related concepts. Query expansion can be done either by interacting with the user (user relevance feedback), i.e. the user rates the relevance of the retrieved documents, or automatically (pseudo-relevance feedback), with the help of statistical approaches or external resources that complete or reformulate the original user query.

The systems in CLEF 2014 implement several techniques for query expansion: GRUIM (Shen et al. 2014) utilize mutual information (Fano 1961), i.e. concepts/terms are considered related if they frequently co-occur in the documents, so the original user queries are expanded with the top mutual concepts/terms. Team IRLABDAIICT (Thakkar et al. 2014) utilized CLEF’s discharge summaries (information concerning patients’ medical history) combined with MeSH terminology for query expansion. Team Nijmegen (Verberne 2014) expands queries with the terms in the discharge summaries and

---

15 https://www.hon.ch/HONcode/, Accessed September 13, 2018
16 the probability that a document is a good match to a user query if the content of the document model is likely generate the query (http://nlp.stanford.edu/IR-book/html/htmledition/language-models-for-information-retrieval-1.html, Accessed December 10, 2015)
the UMLS-thesaurus. KISTI (Oh & Jung 2014) re-rank the retrieved documents with abbreviations (expanding abbreviations in the queries and the documents with their full representations) and assume that the top-ranked documents contain relevant concepts/terms, which are then utilized to expand the original query.

2.3 Knowledge patterns

Knowledge patterns were created for architecture by Christopher Alexander et al. (1977) as knowledge units that applied solutions to recurring design problems. Recurring design problems could be solved with recurring solutions, so the authors developed a formal design documentation language where each pattern defined: 1) the design problem; 2) the context where the problem occurs; 3) the competing forces and constraints involved and 4) the solution to the problem, i.e. a generic list of instructions to apply in order to solve the problem in different situations. Borchers (2001) compared knowledge patterns design to participatory design in software engineering, i.e. the process where the end users are involved directly in the software development cycles since the ultimate goal of the knowledge patterns is to engage citizens in the design of solutions that improve their work and living environment.

It is not surprising that this design documentation methodology was then utilized by the Software Development community (e.g. Gamma et al. 1995, Beck and Cunningham 1987, Hohpe & Woolf 2004) in order to support source code reuse. These patterns tended to be more fine-grained and descriptive than in architecture, describing at a lower level of detail how the software modules/objects interact with each other. Patterns were eventually also utilized by interaction designers in human-computer interaction (HCI) since interaction designers noticed that user interface problems tended to re-occur and that these problems could be solved by applying already known solutions. Compared to other documentation technologies such as guidelines, they were considered more straightforward to interpret (Mahemoff & Johnston, 1998). Mahemoff and Johnston (1998) produced patterns for users and user-interface elements. They argued that HCI patterns should be divided into a) patterns with the descriptions of the tasks to be performed by the users in the system (e.g. “view documents”, “upload a file” etc.); b) patterns of users, i.e. patterns that cover users’ domain of expertise, user behaviour (i.e. user profiles); c) patterns of user-interface elements, i.e. patterns that explain to developers when to implement certain interface elements such as scrollbars or menus and d) patterns of entire systems, i.e. patterns that cover the issues related to the
implementation of specific systems (e.g. word processors, aeroplane monitoring systems etc.).

Nilsson (2009) produced a set of User Interface (UI) design patterns specially adapted for mobile applications. The goal of the knowledge patterns was to provide indications for developing more user-friendly applications on mobile devices. He focused on three main problems: 1) screen space utilization; 2) interaction mechanisms; and 3) design at large. The patterns were then evaluated with the help of formative and summative questionnaires applied in tutorials and workshops.

Tidwell (2010) combined user interface best practices and reusable ideas into design patterns that provided solutions to common design problems for mobile applications, web applications, desktop software and even social media. Her design patterns are characterized by a few essential sections, which makes them easy to follow and implement:

1. **Pattern Name**: Contains the name of the pattern or its reference number
2. **What**: This section explains the problem that the pattern addresses.
3. **Use When**: Describes the context in which the pattern can be applied.
4. **Why**: Corresponds to the rationale of the pattern, i.e. it explains why the solution in the pattern is appropriate in relation to the context and the problem.
5. **How**: This is the section that explains the solution to the problem.
6. **Examples**: Contains visual examples of the proposed solution (screenshots).

Tidwell’s patterns are widely recognized by the UI research community.

Guidelines have also been utilized in earlier research (e.g. Zaphiris et al. 2007, Morrell 2005) for formalizing and documenting user interface knowledge. Guidelines are generally divided into two different categories (Zaphiris et al. 2007): theory-driven guidelines (i.e. generated by academic theories) and guidelines based on practical experiences (i.e. generated by the industry). Guidelines focus mainly on providing indications about how to use interfaces coherently (Granlund et al. 2001). They have also been included in knowledge patterns as guides about “how-to-use” the patterns. Compared to knowledge patterns, guidelines lack information about the forces and the logical reasons that influence design decisions, i.e. the context, the users and the rationale

---

17 http://designinginterfaces.com/, Accessed October 14, 2017
(Granlund et al. 2001). They are also more prone to change over time compared to knowledge patterns.

Knowledge patterns have also been utilized in e-health for presenting business models (Mettler & Eurich 2012, Osterwalder & Pigneur 2010), guiding e-health providers and e-health marketers for commercializing e-health services.

There are also other approaches that model medical knowledge for design and implementation of health care systems: medical thesauri and health care ontologies (Juarez et al. 2009). Ontologies are formal representations of domain knowledge and enrich medical thesauri with semantic information, such as relations among concepts and constraints. Juarez et al. (2009) implemented a new model for acquiring/capturing medical knowledge of physicians and its representation based on ontologies. The model incorporates causal, temporal relations and constraints between diseases and other contextual factors - such as age, environment, risk factors etc. - with the goal to provide a knowledge base for hospital information systems.
Chapter 3 Research framework, materials and methods

This chapter describes the research methodology of this thesis. First, the overall research framework (design science) is described, followed by a description of the materials/data and the research methods of the foundational aspects. The chapter then continues with a description of the evaluation methods of the artefact. The final part of the chapter contains a discussion on the limitations of this research and a discussion on ethical issues.
3.1 Research framework

This research covers a broad spectrum of interdisciplinary fields such as computer science, knowledge management, human-computer interaction (HCI), computational linguistics, health informatics and health care. The research paradigm is exploratory and also to some extent causal. It is exploratory since the purpose of our studies is to find out important starting points for the development of online health portals in the future. It is causal since we also study the effects of independent variables (different techniques for user modelling, different techniques for CLIR and lexicon extraction) on dependent variables (end users, the output of the translation and retrieval process, the extracted lexicons) through experimental set-ups.

Considering the variety of disciplines and studies involved, we applied both quantitative and qualitative evaluation methods (see section 3.2).

This research falls within the design science framework, i.e. “the scientific study and creation of artefacts as they are developed and used by people with the goal of solving practical problems of general interest” (Johannesson & Perjons, 2014 p. 13). Design science is a research framework typical for IT and information systems and is a problem-solving paradigm (Hevner et al. 2004), and within its realm people, organizations and technology cooperate (Lee 1999). Depending on the focus of the research both quantitative and qualitative research methods can be applied in order to develop a theory or solve problems (Klein & Myers 1999, Hevner et al. 2004). Both behavioural and engineering problems might be solved with design science, since people, organizations and technology are interacting actors within IT.

Figure 2 shows a model of the design science research framework (DSRF) (Johannesson & Perjons, 2014). Based on our research, we have created an artefact consisting of a set of knowledge patterns for design and development of health portals. The knowledge patterns were based on portal interaction studies (Article I, Article III), integration of existing knowledge in the field of QA and CLIR (Article II, Section 2.3), and experiments with language technology applications (Article IV, Article V, Article VI). They were then developed and evaluated through workshops and surveys (Article VII).

As a test-bed of our research, we chose a portal with multilingual Frequently Asked Questions (FAQs) about mental health (web4health\(^{18}\)).

Figure 2. A Design Science Research Framework (Johannesson & Perjons 2014, p.77). In the white boxes on the left side the knowledge that worked as basis to the activity, in the rounded boxes on the right side the corresponding chapters in the thesis.
The DSRF consists of five activities (see figure 2) but researchers do not necessarily have to perform all the activities in depth (Johannesson & Perjons, 2014), the researcher can decide to focus on some of the activities instead. In this thesis, we have focused on “Design and Develop Artefact” and “Evaluate Artefact”. The requirements are based on published literature.

Explicate problem
The first activity is about identifying a practical problem that motivates the design and development of the artefact. The practical problem is of general interest or driven by curiosity (Johannesson & Perjons 2014). The practical problem of this research is discussed in Chapter 1, section 1.2. The research questions related to the problem are also discussed in section 1.2 and the limitations of this research in section 3.4. Chapter 2 discusses related research in the field of knowledge patterns, CLIR and UM.

Define requirements (outline the artefact and elicit requirements)
The second activity involves two sub-activities: to choose the artefact type to implement and to elicit the requirements of the chosen artefact.

Outline the artefact type
The artefact that we needed to find for this research should be able to easily encode and share knowledge, with a clear and easily understandable structure that motivates design decisions. It should also be applicable to knowledge from different disciplines, specifically language technology, HCI and e-health.

After literature studies, we chose to utilize knowledge patterns as our artefact type. They were chosen since as explained by Granlund et al. (2001) they are easily understood by readers without technical skills and also provide background motivations behind the proposed solution (the “rationale” of the pattern). They are an effective method to capture and formalize knowledge and then sharing it with to non-experts, as shown with recurring architectural solutions for laypeople (e.g. Alexander et al. 1977). Knowledge patterns have also successfully been utilized in design and best practice solutions in the disciplines of our research, i.e. health care (Persson et al. 2008), HCI (Tidwell 2010) and language technology (Nuzzolese et al. 2011).

Guidelines are another common tool for sharing knowledge. They are particularly useful for formalizing and documenting user interface knowledge (e.g. Zaphiris et al. 2007, Morrell 2005). However, compared to knowledge patterns, guidelines lack background information about the logical reasons that influence design decisions, (i.e. the context and the rationale) and are also more prone to changes over time (Granlund et al. 2001).
Elicit requirements

Once we selected the type of artefact to utilize, the next step was to elicit the requirements concerning the content of the artefact, i.e. the knowledge patterns. The following requirements were in focus concerning the content: usefulness (Rolland et al. 2000), relevance (Asrofah et al. 2010, Zhu et al. 2007), adoptability (Rolland et al. 2000), coherence (Rolland et al. 2000), consistency (Rolland et al. 2000), granularity (Rolland et al. 2000) and comprehensibility/comprehensiveness (Johannesson & Perjons 2014, Niwe & Stirna 2010). The patterns should also be based on results of scientific foundation/contribution (Hevner et al. 2004).

Usefulness and relevance: Since the knowledge gap addresses three research foundations 1) integration and review of existing knowledge in language technology; 2) studies of portal interaction and 3) experiments with language technology applications, the artefact should be useful in providing a relevant contribution for solving portal design problems related to each research foundation.

Adoptability: The artefact should also be adoptable in real situations by designers of health portals for resolving a particular problem of interest in the fields of HCI, language technology and e-health.

Coherence: The artefact should be coherent, i.e. not contain contradictions, and make sense on its own, i.e. the artefact should be a self-contained piece of knowledge.

Consistency: The artefact should also be consistent and conform to existing knowledge and vocabulary in the fields of HCI, language technology and e-health.

Granularity: The artefact should address a problem at an adequate level of detail in the solutions, i.e. not too coarse or too detailed.

Comprehensibility/Comprehensiveness: The artefact should also be easily comprehensible, i.e. easily understood or comprehended by their user.

Scientifically founded: In order to strengthen the value of the proposed solutions, the content of the artefact should be the result of scientific research confirmed by scientific publications in the field of health informatics, information systems and HCI.
**Design and develop artefact**

This is the third activity in the DSRF, and it deals with design and development of the artefact that solves the research problem. The resources utilized in this activity involve knowledge from research literature and knowledge coming from artefacts or practice\(^{19}\) (Johannesson & Perjons 2014).

The document template utilized for our knowledge patterns is the enterprise knowledge patterns (EKP) approach and format (Bubenko et al. 2001, Persson et al. 2008). We chose this specific format since it has already successfully been utilized in documenting design solutions and best practice in health care (Persson et al. 2008).

The problems and the solutions embedded in the templates were based on the scientific results of the three research foundations:

1) User studies of portal interaction: two studies were performed; first a study of how psychological information seekers utilize and experience health portals (Article I), then a comparative study that elicits the differences between retrieval results with different personalized search services and with or without profiling of the user (Article III). The results helped to suggest solutions to interface problems and functionality problems such as how to capture user attention, how users with different search goals experience the published information, how to facilitate search services and how users experience personalized search.

2) Knowledge integration of language technology approaches: the results helped to define the language technology approaches that fit health portals better, suggesting solutions to the choice of QA and CLIR approaches for the health domain (Article II, Section 2.3).  

3) Experiments with language technology applications in the domain of psychology and psychotherapy: we conducted best practice studies about methods for extraction of bilingual Swedish-English lexicons from bilingual texts (Article IV, Article V) and experiments with different CLIR techniques on a health portal (Article VI). The results suggest solutions to the choice of language technology approaches, search engines, text processing and the types of lexicons that better fit the domain.

The methods utilized for data collection and analysis are explained in section 3.2.2. The problems and the solutions were then formally embedded in the

---

\(^{19}\) According to Adler, E., & Pouliot, V. (2011) practice is performance, the act of doing something, which relies on background knowledge and it is patterned, i.e. tends to rely on actions that are repeated over time and space.
EKP documents, which were iteratively refined and categorized with formative workshops with both knowledge pattern experts and domain experts (Article VII). This led to the creation of 20 knowledge pattern drafts that were originally classified in two large groups depending on their content: HCI patterns and language technologies-information retrieval patterns. A mixed-methods approach was then employed for the further development and evaluation of the knowledge patterns (see chapter 5 Evaluation of the knowledge patterns).

Demonstrate artefact

This activity is “proof of concept” of the artefact, i.e., it shows that an artefact can solve parts of a problem or cases from literature or real life (Johannesson & Perjons 2014). The artefact demonstration can also be defined as lightweight evaluation (Johannesson & Perjons 2017), and in some design science guidelines (e.g., Hevner et al. 2004) this activity is not explicitly mentioned since it is considered part of the evaluation activity. We decided to concentrate on performing a thorough evaluation activity (see next section Evaluate artefact) that demonstrated the validity of the artefact—assessing its usefulness, relevance and adoptability in the domain (Article VII).

Evaluate artefact

This activity evaluates the utility of the artefact, its requirements and the knowledge formalized in the artefact (Johannesson & Perjons 2014). It also aims to find opportunities for further improvement. This activity consisted of both formative evaluation workshops and summative evaluation of the artefact (Article VII).

A mixed-methods approach was employed for the further development and evaluation of the knowledge patterns (see section 3.3). First, a formative evaluation (Johannesson & Perjons 2014) was employed with the purpose of improving with iterative steps the patterns during the design process. Then a summative evaluation (Johannesson & Perjons 2017) was employed in order to get a final assessment of the patterns.

In the formative part, four semi-structured participatory workshops (Rolland et al. 2000, Chung et al. 2004) were organized with three knowledge pattern experts and three domain experts in the field of language technology, health and human-computer interaction. The summative evaluation was performed with an online survey assessed by a non-randomized set of international domain experts in the areas of natural language processing, information retrieval, and HCI.

20 A formative evaluation implies that an artefact is evaluated while it is still being designed/developed.
21 A summative evaluation produces a final assessment of an artefact after it has been designed/developed.

human-computer interaction, and e-health. The domain experts were selected from established experienced professionals from both the industry and academia and were active in research communities in different countries (Sweden, Finland and Australia).

Earlier process models for design science (e.g. Takeda et al. 1990) ended with the evaluation of the artefacts and the conclusion of the researcher. However, modern design science models (e.g. Peffers et al. 2007) consider communication of the results through scholarly and professional publications as an important step of the research. Our research follows the latter approach since our results have been communicated and presented in conferences in the field of information systems (Articles II, III and IV), e-health (Articles I, VI), World Wide Web (Article V), or accepted for publication in health informatics journals (Article VII).

3.2 Design science research framework: design and develop artefact - materials and research methodology for foundational aspects

This section provides a description of the materials and the methods utilized to gather data for the foundational aspects of this research.

3.2.1 Materials

The test-bed

The knowledge domain utilized in this research consists of health information in the field of psychology and psychotherapy. The knowledge base (i.e. the medical content) is part of a health portal (web4health.info) developed within an EU-s financed project\(^\text{22}\), with the goal to improve the mental health of European citizens with the help of multilingual health information provided by psychiatrists and psychotherapists from five European countries (Italy, Sweden, Holland, Greece and Germany). We chose web4health for the following reasons: 1) it is a mental health portal available for research purposes; 2) it implements CLIR as well as different search GUIs and 3) the medical content is written by medical professionals and based on medical research.

The knowledge base consists of semantically classified Web pages that cover the following topics (from Andrenucci 2005):

“(1) Eating disorders and obesity.
(2) Psychological obstacles to achieving healthy living habits such as unhealthy eating and substance abuse (e.g. drugs and alcohol).
(3) Psychological problems causing a lower quality of life through despair and inability to work.
(4) Life problems such as marital and interpersonal relationship problems.”

The medical content consists of FAQ (Frequently Asked Questions) entries, i.e. manually and semantically annotated web pages. Every FAQ object encloses a question-answer pair, where the question part consists of a pre-defined pattern/template created to match variations of similar queries (see figure 3).

Users search information on web4health either with natural language queries through a search panel or by manually navigating the taxonomies of the topics covered by the portal. The taxonomies are hierarchical and users navigate the hierarchy downwards, i.e. starting from general items and clicking further to more specific topics.

![Figure 3. Alternative forms of queries for the same FAQ, from Andrenucci (2005)](image)

**The user model**  
The user model utilized to encapsulate user preferences (Article III) is both explicit and implicit, according to the definition of Kass and Finin (1998). The explicit (or direct) profile is created by letting the user submit directly his/her
topics of interest - ranking them in order of relevance - and the search goals with the help of a menu-based form (see figure 4).

The implicit (or indirect) user model is created by monitoring the queries submitted by the user and the topics of the documents retrieved by the system, so-called “learning from observation” (Pohl & Nick 1999). The system first computes the topics and the search objectives that occur more often among the user queries and the retrieved documents. Then it presents the inferred information to the user with the help of a feedback panel, so the user can accept, change or even discard the inferred profile (see figure 5). This information allows the user to always be aware of the UM and puts him/her in control of the encapsulated data.

The topics of interests correspond to diseases and classification categories for the FAQs in the knowledge base of the portal (Andrenucci 2005, p. 44 and 45), while the search purposes match a well-known medical taxonomy (Bader & Theofanos 2003, Bhavnani et al. 2003, White 2000) with five medical terms: Terminology, Risks/Effects, Prevention, Diagnosis, Treatment and Prognosis. Another term (“Causes”) was added to the list after analysing the weblog of the portal; this term was recurrent among the user queries submitted to the search panel. In order to assess whether user profiling enhanced the IR quality (see Article III), the retrieved documents were listed in two columns (with and without user profiling), without any information about the retrieval

Figure 4. The menu-based form for explicit UM, from Andrenucci (2005)
algorithm utilized, not to bias the judgmental work of the participants (blind experiment, Chin 2001).

The parallel corpus
A parallel corpus is a collection of texts in two or more languages, where each text is an exact translation of the original. The parallel corpus of our research is the knowledge base of the web4health portal, which consists of FAQ objects with question/answer pairs. The Swedish corpus consists of circa 135,301 words and the English counterpart of circa 143,118 tokens. Before utilizing the texts we proofread the material and corrected it to ensure its completeness and correctness. We also removed HTML-tags from the text and converted it to plain text.

For the pilot study described in Article IV, we utilized a subset of the corpus (a randomly selected set of FAQs) where the Swedish part consisted of circa 12,800 tokens and the English counterpart of circa 13,000 tokens.

A version of the bilingual corpus was lemmatized with the CST Lemmatizer (Jongejan & Haltrup, 2005) - a rule-based lemmatizer for languages that utilize inflectional suffixes, such as Swedish and English. The Trigrams’n Tags tagger (Brants 2000) was utilized to annotate the POS tagged versions of the

---

23 “tokens” refer to the total number of words in a corpus, not the number of unique words which are called “types”
corpora. The tagger was trained on Swedish (Megyesi 2002) using the Stockholm Umeå Corpus\textsuperscript{24}, and utilized the PAROLE annotation scheme\textsuperscript{25}, a tag set that includes POS and morphological features of the words. The Penn Treebank corpus and its tag set (Marcus et al. 1993) were utilized for the English language. For syntactic parsing of the English texts, we utilized the open source tool GROK\textsuperscript{26}. The Swedish texts were parsed with a context-free grammar parser developed at Uppsala University (Megyesi 2002).

**The bilingual dictionaries**

In order to assess the quality of cross-language information retrieval (Article VI) we utilized three dictionaries (two developed for this research), two machine translation systems and two search engines (see figure 6).

The three dictionaries were:

1) A generic bilingual lexicon of approximately 50,000 entries (Lexin\textsuperscript{27}, utilized as the lexicon baseline).

2) A domain-specific dictionary extracted from the bilingual texts of web4health (as described in Articles IV and VI), approx. 14,000 entries (word alignment lexicon).

3) A domain-specific lexicon based on the English words from the FAQ templates of web4health, approx. 16,000 entries (template-based lexicon).

The lexicon extracted from the web4health’s parallel was part-of-speech tagged and syntactically parsed with the Uplug Toolkit (Tiedemann 2003), i.e. a collection of tools for processing parallel texts, which align texts at the sentence and word level. The template-based lexicon (approx. 16,000 entries) was created utilizing a backward translation approach, i.e. translating the content of the templates in the target language to the source language. For machine translation we utilized two commonly available systems: Google Translate\textsuperscript{28} (through its public API) and Systran\textsuperscript{29} (through an API that was available for our research group). The queries translated with the different methods were then submitted to two different search engines: Google Site Search\textsuperscript{30} and Quick Ask (Sneiders 2002a), which altogether generated ten different results lists based on the different combinations of translation methods and search

\textsuperscript{24} SUC, 1997. SUC 1.0 Stockholm Umeå Corpus, Version 1.0. Umeå University and Stockholm University, Sweden. It has a total of one million words.

\textsuperscript{25} https://spraakbanken.gu.se/swe, Accessed August 5, 2015


\textsuperscript{27} http://folkets-lexikon.csc.kth.se/folkets/folkets_en.html, Accessed January 9, 2015

\textsuperscript{28} https://translate.google.se/, Accessed September 10, 2017

\textsuperscript{29} http://www.systransoft.com, Accessed November 11, 2015

engines. The architecture of the system utilized to experiment with cross-language information retrieval is shown in figure 6.

The knowledge patterns
The results of knowledge integration of language technology approaches, user studies of portal interaction and language technology experiments, were eventually formalized into 14 knowledge patterns that are available at the following URL: http://www.dsv.su.se/~andrea/knowledgepatterns.pdf

Their development and evaluation are described in Article VII. The document template utilized for our knowledge patterns is the Enterprise Knowledge Patterns (EKP) approach and format (Bubenko et al. 2001, Persson et al. 2008). There are several templates for design patterns which are quite similar to EKP (e.g. Gamma et al. 1995, Hohpe & Woolf 2004). Their focus is on object-oriented source code reuse, and their outline contains sections which are not applicable to this research since our knowledge patterns do not provide solutions at the source code level. For example, the sections “Sample code”, “Participant classes and/or objects”, “Structure: graphical representation of the classes and objects”. Other sections in the format of Gamma et al. (1995) have different names but define the same content as the sections in the EKP tem-

Figure 6. The architecture of the CLIR-system.
Plate. For example the sections “Intent” and “Motivation” correspond to “Rationale” and “Context”. Other sections are identical in all formats, e.g. “Pattern Name” and “Consequences”. Table 1 shows the outline of the EKP template that was utilized to formalize the knowledge. The table includes an example in the rightmost column. Sections that included relevant references and pattern classification information were added to the original template.

Table 1. The outline of the Enterprise Knowledge Pattern (EKP) template (Bubenko et al. 2001, Persson et al. 2008) with some modifications (from Andrenucci 2018). The rightmost column shows an example.

<table>
<thead>
<tr>
<th></th>
<th>The name reflects the problem/solution that it addresses.</th>
<th>Empathetic text and answers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td>This section defines the group the pattern belongs to. We have classified the patterns into four groups: language technology patterns, interface patterns, functionality patterns and user role patterns.</td>
<td>User role patterns</td>
</tr>
<tr>
<td><strong>Founded on</strong></td>
<td>Defines the research foundation the pattern is built upon.</td>
<td>User Studies of portal interaction</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>Describes the issues that the pattern addresses within the given context and forces.</td>
<td>Information with unsympathetic/non-empathetic attitude and higher grade of technicality is rejected more easily by people suffering from mental distress (Andrenucci &amp; Forsell 2004, Andrenucci 2006, De Carolis et al. 1996). This might hinder users from finding and assimilating information that could solve their problems.</td>
</tr>
<tr>
<td><strong>Criteria and Context</strong></td>
<td>Describes the preconditions under which the problem and the proposed solution seem to occur.</td>
<td>Users suffering from mental health problems and users without mental health problems have different search goals, are interested in different levels of technical details and appreciate different tones/styles in the text (Andrenucci &amp; Forsell, 2004, Andrenucci 2006). A mental health portal should adapt the content to these different types of information seekers.</td>
</tr>
</tbody>
</table>
| Rationale and Solution | Rationale: Explains why the solution in the pattern is appropriate in relation to forces, context and problem. | In order to deliver and adapt the texts/answers for different information seeking users, it is necessary to encapsulate users’ search goals and their cognitive condition in a user profile.

For information seekers who are suffering from mental distress and are interested in finding a solution to their problem, the texts should be written with an empathetic tone and with a lesser degree of technicality, since details are less important for this category of users. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution: Describes how to solve the problem and to achieve the desired results. It can be expressed in natural language, drawings, images etc.</td>
<td>Solution can be backed up with references to other knowledge sources and other patterns.</td>
<td></td>
</tr>
<tr>
<td>Forces and Influences</td>
<td>Describes the relevant forces and external influences and how they interact or conflict with one another and with the goals we wish to achieve by implementing the solution.</td>
<td>Information seekers suffering from mental health problems are interested in empathetic texts written in an empathetic tone, but other users may prioritize detailed and technical texts.</td>
</tr>
<tr>
<td>Consequences</td>
<td>Describes what the context should be after applying the presented solution.</td>
<td>Providing texts with empathetic tone and with a lesser degree of technicality may help people suffering from mental distress to find and assimilate information that might improve their health condition. Texts with an empathetic tone and a lesser degree of technicality might be uninteresting for information seekers with in-depth knowledge or users without health problems, due to their priority for medical content with a higher level of details.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
<th>References to other authors’ publications.</th>
</tr>
</thead>
</table>

3.2.2 Research methodology for the foundational aspects

In this research, we implemented both quantitative and qualitative research methods. Quantitative methods are generally associated with positivism and their “objective” view of reality, where the observed facts or phenomena can be described and statistically quantified. Positivism sees reality as an own entity that exists independently of human actions and experiences (Johannesson & Perjons, 2014). Positivist research follows the deductive reasoning and it is often related to natural science. Qualitative methods on the other side are often linked to post-positivism or interpretive research (Lillis & Mundy 2005), because the view of reality has a more “subjective” value according to this approach. Reality can be interpreted by the people involved in it, and the interpretations may change over parameters such as time and circumstances (Orlikowski & Baroudi, 1991). Post-positivist research is often based on inductive reasoning and is common in social science.

Quantitative methods were utilized for assessing the information retrieval (IR) performance of the implemented applications (see sections 3.2.2.4 and 3.2.2.5) and in the summative evaluation of the knowledge patterns (see section 3.3). The quantitative methods utilized to assess the IR performance adhere to Cranfield paradigm (Voorhees 2001), which has its origins in the Cranfield experiments in the sixties for evaluating the performance of text indexing systems (Cleverdon 1967) and it has been implemented in natural language processing (NLP) conferences such as TREC\textsuperscript{31} and CLEF\textsuperscript{32}. Systems are evaluated with laboratory testing on “the same set of documents and the same set

\textsuperscript{31} http://trec.nist.gov/, Accessed April 5, 2008
\textsuperscript{32} http://www.clef-campaign.org/, Accessed June 4, 2011
of information needs, and for the use of both precision and recall (see section 3.2.2.4) to evaluate the effectiveness of the search” (Voorhees 2001).

The qualitative methods were utilized for studying how users experienced interacting with the different GUIs of the portal. The studies were based on user observations and interviews (see section 3.2.2.1). We also applied qualitative approaches in the integration of knowledge within QA and CLIR (see section 3.2.2.3) and in the formative and summative evaluation of the knowledge patterns (see section 3.3).

3.2.2.1 User studies - user observations and interviews (Article I, Article III)

An interview is defined by Johannesson & Perjons (2014, p. 57) as a “communication session between a researcher and a respondent, where the researcher controls the agenda by asking questions to the respondent”. In a structured interview there is a pre-defined set of questions with a pre-defined set of answers that the respondent can choose from. In an unstructured interview, the researchers do not use pre-defined questions, but rather let the respondent discuss freely about a certain topic. Semi structured interviews are in between structured and unstructured interviews: there is a pre-defined set of questions, but the respondent can answer “freely” in her own words without choosing from a pre-defined set of answers.

User observations allow researchers to study the behaviour of respondents in practice or while performing a task. The researcher can identify difficulties or problems that respondents face while interacting with GUIs and that users might be unaware of (Johannesson & Perjons 2014, p. 96).

In order to follow the perspective of end users and to find out how they experienced information seeking and the usage of web portals, we decided to utilize qualitative research methods based both on user observations and interviews. Qualitative research methods of this kind are recommended when it comes to evaluating user satisfaction with GUIs and information retrieval systems (Nielsen 1994, Park 1994). They have also been implemented in studies covering information retrieval of medical information on the web (e.g. Eysenbach & Köhler, 2002), usability issues and medical information seeking on health portals (e.g. Gurel & Cagiltay 2012). During the user observation sessions, users were encouraged to “think aloud” (Long & Bourg 1997) in the process of interacting with the portal. The “thinking-aloud” methodology implies that informants verbalize to the researchers their feelings and their thoughts in the process of performing tasks such as interacting with a GUI or submitting information requests to a search engine. This allows the researcher to get insight into the cognitive process of the informants and better understand their seeking behaviour. This technique has been successfully utilized
in projects evaluating users’ health information search processes (Huang et al. 2012), user-system interaction (Griffiths et al. 2002) or developing self-help tools for e-health (Lee et al. 2006).

It was decided to utilize interviews as a complement to user observations for the following reasons:

- Interviews are recommended in cases where a deeper understanding of the users’ benefit of the system is needed (Årsand & Demiris 2008). Considering that users should benefit from the health information retrieved, this is particularly important for our research.

- Interviews tend also to provide more complete responses from the informants. This is possible since probing can be used, misunderstandings can be clarified “live” through follow-up questions, and the respondents can provide explanations in their “own words” (Williamson 2002, p. 244). Interviews are also more suitable than other methods for capturing emotions, experiences, opinions and attitudes (Johanneson & Perjons, 2014).

The interviews in the first study (Article I) were semi-structured and unstructured since we wanted to let the users freely describe how they experienced the GUls of the portal and explore their acceptance of mental health information and psychological counselling services online.

The interviews in the second study (Article III) consisted of one structured and one semi-structured part. The semi-structured part aimed at gathering the user’s general opinion about the interfaces, the user profiling approaches, the type of personal information they were prepared to disclose and the characteristics that users consider important in a health portal. The structured part aimed at quantifying how users rated the retrieved documents with and without the user profile, the preferences of the users about different GUls and different profiling approaches (see next section 3.2.2.2). So even if we applied a qualitative approach overall, there was also a quantitative element.

The target population of our research consists of people without medical knowledge who “consume” health information online, so-called health informatics consumers (Hung et al. 2013). Since sample groups in qualitative research must be representative of the target populations (Grover & Greer 1991), two sample groups among health informatics consumers were identified:

1. Users with mental health problems in need to solve their problems.
2. Users without health issues searching for information in the domain of psychology and psychotherapy.
In our studies, we thus included informants that were representative of the afore-mentioned sample groups and whose age had a span that reflected the diversity of the health consumer population.

In the first study (Article I) our group of informants consisted of twenty-two individuals: twelve individuals, three men and nine women between 17 and 60 years of age, undergoing psychotherapy in a private practice in Stockholm and ten individuals, four men and six women between 17 and 45, without health issues and who had never contact with psychotherapy before. For the selection of the first group, we contacted patients with the help of the psychotherapist in charge of cognitive therapy sessions at the private practice. For the selection of the second group, we reached visitors at a health and sport’s club in Stockholm.

In the second study (Article II) twenty participants were included: Ten individuals, three men and seven women between 23 and 58 years of age undergoing psychotherapy and ten individuals, four men and six women between 25 and 59, without health issues and any previous contact with psychotherapy. For the selection of the first group, we contacted patients as in the case of the previous study. For the selection of the second group, we contacted students, researchers and employees at Stockholm University.

Although the participants had different backgrounds and professions, they shared two common traits: they were all persons without professional medical expertise, and they were experienced Internet users.

All participants were thoroughly informed about the content and the purpose of the studies. They gave their consent orally and they participated in our research projects on a voluntary basis.

3.2.2.2 User studies - evaluation methods for graphical user interfaces, Lickert scales (Article III)

The retrieval performance of the implementations described in Article III (see section 3.2.2.4) was analysed with a standard statistical measure for IR: precision (Salton & McGill 1983), which measures the fraction of the retrieved documents that are relevant. To confirm or reject the statistical IR measures of precision, users rated also the quality of the retrieved information with the help of a four-level Lickert scale. The users provided the rate during the interview.

Lickert Scales are defined as “rating scales on which respondents to a research study are asked to rate their preferences or the frequencies of their activities” (Williamson 2002, p. 332).
It was decided to utilize Lickert Scales since they are very common in HCI evaluations (Mamykina et al. 2001, Ford & Gelderblom 2003, Petrie et al. 2004, Onibere et al. 2001) and are particularly useful in measuring users’ opinions (Williamson 2002, p.237).

Generally, Lickert scales are utilized with five levels (Petrie et al. 2004, Ford & Gelderblom 2003), where the third level includes a “no preference/undecided” choice. Users in our studies rated the GUI and the retrieved information through a four level scale. We opted for four levels in order to make users take a stand and express an explicit preference, instead of selecting an alternative in the middle that provides an “easy way out” during the interview.

Users’ ratings were based on the following template:

<Statement>

<Four-level scale>
1) I strongly disagree
2) I disagree
3) I agree
4) I strongly agree

3.2.2.3 Knowledge integration (Article II and section 2.2)
Marshall and Rossman (1995 p. 28) define literary reviews as a tool “to build a logical framework for the research that sets it within a tradition of inquiry and a context of related studies”. Williamson (2002 p. 62) underlines that “a thorough search of the literature enables the researcher to identify the gaps in previous research and thus justify a proposed study in relation to a demonstrated need”. Furthermore, Williamson (2002 p. 61) pinpoints that “a literature review should include evaluative and critical judgements about the literature, and that it should present a comparison of ideas and research findings tying them together”. These are the approaches that were kept in mind while collecting reports and articles, and in the process of comparing and analysing the material.

In addition to the CLIR techniques, the QA approaches were also investigated for the following reasons:

1) A step towards the QA paradigm is the development of document retrieval systems into passage retrieval systems. This implies that the focus is on retrieving text passages or text chunks rather than entire text documents. Passage retrieval is now a standard component of modern IR-based QA applications in health applications (e.g. Gobeill et al. 2009, Lee et al. 2006).
2) One of the most fundamental issues, when health users intend to search for information on health portals, is how to express oneself to retrieve the most “useful” information (Stanton et al. 2014). According to Liu et al. (2013), the more familiar users are with a search engine, the more complex and question-like queries they tend to submit, which requires the retrieval of more focused/specific text passages rather than entire documents.

3) Mental health care is moving further from traditional visits at medical centres towards e-health services (Webb et al. 2008, Christensen & Hickie 2010a). If health portals aim to implement automatized psychological counselling services, this paradigm cannot be ignored since it transfers the patient-doctor interaction online.

How the knowledge integration was conducted
The literature review was conducted on articles within CLIR and QA. The search for relevant articles was performed on established publication portals such as PubMed33, ACL proceedings34, ACM digital library, Google Scholar35, CLEF36 and relevant highly referenced publications from the papers. The search was based on keywords such as: “cross-language information retrieval in health domain”, “cross-language information retrieval in health portals”, “cross-language information retrieval in psychology, psychotherapy”, “dictionary translation for medicine”, “machine translations in health domain”, “bilingual lexicons in health domain”, “question-answering in health portals”, “question-answering and the health domain”, “question-answering in psychology, psychotherapy”. The literature review of articles within CLIR was conducted on papers from 1997 to 2016, emphasizing recent publications (2012-). The literature review of articles within QA was conducted 1995 on articles from 1980 to 1995 but its results are still actual and the Article (II) has good citations rates in scientific publications. Even later review articles (e.g. Mollá & Vicedo 2007) also discuss some of the analysis and arguments covered in our article, focusing on QA in open and restricted domains (see section 4.2).

The literature review resulted in the definition of three major approaches in QA and CLIR, respectively, and the description of the distinctive features of each approach. The analysis is qualitative and discusses the different techniques from points of view such as reliability of the retrieved results, portability on web applications, scalability on restricted domains, suitability for advice-giving and decision support domains. Quantitative parameters defining the retrieval performance of the systems are also considered in the evaluation.

36 https://sites.google.com/site/clefehealth2014/, Accessed May 10, 2017
3.2.2.4 Language technology experiments - evaluation methods for information retrieval (Article III and Article VI)

According to the Cranfield paradigm (Voorhees 2001), three distinct components are generally utilized in the assessment process: the documents or the objects from a knowledge base or collection, the statements of information needs (user queries/ user questions), and a set of relevance judgments. The relevance judgements can be done “manually”, with the help of a pre-selected group of human assessors, or “automatically”, with the help of pre-compiled referenced data. We utilized the first approach in the evaluation of retrieval results (Articles III and VI) and the second approach in the evaluation of extracted lexicons (Articles IV and V, see section 3.2.2.5).

The relevance choice in the implementation of the paradigm has been binary: either a document/object was relevant or not relevant, following the approach of other evaluation conferences (e.g. TREC).

We complemented the statistical data with qualitative data analysis to reach a deeper understanding of the retrieval results.

The retrieval performance of the implementations described in Article III was analysed with a standard statistical measure for IR: precision (Salton & McGill 1983). We focused on the top five and ten retrieved documents, so we utilized precision at a cut-off of five documents (P@5) and precision at a cut-off of ten documents P@10, which define respectively the proportion of top five and top ten documents that are relevant.

Recall measures the ability of IR systems to retrieve as many relevant documents as possible from a corpus, repository or website: it is the fraction of the relevant documents that have been retrieved. Precision instead measures the ability of IR systems to recover only relevant documents: it is the fraction of the retrieved documents that are relevant.

If RETR is the set of retrieved documents by a system and REL the set of relevant documents on a website, repository or database, recall R and precision P are then defined as follows (Baeza-Yates & Ribeiro-Neto 1999):

\[
R = \frac{|RETR \cap REL|}{|REL|}
\]

\[
P = \frac{|RETR \cap REL|}{|RETR|}
\]
We utilized both quantitative and qualitative methods when evaluating the performance of the different query translation approaches (Article VI). To assess the retrieval results, two standard evaluation measures were used: mean average precision (MAP) (Voorhees & Harman 2005) and precision at a cutoff of five documents (P@5). The main advantage of MAP is that it computes precision and recall at every position in the ranked list of documents, thus taking into the calculation even the position of the relevant ranked documents. Mean average precision is the most common summary measure (Volk et al. 2002) and it is utilized for assessing TREC experiments in Cross-Language Text Retrieval (Gaussier et al. 1998). Both P@5 and MAP are generally utilized for assessing binary relevance, i.e. either a document is relevant or not (e.g. 1 or 0), unlike other assessment methods such as Discounted Cumulative Gain (DCG) or Normative Discounted Cumulative Gain (NDCG) (Jarvelin & Kekäläinen 2000) that are generally utilized for assessing multiple levels of relevance (e.g. 1-4) (Kekäläinen 2005).

Relevance assessment of the retrieved results was performed by three human assessors, two information retrieval experts and one mental health expert. Even in this case, the judgement was binary, a retrieved document was considered either relevant or not.

We also performed a qualitative analysis of the quality of the user query translations. The translations were scanned in detail evaluating how lexical, grammatical and semantical differences in the texts influenced the search results. The qualitative analysis was performed as a complement to the quantitative results.

As input data we utilized 100 user queries randomly extracted from web4health’s input log: 50 single-word units (SWUs\textsuperscript{37}) and 50 multi-word units (MWUs).

3.2.2.5 Language technology experiments - evaluation methods for lexicon extraction (Article IV and Article V) There are two main evaluation methods for assessing the quality of alignments at the word level (Ahrenberg et al. 2000): 1) automatic evaluation with the help of reference data, so-called gold standard, which is compiled before the alignment process and then compared to results; 2) a manual evaluation on sample data performed by experts after the alignment process. We opted for automatic evaluation, since it allows more control and flexibility over the reference data selection, for instance, the choice of word types and words with specific frequency ranges (Merkel 1999).

\textsuperscript{37} Compound words were included in this category
As reference data, we included phrases and content words – e.g. nouns, verbs, adjectives i.e. words with a meaning of their own - randomly sampled from specific frequency ranges. This approach is standard for this type of evaluation (Ahrenberg et al. 2000).

The gold standard in the pilot study (Article IV) consisted of two parts: in the first part, words and phrases were randomly selected from a sample corpus of web4health, and, in the second part words and phrases were randomly extracted from the log of user queries submitted to the portal. Each part consisted of 130 entries with the following ranges: 10 entries with a frequency above or equal to 10, 30 entries with frequency 7-9, 30 with frequency 5-6, 30 with frequency 3-4 and 30 with frequency 1-2. Related research (e.g. Dalianis et al. 2009, Charitakis 2007) utilized coarser frequency ranges (e.g. 1-2, 3-10 and >10) but we opted for more fine-grained frequency ranges to see if there were differences for the results of the frequencies 3-4, 5-6, 7-9.

The gold standard in the follow-up study (Article V) was compiled by randomly selecting phrases and words from the corpus (200 entries, 100 SWUs and 100 MWUs). In this case as well, a fine-grained frequency-balanced approach was chosen with the following frequency ranges: 40 entries with a frequency above or equal to 10, 40 entries with frequency 7-9, 40 with frequency 5-6, 40 with frequency 3-4 and 40 with frequency 1-2.

The links included in both studies were of types: regular, fuzzy (partly overlapping) and null (i.e. omissions).

Since word alignment can be defined as a retrieval problem (Ahrenberg et al. 2000), it was decided to apply performance metrics from IR in the evaluation of the extracted word-relations. The most common metrics in IR are precision and recall. Precision is the ratio of correctly aligned items in relation to the number of aligned items while recall is the ratio of correctly aligned items in relation to the number of correct items. A drawback of these metrics is that they work well with alignments of SWUs, but not with MWUs (Tiedemann 2003 p. 26), since they consider links as entirely correct or entirely wrong. For example, the alignment “barndomsupplevelser – experiences” would be considered entirely wrong and not partly correct compared to “barndomsupplevelser – childhood experiences”.

Since MWUs were included in our reference data and we needed to handle partially correct links, refined versions of these metrics (see Tiedemann 2003, p. 68) were applied in order to measure the degree of correctness of the alignments. The metrics calculate a partiality value Q according to the following formulas:
\[
Q_x^{\text{precision}} = \frac{|\text{aligned}^x_{\text{src}} \cap \text{correct}^x_{\text{src}}| + |\text{aligned}^x_{\text{trg}} \cap \text{correct}^x_{\text{trg}}|}{|\text{aligned}^x_{\text{src}}| + |\text{aligned}^x_{\text{trg}}|}
\]

\[
Q_x^{\text{recall}} = \frac{|\text{aligned}^x_{\text{src}} \cap \text{correct}^x_{\text{src}}| + |\text{aligned}^x_{\text{trg}} \cap \text{correct}^x_{\text{trg}}|}{|\text{correct}^x_{\text{src}}| + |\text{correct}^x_{\text{trg}}|}
\]

Where \( \text{aligned}^x_{\text{src}} \) is the set of source language words and \( \text{aligned}^x_{\text{trg}} \) the set of target language words in link proposals for a reference link \( x \) in the reference data set. \( \text{correct}^x_{\text{src}} \) and \( \text{correct}^x_{\text{trg}} \) are the sets of source and target words of reference link \( x \).

Precision (\( P \)) and recall (\( R \)) are then defined using \( Q \), where \(| \text{aligned} |\) is the total number of correct, incorrect and partially correct links in relation to the reference data and \(| \text{correct} |\) is the size of the reference data set.

\[
R = \sum_{x=1}^{x} \frac{Q_x^{\text{recall}}}{|\text{correct}|}
\]

\[
P = \sum_{x=1}^{x} \frac{Q_x^{\text{precision}}}{|\text{aligned}|}
\]
3.3 Design science research framework: evaluate artefact - evaluation of the knowledge patterns

Figure 7 shows the workflow of the construction, design and evaluation of the knowledge patterns where the grey boxes are the outcome in each work phase.

As mentioned earlier, a sequential mixed-methods approach was utilized for development and evaluation of the knowledge patterns (Article VII). First, a formative evaluation (Johannesson & Perjons 2014) and then a summative evaluation were employed. The primary rationale behind this choice of a sequential mixed-methods approach was to achieve complementarity - to increase the validity of the overall results, as well as to minimize the risk of method bias in the end result.

The formative evaluation was employed to refine the material and gain unexpected insights (Driscoll et al. 2007) in the formative phase, before submitting the material to the final assessment phase. In this part, four semi-structured participatory workshops (Rolland et al. 2000, Chung et al. 2004) were organized inviting three knowledge pattern experts and three domain experts in the
field of language technology, e-health and HCI. The workshops were documented through note-taking, including action points for edits and improvements. The notes were then later utilized as a base for editing all patterns. A subset of the edited patterns was circulated to workshop participants with the purpose to reach a consensus on the appropriateness of the edits. The workshops led to a better understanding of the template sections and led to a classification of the patterns in the following groups: 1) functionality of the portals; 2) *user interfaces* that give access to the functionality; 3) *user role*, i.e. the role of the user in the interaction with the portal; 4) *language technologies* that make the functionality possible. They also led to an initial ex-ante evaluation\(^3^8\) (Johannesson & Perjons 2014) of the content, and an improvement of the knowledge and the language embedded in the patterns. Patterns with related or similar content were merged together, and the total number of patterns was reduced from 20 to 14 (see figure 9).

The *summative evaluation* (Johannesson & Perjons 2014) was employed in order to get a final assessment of the patterns. The summative evaluation was performed with an online survey assessed by a non-randomized set of international domain experts in the areas of NLP, IR, HCI, and e-health. The online surveys utilized both close-ended question and open commentary fields. The domain experts were selected from experienced international professionals from both the industry and academia, established in research communities in different countries (Sweden, Finland and Australia).

The focus of this part was to evaluate two main features in the knowledge patterns (Rolland et al. 2000, Stirna et al. 2006): 1) the knowledge embedded in the patterns and 2) the language utilized to mediate the knowledge to the users. *Eight* hypotheses (corresponding to the artefact requirements in section 3.1) were specified and evaluated against given evaluation criteria and metrics (Rolland et al. 2000) as shown in table 2.

---

\(^3^8\) An ex-ante evaluation implies that an artefact is evaluated before its implementation or before it is fully developed.
Table 2. The features, criteria and hypotheses of the summative evaluation from Andreucci et al. 2018

<table>
<thead>
<tr>
<th>Feature 1: The knowledge embedded in the patterns</th>
<th>Criteria</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness: the degree to which the usage of the pattern would provide a substantial contribution in the context of a real problem-solving application</td>
<td>H1: The knowledge embedded in the pattern provides a substantial contribution for developers of mental health portals to solve an existing problem.</td>
<td></td>
</tr>
<tr>
<td>Relevance: the degree to which a pattern addresses a significant problem for health portals and the research areas of language technology and HCI</td>
<td>H2: The pattern addresses a problem that is significant and relevant for health portals and research within e-health, human–computer interaction and language technology</td>
<td></td>
</tr>
<tr>
<td>Adoptability: the degree of acceptance of the pattern to be used by domain experts for resolving a particular problem of interest</td>
<td>H3: The pattern is likely to be used in a real situation by designers of health portals, researchers within e-health, human computer interaction and language technology for resolving a particular problem of interest.</td>
<td></td>
</tr>
<tr>
<td>Coherence: the degree to which the pattern constitutes a coherent unit including correct relationships with other patterns</td>
<td>The pattern is coherent in the sense that it makes sense on its own and does not contain any contradiction. The hypothesis with questions about coherence of the embedded knowledge was split into two hypotheses: one covering H4: the coherence of the description of the problem and the other one covering H5: the coherence of the description of the proposed solution.</td>
<td></td>
</tr>
<tr>
<td>Consistency: the degree to which the pattern conforms to existing knowledge and vocabulary used in the fields of health portals/e-health, language technology and HCI</td>
<td>H6: The pattern fully conforms to the existing knowledge and vocabulary in e-health, health portals, Human Computer Interaction or language technology</td>
<td></td>
</tr>
<tr>
<td>Granularity: the level of detail at which the pattern addresses a problem</td>
<td>H7: The pattern provides a solution with a level of detail that adequately reflects the problem that is addressed.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature 2: The pattern language</th>
<th>Criteria</th>
<th>Hypothesis</th>
</tr>
</thead>
</table>
Comprehensiveness: the degree to which the language used in the pattern is precise in its description of the problem and the proposed solution

H8: The language used in the pattern is precise in its description of the problem and the proposed solution.

The questions in the survey evaluated each hypothesis against a given criterion. Five-level Lickert scales were utilized to grade each criterion, where N=1 was lowest and N=5 highest. Each evaluation produced evaluation tuples in the following form: <knowledge pattern name, feature, hypothesis, criterion, metric>, see table 3 for an example of the average result of an evaluation tuple.

Table 3. Results for the tuple <"Choice of QA approaches", "The knowledge embedded in the pattern", "The patterns is coherent in the sense that it makes sense on its own and does not contain any contradiction", "Coherence of the problem", "1-5">, where “1 to 5” is the coherence grading (from Andrenucci et al. 2018).

<table>
<thead>
<tr>
<th>Pattern name: Choice of question-answering (QA) approaches</th>
<th>Number of responses (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis: The patterns is <strong>coherent</strong> in the sense that it makes sense on its own and does not contain any contradiction</td>
<td></td>
</tr>
<tr>
<td>1 Incoherent</td>
<td>0 (0 %)</td>
</tr>
<tr>
<td>2</td>
<td>0 (0 %)</td>
</tr>
<tr>
<td>3</td>
<td>2 (20 %)</td>
</tr>
<tr>
<td>4</td>
<td>1 (10 %)</td>
</tr>
<tr>
<td>5 Coherent</td>
<td>7 (70 %)</td>
</tr>
<tr>
<td>I can’t answer</td>
<td>0 (0 %)</td>
</tr>
<tr>
<td>Average</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Lickert scales were utilized since they are standard in evaluations of knowledge patterns (e.g. Nilsson 2009, Rolland et al. 2000, Chung et al. 2004) and they are particularly useful in measuring users’ opinions (Williamson 2002, p.237). To consider a hypothesis verified, we used a threshold of 4, since in a five-level Lickert scale the value of 3 is considered to have more of a
neutral value, i.e. the third level can be more of a “no opinion/no preference” statement from the evaluator, while the fourth level is generally equal to agreement to a statement (Sullivan & Artino 2013). Scores between 3 and 4 were considered only partially to validate the hypotheses. The option to write open comments was also included, in order to allow the evaluators to submit qualitative feedback. Along with the quantitative analysis, a qualitative analysis of the comments was also performed. The comments were first thoroughly read and analysed, and then summarized and exemplified in conjunction with the quantitative results.

3.4 Limitations

The Knowledge integration of language technology approaches within CLIR (section 2.2) and QA systems (Article II) was carried out by one person only, namely the author of this thesis. The author could have misinterpreted or misunderstood the content, or some of the concepts in the reviewed articles.

No social and cultural parameters were considered in the user studies discussed in this research (Article I, Article III) - for instance, no particular consideration was given to what is typical for cultures and behaviour in Scandinavia, southern Europe or other parts of the continent. One reason for this approach is that the portal utilized in this research was developed as part of a project financed by the European Union, and one of the major goals of EU projects is to gather together the different European cultures, instead of trying to elicit their differences. Results based on the different professions of the participants were not considered either for the same reason.

The relevance assessment of the retrieved documents described in Article III and Article VI was based on subjective judgements of end users (Article III) and of assessors (Article VI). Other users or assessors may disagree about the relevance of the documents. The assessment of the knowledge patterns described in Article VII was also based on the subjective judgements of the assessors. Other assessors may disagree on the results.

Researchers were present during the study discussed in Article I and Article III. This might not be the typical search environment for health portal users who aim at searching for information anonymously, and the presence might have affected the results of the study. An alternative method for gathering quantitative data could have been an online survey, which could avoid any potential bias caused by the presence of the researcher. However, a survey could not offer the opportunity to directly ask follow-up questions related to the informants’ choices.
It is always desirable to have large groups of test users and assessors, but, unfortunately, time and resources tend to be common limitations in all research areas.

The sample groups discussed in Article I (22 individuals) and Article III (20 individuals) consisted of people of different ages and backgrounds, making them a reliable sample of the target users of health portals.

The evaluation of the alignments at the word level could have been done manually by language experts and not automatically with the help of reference data. However, such an approach does not offer the control and the flexibility that an approach based on reference data offers, which allows focusing in advance on certain word types or words within specific frequency ranges.

The assessors of Article VI are two IR-experts and one psychologist. They were part of the research group, but they evaluated the results blindly, i.e. the result lists were presented to the assessors without any information about the translation methods or the search engine that returned the results. It would have been interesting to involve a larger number of evaluators. Each evaluation session was time-consuming since it implied the assessment of 10 result lists per query (100 queries in total), and only highly motivated assessors interested in this research could provide quality feedback.

The results in Article VI are based on 100 randomly extracted user queries. A larger number of input queries might have provided more indications concerning the quality of the translations or the statistical results. CLIR and lexicon extraction is limited to Swedish as source language and English as target language. No further studies were performed.

81 potential respondents were contacted for the summative evaluation of the knowledge patterns (Article VII) and 25 responded, with an average response rate slightly over 30%. A larger number of evaluators would have provided more qualitative and quantitative data for the evaluation.

The summative evaluation could have been replaced or complemented by interviews or participatory (interactive) workshops, but this approach was not chosen in order to avoid logistical or locational constraints and to allow the evaluators to respond in their own time.

The texts in the EKP sections are not as concise as EKP patterns tend to be in software engineering books (e.g. Gamma et al. 1995). The utilization of extended text in the sections was a conscious choice made by the researcher since the patterns present a summarization of the results of the research foundations. The inclusion of more detailed information could give the reader/user
a better understanding of the content in the key sections of the patterns: the “problem”, the “rationale” and the “solution”.

3.5 Ethical issues

Our research adhered to the ethical research principles for design science as stated in Johannesson and Perjons 2014 (p. 190).

Principle 1 The Public interest: our studies aimed at finding implications of design and guidelines for the development of health portals for the general public. Health care is a common matter of interest for all the citizens so in our research we did not consider nor deal with private companies’ benefits.

Principle 2 Informed consent: the informants involved in our research were thoroughly informed about the content and the purpose of the studies. They were asked for permission to take notes. They gave their consent orally (user interaction studies, evaluation of query translation approaches) or replying to the evaluation survey (evaluation of knowledge patterns). They all took part in our research projects on a voluntary basis.

Principle 3 Privacy: all the questionnaires, interviews and user observation data were anonymous, i.e. no names or personal information linked to specific individuals could be deducted. The summative evaluation results were protected by password log-in.

Principle 4 Honesty and accuracy: the research was performed openly and honestly towards the informants and towards the research community in the scientific publications.

Before starting the portal interaction studies, the participants were informed about how the experiment was structured, and a visual demonstration of the GUI was provided. The participants were informed that they could interrupt the study at any time if they felt uncomfortable.

The query translation methods evaluators were informed by mail with details concerning the goal of the research and the evaluation process. The knowledge pattern evaluators were informed both by mail and orally (formative evaluation), or by mail (summative evaluation).

The researcher applied unbiased analysis, interpretation and evaluation of all the results.
Principle 5 Property: there was no need to sign an agreement about ownership of property rights since the corpus utilized is “open source”, i.e. it consists of the knowledge base of the web4health portal, a publicly available resource of information.

Principle 6 Quality of the Artefact: the data was collected in a test environment based on the knowledge base of web4health, i.e. a publicly available health portal implemented within the framework of an EU financed project. The IR systems utilized were Quick Ask (Sneiders 2002a) and Google Search. The query translation methods were based on MT and dictionary look-up translations. Two commonly available MT systems were employed for machine translation: Systran and Google Translate. Three dictionaries were employed for dictionary look-up translations 1) a public, generic online lexicon (Lexin); 2) a domain-specific lexicon developed with the help of scholarly recognized tools (Uplug, Tiedemann 2003); 3) a domain-specific lexicon based on the English words from the FAQ templates of web4health. The knowledge patterns produced were based on well-established templates (Bubenko et al. 2001, Persson et al. 2008) and contain the results of scientific publications in the field of e-health, information systems and HCI.
Chapter 4 Results and discussion

This chapter contains a discussion that answers the research questions described in chapter 1. The results of this research provide indications about:

- How information seekers experience different GUIs and user tailored search services.
- How information seekers search for mental health information on online portals.
- The QA approaches and the query translation methods that better fit health portals.
- The best practice methods for extraction of bilingual lexicons.

The chapter also contains a summary of the construction steps of the artefact that embodies the results of the foundations.

4.1 Research questions 1 and 2 (Article I, Article III)

4.1.1 Research question 1: how do information seekers experience interacting with a mental health portal?

Studies from Articles I and III showed that the two groups experienced the portal differently. Users undergoing psychotherapy saw the health portal of the study more as something or someone they could confide with, a friend, thus trying to reach a more personal contact in the interaction process. They also tended to submit queries related to their problems or personal lives. The group without health issues tended instead to search for general information, rather than seeking advice to solve own problems, and treated the portal like a medical encyclopaedia. The queries were more towards finding out facts about health topics rather than solving health issues.

Information with unsympathetic or non-empathetic texts, or texts with a high level of technical detail, was rejected more easily, particularly by the patients group. This group emphasized easy-to-comprehend popular scientific language, with an empathetic tone. The group of users that searched for general
information prioritized detailed information and factoid texts from experts. Information that confirmed previous knowledge was accepted more easily.

The patients’ group argued that the portal could not motivate them to make radical changes in their lives since it did not have “the eye-contact and the human warmth” that a real person could give. So the advice provided online did not seem to have the same impact as a therapy session with a professional therapist. This confirms the findings of similar research projects, where information seekers missed human contact with either peers (Zhang 2010) or therapists (Simpson 2004, Lal & Adair 2014).

Patients who had suffered for a long time and still had not solved their problems were more open to unexpected answers/texts. This spurred the reflective process, making them read about topics that in some cases gave new insights into their information needs. For instance one participant searching information about “tiredness” received texts about “depression” and “anxiety”, and another searched about “prevention of depression” and retrieved documents containing information about “exercise and workout”, thus discovering the connection between those topics.

Both groups appreciated the possibility to choose navigation paths among FAQs – indicating that navigation variety is the best option. Anonymity and confidentiality were considered important by both groups. In their opinion it allowed them to be more honest when it came to revealing personal details or confidential matters.

Natural language interfaces and structured menus as GUIs
Information in the portal can be searched with the help of two GUIs: structured menus and a panel for natural language (NL) input. The structured menus consist of a form with tree structured menus, where the user can select his/her topics of interest from a taxonomy corresponding to the categories of the knowledge base. The user can navigate from the generic categories into more specific subcategories, and related topics, by clicking further until s/he finds the FAQ of interest.

The NL panel consists of a standard panel as seen in global search engines, where users can submit information requests in natural language, i.e. the input does not have to be in a structured form or follow given input patterns. After receiving the input, the system returns a list of relevant entries ranked in descending order of relevance (the closer to the top, the more relevant). Each returned entry contains the FAQ heading, a short description of the content (preamble), as well as a link to the whole document.
Participants of both groups valued the possibility of navigating through the FAQs in several ways, either navigating back and forth from the retrieved answer headings/descriptions and their bodies (breadth-first navigation, Berendt & Brenstein 2001), or clicking on the links to related FAQ objects at the end of each text (depth-first navigation, Berendt & Brenstein 2001). The latter path was followed when users found very informative content in the retrieved document.

Users, in general, were also pleased to be able to choose between different GUIs since it allowed the user more flexibility in the search process.

The search panel allowed users to submit more personal and more intimate queries with the aid of NL input. This fact allowed users to vent their feelings and work off worries (“I like the fact that you can submit questions in natural language. You can write exactly what is on your mind”, “you can steer the dialogue better and ask what you exactly want”). This outcome was quite typical for users with mental health problems.

After submitting their input to the NL panel, several participants applied the following pattern: they first scanned the list with retrieved documents, reading the title and the description, and then selected the documents that seemed most relevant. Once the selection was made, the informants followed the link to the body of the chosen document. If the title or the description did not explicitly contain the words they submitted in their NL input, they found it difficult to consider an entry as relevant. This problem could have been avoided if the title or description of the document was annotated with the users’ query terms, explicitly coupling the information request to the content of the FAQ.

Uses who were more interested in generic information, or were unsure/not aware of which input to submit, appreciated the insight into the document collection provided by the menu-based interface (“...I prefer the menu-based interface, because I can see what is available in the database...”, “…if you do not know what you are really looking for, it is perfect with the menus, otherwise I would go for the NL interface…” as two users described it). This outcome was more evident for users without health problems, since they used the portal more like a medical encyclopaedia. It was also recurrent among the elderly seekers of both groups, which confirms the findings of other research projects (Stronge et al. 2006, Slone 2003), which showed that older adults prefer web links navigation in the search process rather than submitting search queries to search engines.
**Important properties on a mental health portal**

The informants were asked to select the attributes that according to their opinion characterize a high-quality health portal (see Article I and Article III for details).

Users from both groups considered that the information should be provided by **reliable medical sources**. This was considered important by nineteen out of twenty participants. Readily comprehensible text (seventeen users) and up-to-date content (sixteen users) were also considered as important features, i.e. it was a common choice for both groups.

The two groups prioritized differently when it came to the possibility to submit questions to a human expert. Eight participants undergoing psychotherapy considered it crucial, while only three participants from the healthy group shared the same opinion. Related research confirms this priority among health portal users (Sillence et al. 2006), in particular in crisis situations (Farrell et al. 2005, Koivunen et al. 2007).

Anonymity and confidentiality were also more important for patients than for the healthy group as well as the possibility of receiving information that was not influenced by business interests. The healthy group on the other side prioritized a high level of detail for the information provided.

### 4.1.2 Research question 2: which user profiling methods do psychological information seekers prefer?

Most of the participants (80%) preferred submitting information about themselves directly instead of letting the system infer their preferences and interests. Users were very sceptical towards letting a computer program infer so sensitive information by monitoring their information requests. The participants undergoing psychotherapy (the therapy group) were more influenced by emotional aspects and did not like to be monitored by a “big brother” system. They were afraid of losing control over the search process (“…I want to be able to choose my profile myself….I don’t like the fact that the system infers my profile behind my back…”, as one user described it). Users from the “healthy group”, i.e. users who did not have any contact with psychotherapy before, were more influenced by technical matters (“…the direct profile avoids typos and misspellings”, “…the direct profile approach is easier if you are aware of your own interests…”) and practical matters (“…it took me less time to create the profile”, “…I can quickly re-edit my profile…”).
It is interesting to underline that despite the scepticism towards the indirect profiling approach, 90% of the participants were satisfied with the accuracy of the inferred profile.

Users appreciated *the affordance*\(^{39}\) (Norman 1999) of the menu-based interface of the direct profiling approach since it provided an overview of the topics of the database, and eased their cognitive workload. As stated earlier it avoided misspellings and typos, but it did not provide the flexibility of the natural language input which allowed users to write “… exactly what is on their mind…”, as one user described it.

The user profile, in general, produced higher precision results for the top five documents\(^{40}\) and also had a higher precision @ 1, i.e. it managed to provide a higher fraction of relevant documents on the number one position in the document list\(^{41}\).

Users of both groups confirmed these results in the interviews, stating explicitly that according to their opinion it enhanced the quality of the retrieval results.

The user profile was useful in sorting results for generic queries, or non-exhaustive queries, i.e. queries with few keywords that did not give a complete picture of the user information needs. The information about the topics of interest and the search objectives improved the quality of the ranking and clarified the information of the queries. For example the query “I cannot sleep” gave higher ranks to documents about “ADHD\(^{42}\) and sleep problems” for users with the topic of interest “ADHD” in the profile, rather than generic documents about stress and insomnia.

Similarly, users who were interested in the causes of stress received higher ranks for documents about the factors that cause stress (e.g. family problems) rather than treatments of this problem.

\(^{39}\) What the design of a GUI intuitively allows users to do in the interaction process, for example pressing a button to submit a query

\(^{40}\) Precision at the top five documents for the healthy group: 58% with user profile and 51% without - for the therapy group: 47% with user profile and 44% without (from Andrenucci 2005, p. 63)

\(^{41}\) Healthy group p@1 results were 68% with user profile and 50% without, Therapy group: 65% with user profile and 63% without (from Andrenucci 2005, p. 63)

\(^{42}\) Attention Deficit Hyperactivity Disorder
4.2 Research question 3: which are the main approaches in QA and CLIR and which approaches fit online health portals (Article II and section 2.2)

Main approaches in QA

The question-answering paradigm was introduced in the late 1960s and early 70-s and simulated human intelligence trying to understand natural language input to complex, computationally expensive systems. QA-systems worked as natural language front-ends to databases (Woods et al. 1977), dialogue systems (Winograd 1972) and story comprehension systems (Lehnert 1977), and relied mainly on heavy NLP techniques that converted input text into formal representations of meaning such as logic and semantic networks. With the advent of the World Wide Web research has moved towards computationally lighter techniques that focus on extracting text chunks from large document sets (information extraction, Gaizauskas & Wilks 1998), retrieve documents that are relevant to users’ information needs (information retrieval, Salton 1971) or utilize templates that match different variations of similar questions to conceptual models of the database (e.g. Sneiders 2002a).

To our knowledge, there are not many literary reviews covering QA. The studies found do not explicitly discuss the health domain or online portals (e.g. Mollá & Vicedo 2007, Liu & Kang 2009, Androutsopoulos et al. 1995, Hirschman & Gaizauskas 2001). Related reviews generally divide the QA approaches into structured-based QA (i.e. with NL interfaces to databases, as described by Androutsopoulos et al. 1995) and free-text based QA (i.e. with NL interfaces to documents, as described by Zajac 2001) or between restricted domains and open domains (Mollá & Vicedo 2007). The review in this study is the first work that defines and covers three major QA approaches: 1) formal semantics-based (or deep) Natural Language Processing (NLP), i.e. text input is converted into formal representations of meaning such as logic and semantic networks; 2) Information Retrieval enhanced by shallow NLP, i.e. “lighter” NLP techniques that do not represent or analyse the “meaning” of the input; 3) template-based QA, i.e. where the understanding of the input is precompiled in a collection of manually created question templates.

In the review of the approaches, parameters such as the quality of the answers, portability and scalability issues, and the application domain, were considered.

Formal semantics-based NLP techniques are less portable and computationally heavier. A change of the application domain requires a chain of transformations that involves both the knowledge specific components and the language-specific components. However, they provide more accurate answers (as
shown for instance in the best performing QA system in several TREC conferences, Power Answer, Moldovan et al. 2004). The answers are also more reliable since they can be enhanced by logical proof (e.g. Athenikos et al. 2009, Harabagiu et al. 2000). Power Answer proved to achieve results around 83% accuracy in TREC 02 and 70% in TREC 03. Accuracy is defined as the fraction of correct answers. Mollá et al. (2003) achieved similar results (with a Mean Reciprocal Rank\(^{43}\) of 0.63) with higher precision and lower recall than standard IR approaches. A medical QA system that utilizes formal semantics-based NLP techniques is the ExtrAns system (Rinaldi et al. 2004). This system creates logic representations of user questions and the documents in the repository. The documents are analysed offline, and their semantic representations are stored in a database. User questions are converted into semantic representations online and are then compared to the representations of the documents in the matching process.

The IR approach is generally language and domain independent, which makes it easier to migrate to new domains or apply it to large domains such as the World Wide Web. It utilizes large text corpora and exploits statistical methods based on data redundancy\(^{44}\) in order to find and extract the answer from the texts. Mollá and Vicedo (2007) also confirm these conclusions, underlining that open domain QA rely on data redundancy and that data redundancy techniques are less appropriate for restricted domains in general, since fewer sentences contain the answer. Systems of this kind do not “comprehend” nor analyse the “meaning” of the user input or the texts, which implies that the answers provided cannot be supported or justified by logical proof, as in the case of the formal semantics-based NLP approach. The focus of this technique is more towards factoid questions (e.g. “when was the polio vaccination invented?”). IR systems that are not enhanced by NLP tend to provide lower precision but higher recall (e.g. the SMART system, Salton 1971, achieved a Mean Reciprocal Rank of 0.46).

New techniques implementing machine learning models based on neural networks have been applied successfully in question-answering, these techniques are called deep learning (Stroh & Mathur 2016). Deep learning utilizes both classical IR techniques (such as vector space representations) to represent queries and text passages/documents/answers. Syntactic analysis based on parse trees (e.g. Jousse et al. 2005) and semantic information are also utilized (Minaee & Liu 2017). Deep learning relies on neural networks in order to compare the representations of questions and candidate passages/answers, and compute their similarity (e.g. Severyn & Moschitti 2015) in order to predict

---

\(^{43}\) The rank \(k\) is the position of the first correct answer in the retrieved list. Its reciprocal rank is \(1/k\). Mean Reciprocal Rank is the mean counted on all answer lists.

\(^{44}\) A large number of similar text fragments/chunks containing the answer to user questions
Deep learning models generally need large amounts of training data to be effective (Stroh & Mathur 2016), for example with question-answer pairs where the relevant answers are marked. Systems of this kind achieved results equal or even higher than formal semantics-based NLP, i.e. with Mean Reciprocal Rank values higher than 70% (e.g. Tan et al. 2015).

The Template-based approach is a pattern matching approach where the semantic analysis of the user questions is done in advance, i.e. before implementing the system. Question templates have open concepts that are mapped to the conceptual model of the knowledge domain. The open concepts are instantiated with the terms contained in the user questions. Jacquemart and Zweigenbaum (2003) for example utilizes templates for classifying user questions in a QA system in the oral surgery domain. The templates are based on triples that contained two concepts and their relation (Concept A – Relation – Concept B), where the relation and the concepts are defined with the help of the UMLS semantic network. Several QA systems (e.g. Niu et al. 2003) also utilize the PICO-format (Sackett et al., 2000) with templates that classify natural language input instantiating the following concepts: problem (the problem of the patient), intervention (medication or therapeutic procedure), comparison (of the actual intervention to other possible interventions) and outcome (the effect of the intervention).

The template-based approach does not provide logical proof to the answers and is domain dependent, which causes portability difficulties. A fine-tuned system of this can achieve recall values above 80% (Sneiders 2002b).

Table 4 summarizes the three QA approaches from the application point of view. The techniques that are more suitable for QA in narrow domains (such as the health domain) are formal semantics-based NLP and template-based QA. They can handle more advanced questions (for instance “yes-no”-questions, “how to do”-questions, “when”-questions based on temporal time, i.e. “when should I take my medication?”45) and they can be implemented as interfaces to structured databases. Both techniques work well in counselling and advice matters. However, the formal semantics-based NLP approach may also implement argumentation techniques and discourse planning models, in order to make the systems more persuasive in counselling and advice-giving matters (e.g. Cawsey et al. 1999).

---

45 Factoid questions generally deal with absolute time/dates instead, e.g. “when was America discovered?”
Table 4. Comparison of the QA approaches from the application point of view, from (Andrenucci & Sneiders 2005)

<table>
<thead>
<tr>
<th></th>
<th>Formal semantics-based (Deep) NLP</th>
<th>IR &amp; shallow NLP</th>
<th>Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire Web</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Structured data</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Facts from text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Advice-giving</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Reliability % 46</td>
<td>Accuracy &gt; 70</td>
<td>Accuracy &gt; 56</td>
<td>Recall &gt; 80</td>
</tr>
<tr>
<td>Small domains</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The formal semantics-based (deep) NLP approach provides reliable answers with justification for the answers retrieved. These characteristics fit systems that support medical professionals in decision support situations that require evidence for taking critical decisions (Lin & Demner-Fushman, 2005).

Template-based QA fits health portals better (Andrenucci 2008): 1) it supports multilingual content and the retrieval of answers in different formats (e.g. images, audio-files and FAQ-objects), 2) it is easier to maintain and update, since every template is an individual piece of knowledge that can be updated independently, and 3) it solves word-sense disambiguation issues without computationally heavy calculations.

IR QA fits large open domains such as the entire web and facts extraction from texts.

**Main approaches in CLIR**

According to the research literature (Kishida 2005), translation methodologies within CLIR can be classified into three main approaches: 1) query translation, i.e. the user queries in the source language are translated into the target language, i.e. the language of the documents in the repository; 2) document translation, i.e. the documents are translated into the language of the user queries; 3) interlingua translation, i.e. both the user queries and the documents are converted into an intermediate language or semantic representation for matching purposes.

46 The accuracy results are from Voorhees & Dang (2003)

64
The query translation method is the most common translation approach since it is less computationally costly and easier to maintain (Kishida 2005, Rosemblat et al. 2003). There are three main approaches for translating queries from the source language to the target language (Gey et al. 2005, Kishida 2005): 1) bilingual dictionary search, i.e. word by word translations of queries by using bilingual dictionaries, b) machine translation (MT), i.e. user input is translated with software for machine translation, and c) parallel corpora search, i.e. translations of the source language are found in bilingual texts/documents aligned at the phrase or word level.

The trends in recent CLIR research show that researchers tend to combine those approaches and implement several techniques even within the same approach (e.g. Wu et al. 2008): this holistic view provides better translation and retrieval results overall. For instance, the best performing team at CLEF 2014 (SNUMedinfo, Shen et al. 2014) uses an MT system (Google translate) to translate the queries into different languages. The system also utilizes the UMLS metathesaurus - i.e. a multi-purpose and multi-lingual concept-based dictionary, with synonyms and information about the relationships between the concepts - for expanding the queries with related terms. One of the systems that used only standard IR techniques (Team Miracl, Ksentini 2014), such as the vector space model, achieved fairly low retrieval results instead. Research outside CLEF also confirms this trend. Volk et al. (2002) annotated texts in a German-English parallel corpus, and user queries with UMLS concepts and their semantic relations. They also created a similarity thesaurus that contained words from the corpus associated with a set of words with similar meaning, or from similar contexts. This combined approach enhanced precision both in CLIR and monolingual IR.

Wu et al. (2008) combine both parallel corpora and bilingual dictionaries. The researchers first utilize non-domain specific corpora in order to train the system, and then utilize domain-specific bilingual dictionaries and domain-specific monolingual corpora to improve the domain-specific translations. This approach based on combined resources improved the quality of the translations between Chinese and English and between English and French.

Tran et al. (2004) utilize a mixed translation approach for French-English CLIR in the biomedical field. They implemented a hybrid MT approach based on the general MT system Systran, enhanced with tailor-made heuristic rules for the biomedical domain and French-English translation patterns. They also implemented a parallel dictionary look-up based on UMLS-metathesaurus. Their results showed that their mixed MT-approach provided better retrieval

---

results than standard MT and the thesaurus-approach used separately. Furthermore, their results showed that the thesaurus-based look-up worked best with short queries (single term or two-word term queries), while MT-based performed better with long queries. Similar conclusions have also been presented of Oard & Diekema (1998). When it comes to the quality of the translations, Wu et al. (2016) showed that among MT systems, Google Translate is the system that produces state-of-the-art results, with translation quality that matches translations produced by human translators, in particular for the language pairs English-French, English-Spanish, English-German and English-Chinese. Our experiments (Andrenucci 2016) confirm these results even for Swedish-English.

For health applications that are not aimed at processing longer input queries a dictionary look-up approach is recommended - for instance, systems that have to support a keyword-based search of medical articles. Previous research (e.g. Pirkola 1998) has also shown that if longer phrases or multi-word units (MWUs) are not identified properly in dictionary-based look-up systems, then the phrase/MWU constituents are translated separately, and the sense/meaning of the whole phrase/MWU is somehow lost. Thus for systems that have to process longer user queries, such as QA systems or systems that process longer natural language inputs, the MT-approach is recommended.

4.3 Research question 4: which query translation methods work better for health portals and which are the best practice methods for extraction of bilingual lexicons? (Article IV, Article V, Article VI)

4.3.1 Query translation methods and health portals (Article VI)

As discussed in section 3.2.1, the following tools were utilized in the CLIR experiments: two MT systems (Systran and Google Translate), three dictionaries (the generic dictionary Lexin as lexicon baseline, the word alignment lexicon and the template-based lexicon) and two search engines (Google Site Search and Quick Ask).

Machine translation

Our experiments indicate that Google Translate achieved better retrieval results than Systran among MT approaches with both search engines (see table 5 for precision and MAP results, highest results in bold). The qualitative analysis of the translations showed that Google Translate managed word-sense
disambiguation better (see table 6 for translation examples), with translations that better fit the psychotherapeutic context (e.g. näringstillskott – “nutritional supplements” compared to Systran’s translation “industry boosts”, or vätskedrivande medel – “diuretics” vs Systran’s “liquids”).

Google Translate also provided linguistically more correct translations overall, even compared with the dictionary look-ups, in particular with compound words and MWUs. This outcome confirms our knowledge integration results and the results of earlier research: Wu et al. (2016) showed that among MT systems, Google Translate is the system that produces state-of-the-art results, with translation quality that matches translations produced by human translators. Liu & Cai (2015) also utilized Google Translate and Microsoft Bing as MT systems, comparing them to their corpus-based bilingual lexicon (Spanish - English). Their results also showed that Google Translate outperformed their system. Oard and Diekema (1998), Tran et al. (2004) showed that MT-based approaches performed better than dictionary look-ups with long queries. Google Translate utilizes Neural Machine Translation (NMT, Bahdanau et al. 2015), which has lately become the more popular paradigm in MT (Etchegoyhen et al. 2018).

**Dictionary look-up**

Word alignment provided better statistical results for both SWUs and MWUs (see table 5) and produced better compound word translations among dictionary look-ups (see table 6). Both the statistical results and the qualitative analysis in our study confirmed that domain-specific lexicons improved the quality of IR and the quality of the translations for both SWUs and MWUs, compared to the generic lexicon. Their implementation helped to find the translations that were more appropriate for the psychological context (e.g. medel – medicines vs funds, besvär – disorders vs complaints, ond – pain vs evil). Word alignment lexicons, in particular, proved to be very useful with compound words and MWUs, which are very common in the Swedish language and are also common in the medical domain (Rinaldi et al. 2004). So our conclusion is that it is advisable to implement domain-specific resources on top of generic lexicons.

The quality of the compound word translations was poor in the baseline lexicon, which affected the quality of the information retrieved with lower precision results.

---

48 The compound words were included in the SWUs
Table 5. Statistical results for Mean Average Precision (MAP) and Precision at a cut-off of 5 (P@5), from Andrenucci (2016), min-max values in parenthesis

<table>
<thead>
<tr>
<th>Translation Method/ Search Engine</th>
<th>MAP SWUs</th>
<th>MAP MWUs</th>
<th>P@5 SWUs</th>
<th>P@5 MWUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexin/ Quick Ask</td>
<td>0.15 (0.13-0.20)</td>
<td>0.27 (0.17-0.34)</td>
<td>0.34 (0.29-0.44)</td>
<td>0.34 (0.20-0.43)</td>
</tr>
<tr>
<td>Lexin/ Google Search</td>
<td>0.17 (0.14-0.20)</td>
<td>0.13 (0.10-0.17)</td>
<td>0.33 (0.29-0.20)</td>
<td>0.11 (0.04-0.16)</td>
</tr>
<tr>
<td>Word Alignment /Quick Ask</td>
<td>0.27 (0.23-0.32)</td>
<td>0.37 (0.30-0.43)</td>
<td>0.46 (0.37-0.57)</td>
<td>0.46 (0.36-0.55)</td>
</tr>
<tr>
<td>Word Alignment /Google Search</td>
<td>0.31 (0.26-0.38)</td>
<td>0.38 (0.32-0.43)</td>
<td>0.49 (0.28-0.68)</td>
<td>0.36 (0.25-0.44)</td>
</tr>
<tr>
<td>Template-Based /Quick Ask</td>
<td>0.18 (0.16-0.21)</td>
<td>0.36 (0.25-0.45)</td>
<td>0.36 (0.32-0.44)</td>
<td>0.44 (0.28-0.55)</td>
</tr>
<tr>
<td>Template-Based /Google Search</td>
<td>0.20 (0.18-0.23)</td>
<td>0.13 (0.07-0.18)</td>
<td>0.36 (0.29-0.44)</td>
<td>0.17 (0.09-0.24)</td>
</tr>
<tr>
<td>Systran /Quick Ask</td>
<td>0.15 (0.13-0.19)</td>
<td>0.28 (0.16-0.36)</td>
<td>0.35 (0.29-0.44)</td>
<td>0.37 (0.22-0.46)</td>
</tr>
<tr>
<td>Systran /Google Search</td>
<td>0.16 (0.14-0.20)</td>
<td>0.12 (0.06-0.17)</td>
<td>0.32 (0.29-0.37)</td>
<td>0.17 (0.08-0.24)</td>
</tr>
<tr>
<td>Google Translate /Quick Ask</td>
<td>0.23 (0.19-0.28)</td>
<td>0.39 (0.33-0.45)</td>
<td>0.45 (0.41-0.54)</td>
<td>0.49 (0.38-0.58)</td>
</tr>
<tr>
<td>Google Translate /Google Search</td>
<td>0.30 (0.26-0.34)</td>
<td>0.23 (0.12-0.32)</td>
<td>0.55 (0.51-0.64)</td>
<td>0.31 (0.18-0.40)</td>
</tr>
</tbody>
</table>

Machine translation vs dictionary look-up

Word alignment in some cases translated adjectives instead of verbs or nouns, but that did not affect the quality of IR since the meaning of the words/phrases was semantically correct (självskada – self-injurious vs Google’s more correct “self-harm”). The translation of conjunctions and prepositions was also worse than Google Translate (e.g. “consequences for gambling addiction” vs Google’s “consequences of gambling addiction”, “signs on eating disorders” vs “signs of eating disorders”, or “after to you stop smoking” vs Google’s “after you quit smoking”), but it did not affect the IR quality. Google Translate managed to provide highly qualitative translations and high IR results with SWUs and compound words as well (e.g. mättnadskänsla - satiety, sömnstörningar - sleep disorders).

Extracting a bilingual lexicon with word alignment requires time and access to bilingual resources. If a health portal does not have access to bilingual corpora, and it is not possible to invest time and financial resources on such a project, it is advisable to utilize Google Translate for Swedish to English CLIR.

When it comes to the search engines, Quick Ask gave better precision results with MWUs compared to Google Site Search. One explanation of this result is that Quick Ask implements domain-specific templates created to match NL questions in the field of psychology and psychotherapy. In other words, Quick Ask is more similar to a QA system than a search engine, and its templates were fit to match longer inputs/phrases in natural language.
<table>
<thead>
<tr>
<th>Query in source language</th>
<th>Lexin (baseline)</th>
<th>Word Alignment</th>
<th>Template-Based</th>
<th>Systran</th>
<th>Google Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td>näringstillskott</td>
<td>sustenance charge</td>
<td>nutritional supplements</td>
<td>Vitamin industry boosts</td>
<td>nutritional supplements</td>
<td></td>
</tr>
<tr>
<td>småhetsideal</td>
<td>narrow agitation ideal</td>
<td>slim ideals</td>
<td>slim treaty</td>
<td>narrow ideal</td>
<td>ideals of slimness</td>
</tr>
<tr>
<td>minskad minnesförmåga</td>
<td>decrease N/A</td>
<td>decreased memory</td>
<td>depressed mind</td>
<td>slimmed-down memory ability</td>
<td>reduced memory capacity</td>
</tr>
<tr>
<td>minnesproblematik</td>
<td>N/A</td>
<td>memory impairments</td>
<td>mind complex of problems</td>
<td>N/A</td>
<td>memory problems</td>
</tr>
<tr>
<td>sömnstörningar</td>
<td>sleep drive</td>
<td>Insomnia</td>
<td>sleep drive</td>
<td>torpor disturbances</td>
<td>sleep disorders</td>
</tr>
<tr>
<td>måttudskänsla</td>
<td>N/A</td>
<td>satisfaction</td>
<td>N/A</td>
<td>saturation feeling</td>
<td>Satiety</td>
</tr>
<tr>
<td>spelmisbruk</td>
<td>courtship abuse</td>
<td>pathologic gamblers</td>
<td>courtship abuse</td>
<td>playing misuses</td>
<td>gambling</td>
</tr>
<tr>
<td>skilsmässorådgivning</td>
<td>N/A</td>
<td>divorce advice</td>
<td>N/A</td>
<td>divorce advice</td>
<td>divorce advice</td>
</tr>
<tr>
<td>amfetaminmisbruk</td>
<td>N/A</td>
<td>drug abuse</td>
<td>amphetamines abuse</td>
<td>N/A</td>
<td>amphetamine abuse</td>
</tr>
<tr>
<td>vilka effekter har droger på människor</td>
<td>what effect be provided with drug at human being</td>
<td>what effects have drugs on people</td>
<td>what effects take drug of men</td>
<td>which effects have drugs on people</td>
<td>what effects drugs have on people</td>
</tr>
<tr>
<td>tecken på ätstörningar</td>
<td>character at eating disorder</td>
<td>signs on eating disorders</td>
<td>character of anorexia</td>
<td>sign indicating eating disorders</td>
<td>signs of eating disorders</td>
</tr>
<tr>
<td>antisocial personlighetsstörning</td>
<td>anti-publik personlighetsstörning</td>
<td>antisocial personality disorders</td>
<td>N/A</td>
<td>antisocial figure disturbance</td>
<td>antisocial personality disorder</td>
</tr>
<tr>
<td>konsekvenserna till spelberoende</td>
<td>consequences for gambling addictions</td>
<td>effect of courtship addiction</td>
<td>the consequences to playing depend</td>
<td>consequences of gambling addiction</td>
<td></td>
</tr>
<tr>
<td>överdrivet behov av släktingar</td>
<td>excessively necessity for relatives</td>
<td>excessive needs of relatives</td>
<td>excessively necessity of relations</td>
<td>excessive need for relatives</td>
<td>excess need for relatives</td>
</tr>
</tbody>
</table>
4.3.2 Best practice methods for extraction of bilingual lexicons: quality of word relations extracted from different versions of the corpora (Article IV, Article V)

Dictionary look-up is one of the query translation methodologies that can be utilized within CLIR. It implements word by word translations of queries by using bilingual dictionaries. The extraction of bilingual dictionaries from parallel corpora is a way to produce bilingual lexical resources (Kishida 2005) for query translation in CLIR. A bilingual dictionary can be extracted from bilingual texts with the help of word alignment (Tiedemann 2003), i.e. identifying translation relationships in bilingual texts (e.g. doctor -doctor, ästörningar- eating disorders) with the help of NLP tools.

Two versions of the corpora were utilized in our experiments. One version of consisted of as-is words (i.e. inflected words) and one version consisted of lemmatized words. Lemmatization is the process of morphological analysis of a word to infer the base form, i.e. the lemma of the word. Each version of the texts was also processed in the following ways: 1) part-of-speech (POS) tagged, i.e. the part of speech of the words were marked up, 2) part-of-speech tagged and parsed syntactically, i.e. the types of phrases (i.e. verb phrases, noun phrases) and their relation were tagged in the sentences of the texts.

Our experiments indicate that word inflections provided better statistical results than lemmas overall. The truncation applied in the lemmatization process resulted in worse alignments at the sentence and paragraph levels since the utilized tool (Uplug) applied a sentence alignment algorithm (Church 1993) that exploits the length of the sentences. This was quite evident in the follow-up study (Article VI), where we utilized the whole corpus of web4health in the alignment process, i.e. a much larger set of texts compared to the sample corpus that we utilized in the pilot study (Article IV). The removal of gender, number and definiteness through lemmatization caused a coarser POS tagging, which led to less accurate alignments at the word level. For example, removing the gender suffix -t in the adjective dåligt (bad) made it more difficult to find the nouns the adjective referred to (the so-called “ett-words” in Swedish), so the system aligned “uppförande – misbehavior” instead of the
correct alignment “dåligt uppförande - misbehavior”. This outcome was particularly evident for MWUs. Furthermore, wrong lemmatizations propagated errors to the POS tagging process which also led to worse alignments.

POS tagged and syntactically parsed words achieved better precision and recall than “as-is” inflected words, for both MWUs and SWUs with lower frequency rates (1-2 and 3-4). The information from the POS tagging and the syntactic analysis was useful in the alignment of words with dissimilar strings and low co-occurrence frequency, but sharing:

1) The same part of speech (e.g. two nouns matstrupe – oesophagus vs a noun and an adjective överkänslig – oesophagus),
2) Singular or plural number (e.g. barndomsupplevelser – childhood experiences\(^{49}\) vs barndomsupplevelser – childhood)
3) The same phrase type (e.g. verb phrases consisting of a verb and its particle tänka ut – decide VS tänka – decide, or verb phrases in passive forms uppfattas – are converted VS uppfattas - converted).

The morphological information also helped in linking correctly nouns that together built a conceptual unit (kostrådgivning – diet counselling VS kostrådgivning – counselling, aptitlöshet – loss of appetite VS aptitlöshet – appetite) and produced more fine-grained alignments (möta – cope, strategier – strategies VS strategier möta – strategies cope).

In higher frequency-rates (7-9, >=10), the statistical results of as-is word inflections were higher than POS tagged and statistically parsed word inflections. The lack of morphological information was compensated by the high co-occurrence coefficient values (the Dice Coefficient, Church et al. 1991) and the length similarity values, which provided enough clues for correct alignments. These results confirm partly the results of related research (e.g. Dalianis et al. 2009), where word inflections with higher frequency-rate were aligned with better precision. In our case this was clear for SWUs without POS and syntactic information, but not for MWUs.

Alignments of proper nouns with almost identical strings in both Swedish and English were generally correct in all frequency rates and for both lemmas and inflected words (e.g. Trifluoperazion - Trifluoperazione, Benzodiazepin - Benzodiazepine).

\(^{49}\) Barndom in Swedish means childhood, upplevelser means experiences
Version of a parallel corpus that is more suitable for extracting bilingual lexical resources in mental health

POS tagging and syntactic information improved the quality of the alignments for low frequency units (both SWUs and MWUs with frequency 1-4). Considering that the health domain often consists of multi-word terms with domain-specific meaning (Rinaldi et al. 2004), it is advisable to extract lexicons from word inflections that have been POS tagged and syntactically parsed, when the target is low frequency MWUs. When it comes to words with high frequency-rates (7-9, >=10), the morphological and syntactic clues are not necessary, since the statistical and the length-similarity clues provide enough information to extract correct alignments.

It is not advisable to lemmatize corpora before extracting a bilingual lexicon with the toolkit utilized in this research, or with similar techniques, since it caused problems at the sentence, paragraph and word level. This lowered the quality of the alignments. Lemmatization could be applied on the extracted lexicon afterwards, assembling words with the same base form in the source language or the target language, in order to produce lists with synonyms.

4.4 The artefact: knowledge patterns for online health portals development (Article VII)

Figure 8 shows the workflow of the construction and design of the knowledge patterns where the grey boxes are the outcome of each work phase. The three research foundations, the research questions and their results were synthesized and formalized first into twenty knowledge pattern drafts, originally classified in two large and coarse groups depending on their content: HCI patterns and Language Technologies-Information Retrieval patterns. During the formative evaluation (chapter 5), patterns with related or similar content were merged together or discarded. The total number of patterns was reduced from twenty to fourteen (see figure 9 and appendices). The knowledge patterns are the artefact produced in the Design Science Research Framework.
The formative evaluation also led to a new, more fine-grained classification of the patterns in the following groups 1) functionality of the portals; 2) user interfaces that give access to the functionality; 3) user role, i.e. the role of the user in the portal interaction; 4) language technologies that make the functionality possible. They also led to an initial ex-ante evaluation\(^{50}\) (Johannesson & Perjons 2014) of the content, and improvement of the knowledge and the language embedded in the patterns.

The information in the “problem” and “context/criteria” sections originate partly from the research questions, and partly from discoveries from the portal interaction studies, the knowledge integration and the language technology experiments. One example of the former is knowledge pattern 4, “Choice of direct or indirect user profiling”, which is directly related to research question 2 “Which user profiling methods do psychological information seekers prefer?”. One example of the latter is knowledge pattern 5, “Attention capturing answer headings”, where user studies of portal interaction indicated that participants found it difficult to see how relevant a retrieved document was, if its title or short description did not literally include the topics or keywords that they had explicitly submitted in their queries.

The rationale/solution parts consist of the synthesized results of the three research foundations.

\(^{50}\) An ex-ante evaluation implies that an artefact is evaluated before its implementation or before it is fully developed.
Chapter 5 Evaluation of the knowledge patterns

This chapter contains the results of the evaluation of the knowledge patterns, a comparison with related research and the lessons learned. The chapter ends with a discussion of the quality of the research.

5.1 Evaluation results

The evaluation activity regarding the knowledge patterns consisted of a formative evaluation and a summative evaluation.

The *formative evaluation* was employed to refine the material and gain unexpected insights (Driscoll et al. 2007) in the formative phase, before submitting the material to the summative evaluation. In this part, four semi-structured participatory workshops (Rolland et al. 2000, Chung et al. 2004) were organized involving three knowledge pattern experts and three domain experts in the field of language technology, e-health and HCI.

The feedback from the knowledge pattern experts was important to achieve a better understanding of the sections of the templates. The sections “problem”, “solution” and “consequences” were thoroughly reformulated, e.g., clarifying that “actors” have a problem and “actors” embrace both positive and negative “consequences” with the proposed “solution”. It also led to more effective and informative pattern naming, based on shorter noun phrases or verb phrases (see figure 9). The name of the pattern should focus on 1) the problem, 2) the activity in which the problem arises, or 3) the solution. The workshops also improved the classification of the patterns. The workshops led to a better understanding of the template sections and a classification of the patterns in the following groups 1) *functionality* of the portals; 2) *user interfaces* that give access to the functionality; 3) *user role*, i.e. the user role when interacting with the portal; 4) *language technologies* that make the functionality possible. The new categorization is an example of unexpected insight achieved during the formative phase.
The workshops with domain experts were important for merging together patterns with similar content and discarding redundant patterns. Thus the total number of patterns was reduced from 20 to 14 (see figure 9).
Figure 9. Result of formative evaluation, merging of 20 patterns to 14. The original pattern names and groups on left side, on the right side the pattern names and groups after the formative evaluation. The bracket shows a merge between two or more patterns, from Andrenucci et al. (2018).

<table>
<thead>
<tr>
<th>Original Patterns</th>
<th>Patterns after the formative workshop</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HCI patterns</strong></td>
<td><strong>Interface Patterns</strong></td>
</tr>
<tr>
<td>1) Search GUIs</td>
<td>Search and Navigation Paths (3)</td>
</tr>
<tr>
<td>2) Implement different navigation paths</td>
<td></td>
</tr>
<tr>
<td>3) Representation of retrieved information headings and links</td>
<td>Attention Capturing Answer Headings (8)</td>
</tr>
<tr>
<td>4) Implement exhaustive help for lay users</td>
<td>Exhaustive Help</td>
</tr>
<tr>
<td>5) Implement different layout metaphors</td>
<td></td>
</tr>
<tr>
<td>(was discarded since it was vague)</td>
<td></td>
</tr>
<tr>
<td>6) Direct of indirect user profiling</td>
<td>Choice of Director Indirect User Profiling (3)</td>
</tr>
<tr>
<td><strong>Functionality Patterns</strong></td>
<td></td>
</tr>
<tr>
<td>7) Style and content adaptation of the information</td>
<td>Empathic Text and Answers (4)</td>
</tr>
<tr>
<td>8) Functionality to ask mental health experts</td>
<td>Mental Health Expert Advisor (4)</td>
</tr>
<tr>
<td>9) Allow anonymous information access</td>
<td>Anonymous Information Access (4)</td>
</tr>
<tr>
<td>10) Utilize Reliable medical sources</td>
<td>Reliable Medical Sources (4)</td>
</tr>
<tr>
<td><strong>Information Retrieval Patterns</strong></td>
<td></td>
</tr>
<tr>
<td>11) Implement dynamic user profiling</td>
<td>Dynamic User Profiling (2)</td>
</tr>
<tr>
<td>12) Overcome N/A in fuzzy queries with user profiling</td>
<td>Vague Queries (2)</td>
</tr>
<tr>
<td><strong>Language Technologies Patterns</strong></td>
<td></td>
</tr>
<tr>
<td>13) Search engine</td>
<td>Choice of Search Engines (2)</td>
</tr>
<tr>
<td>14) Implement native languages support</td>
<td></td>
</tr>
<tr>
<td>15) Query translation approaches</td>
<td>Choice of QT Approaches (1)</td>
</tr>
<tr>
<td>16) Dictionary look-up techniques</td>
<td></td>
</tr>
<tr>
<td>17) Domain specific lexicons in mental health</td>
<td></td>
</tr>
<tr>
<td>18) Annotation and processing of mental health texts</td>
<td>Bilingual Text Processing (1)</td>
</tr>
<tr>
<td>19) Lemmatization of words in mental health texts</td>
<td></td>
</tr>
<tr>
<td>20) Question Answering approaches</td>
<td>Choice of QA approaches (1)</td>
</tr>
</tbody>
</table>

(1) evaluated by NLP experts, (2) evaluated by IR experts, (3) evaluated by HCI experts, (4) evaluated by e-health experts.
The feedback was very valuable and helpful in order to improve and update the knowledge in the sections covering the rationale and the solution. The domain experts also suggested enhancing, when possible, the solution sections with example screenshots in order to better illustrate the solution. This feedback was more frequent for user-interface patterns. The vocabulary was also updated with more up-to-date terms, and suggestions for related research were given.

The **summative evaluation** was employed in order to get a final assessment of the patterns. The summative evaluation was performed with an online survey assessed by a non-randomized set of international domain experts in the areas of NLP, IR, HCI and e-health. The evaluators received by email a file with the patterns to evaluate (these were also available through web links on the survey page) and a description of the pattern notation/format. They were also informed about the background and goal of this research. A web link to the survey published on Stockholm University’s tool “Survey and Report” was also included. 81 potential respondents were contacted for the summative evaluation of the knowledge patterns, and 25 responded, with an average response rate of slightly over 30%. Figure 10 below shows the average response values per hypothesis/criteria, figure 11 the average values per hypothesis/criteria in each group of knowledge patterns and figure 12 the average values for each knowledge pattern and each pattern group.

![Figure 10. Total average values per criteria/hypothesis (min-max averages in parenthesis), from Andrenucci et al. (2018)](image-url)
Figure 11. Average values per hypothesis/criteria for each group of knowledge patterns (min-max in parenthesis), from Andrenucci et al. (2018).
The results of the summative evaluation indicate that the following results apply for all pattern groups (average score above 4) for the following criteria: relevance (H2), coherence of the problem (H4) and comprehensiveness (H8). This indicates that for all pattern groups the following hypotheses are verified: H2) the patterns address a problem that is significant and relevant for the knowledge domain; H4) the description of the problem is coherent and without contradictions; H8) the language used in the pattern is precise in its description of the problem and the proposed solution.

Usefulness (H1), and coherence of the solution (H5) achieved average scores above 4 for all pattern groups except for NLP patterns: IR patterns (4.47 for H1, 4.37 for H5), interface patterns (4.40 for H1, 4.43 for H5) and functionality patterns (4.65 for H1, 4.30 for H5). Adoptability (H3) achieved average scores above 4 for IR patterns (4.13) and Interface patterns (4.13) while both NLP and functionality patterns achieved results below 4 (3.60 for NLP, 3.86 for functionality). The results indicate that the hypotheses below are verified for all pattern groups but only partially verified for NLP patterns: H1) the knowledge embedded in the pattern provides a substantial contribution for developers of mental health portals to resolve an existing problem; H5) the

---

**Figure 12.** Average values for each knowledge pattern and for each pattern group (min-max averages in parenthesis), from Andrenucci et al. (2018)

<table>
<thead>
<tr>
<th>Knowledge Pattern</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Language Processing</td>
<td>3.85</td>
</tr>
<tr>
<td>&quot;Bilingual Text Processing&quot;</td>
<td>3.84 (3.44-4.22)</td>
</tr>
<tr>
<td>&quot;Choice of Query TR Approaches&quot;</td>
<td>3.81 (3.13-4.00)</td>
</tr>
<tr>
<td>&quot;Choice of QA Approaches&quot;</td>
<td>3.89 (3.22-4.00)</td>
</tr>
<tr>
<td>Information Retrieval</td>
<td>4.47</td>
</tr>
<tr>
<td>&quot;Dynamic User Profiling&quot;</td>
<td>4.37 (3.8-4.8)</td>
</tr>
<tr>
<td>&quot;Vague Queries&quot;</td>
<td>4.58 (3.8-4.8)</td>
</tr>
<tr>
<td>&quot;Choice of Search Engines&quot;</td>
<td>4.45 (4.5)</td>
</tr>
<tr>
<td>User Role and Interface</td>
<td>4.27</td>
</tr>
<tr>
<td>&quot;Choice of Direct or Indirect User Profiling&quot;</td>
<td>4.06 (3.6-4.5)</td>
</tr>
<tr>
<td>&quot;Search and Navigation Paths&quot;</td>
<td>4.36 (3.6-4.5)</td>
</tr>
<tr>
<td>&quot;Attention Capturing Answer Headings&quot;</td>
<td>4.37 (3.5-4.8)</td>
</tr>
<tr>
<td>Functionality</td>
<td>4.38</td>
</tr>
<tr>
<td>&quot;Empathetic Text and Answers&quot;</td>
<td>4.30 (3.6-5)</td>
</tr>
<tr>
<td>&quot;Reliable Medical Sources&quot;</td>
<td>4.47 (4.2-4.8)</td>
</tr>
<tr>
<td>&quot;Anonymous Informat. Access&quot;</td>
<td>4.40 (4.2-5)</td>
</tr>
<tr>
<td>&quot;Mental Health Expert Advisor&quot;</td>
<td>4.18 (3.6-4.8)</td>
</tr>
</tbody>
</table>
pattern fully conforms to the existing knowledge and vocabulary in e-health, health portals, HCI or language technology. The results also indicate that the following hypothesis is verified for IR patterns and Interface patterns but only partially verified for NLP patterns and functionality patterns: H3) the pattern is likely to be used in a real situation by designers of health portals and experts in the field of e-health, HCI and language technology.

Consistency (H6) achieved average scores above 4 for IR patterns (4.63) and functionality patterns (4.37) but below 4 for NLP patterns (3.41) and interface patterns (3.90). These results indicate that the hypothesis is verified for the IR patterns and functionality patterns and partially verified for the other groups: H6) the pattern conforms to the existing knowledge and vocabulary in e-health, health portals, HCI or language technology. Granularity (H7) was also fully verified for two pattern groups, interface patterns (4.26) and functionality patterns (4.33), but only partially verified for NLP patterns (3.89) and IR patterns (3.86) (H7: the pattern provides a solution with a level of detail that adequately reflects the problem that is addressed).

All patterns scored an average score above 3 and 10 patterns out of 13 achieved average values above 4 (see figure 12). The lowest average scores were achieved by the Natural Language Processing (NLP) patterns (3.85 as overall average score), where 5 out of 8 criteria achieved an average score below 4 (see figure 11): usefulness (3.77 for H1), adoptability (3.60 for H3), coherence with the solution (3.64 for H5), consistency (3.41 for H6) and granularity (3.89 for H7). The open comments provided by the evaluators confirmed the statistical results since they underlined that the level of technicality was too high in some of the patterns. The technical terms made the knowledge difficult to understand without further explanations. Some solutions were not considered fully explanatory on their own either. Some evaluators also suggested supplementing the solution sections with alternative solutions, or expanding them with alternatives designed for people with physical disabilities, such as visual impairments. These results indicate that this is the pattern group that needs further development.

Information Retrieval (IR) Patterns achieved high average scores for the following criteria: relevance (4.85 for H2), coherence of the problem (4.80 for H4), consistency (4.63 for H6) and comprehensiveness (4.66 for H8). The results for granularity were a little bit lower (3.86 for H7). Some comments from the evaluators gave valuable information for further improvement of the patterns, for example, to provide a specific guide about how to choose the most suitable search engine based on cost/benefit instead of comparing the retrieval quality of different search engines. A deeper level of detail was suggested in the solution part for some patterns (e.g. in the pattern “Vague queries”), since the solution was considered quite abstract.
Interface Patterns achieved their highest average scores for usefulness (4.40 for H1), relevance (4.50 for H2), coherence of the problem (4.53 for H4) and coherence of the solution (4.43 for H5). All average scores were above or equal to 4 except consistency (3.90 for H6). Even in this case, the open comments of the evaluators gave valuable feedback to understand the results. Two evaluators suggested including the knowledge of the patterns of Tidwell (Tidwell 2010) in the pattern “Search and navigation paths”. Tidwell’s knowledge patterns are specific for HCI and the knowledge in the pattern could benefit from it. Another interesting feedback was that the knowledge in the same pattern could be generalized to other domains than mental health. Other interesting suggestions were made regarding the utilizations of more precise terms in the solution section of the pattern “Reliable medical sources”, which advises the utilization of content delivered by reliable medical sources. The word “utilize” was not considered clear enough and the meaning of “utilization” should be further explained in the text.

Functionality Patterns achieved their highest average scores for usefulness (4.65 for H1) and relevance (4.75 for H2) and all average scores were above 4, with the exception of adoptability (3.86 for H3). The open comments helped the researcher to discover some possible improvements concerning adoptability. Two evaluators suggested considering financing issues in the pattern “Mental health expert advisor”, which advises developers to implement a functionality where users may consult human experts. The comments suggested investigating how much this kind of service would cost and whether it could be financed in part by the state where the portal is implemented. Another evaluator raised the need to describe in detail how to organize the financing of the portal. These were two points that had not been taken into consideration when developing the patterns. Another interesting comment concerned adoptability issues related to anonymity/confidentiality in the patterns that recommend “Anonymous information access” and a “Mental health expert advisor”. An evaluator raised the issue that for example in Sweden a mental health expert’s advice is subject to control by different governmental authorities such as the National Board of Health and Welfare or the Swedish Psychological Association. This issue may cause obstacles for portal users that need the content of their written questions to be kept confidential.

The pattern “Empathetic text and answers” advises the adaptation of the retrieved text and answers to the information seekers background, focusing on empathetic tone and less technical terms for people suffering from mental distress. One evaluator suggested being more specific about where to draw the line between empathetic and technical texts, in particular when it comes to text quality and text composition.
5.2 Comparison with related research

The quantitative results of the summative evaluation are comparable with the results of Nilsson (2009) and Rolland et al. (2000). Nilsson (2009) produced a set of user interface (UI) design patterns specially adapted for mobile applications. The goal of the knowledge patterns was to provide indications for developing more user-friendly applications on mobile devices. For the evaluation, Nilsson utilized a 6-point scale questionnaire and focused on two criteria (relevance and usefulness), while this research considers eight criteria. Forty-eight participants evaluated the patterns and produced average scores around 5.0. Rolland et al. (2000) developed and evaluated knowledge patterns that focused on modelling business processes and changes in organizations. This research has utilized the same evaluation approach and the same format for the knowledge patterns (Enterprise Knowledge Patterns (EKP), Bubenko et al. 2001, Persson et al. 2008). While this research focused on assessing the content and the language of the knowledge patterns, Rolland et al. (2000) assessed the patterns’ content, language and the methodology for developing the patterns. Twenty-one evaluation criteria were utilized, and evaluation data were collected through a questionnaire that was compiled during interactive workshops. The average scores per hypothesis were around 4.0 for most patterns, which is comparable to our results.

The quantitative results of this research are a little bit higher than the results achieved by Chung et al. (2004), who produced patterns for ubiquitous computing and evaluated them with iterative and interactive workshops. Three pattern groups were iteratively improved after each workshop, and the summative results scored average scores of 3.1, 3.6 and 3.8 on a five-point scale.

Kohler & Kerkow (2008) developed knowledge patterns in the field of workflow modelling. Their evaluation criteria were usefulness, understandability (i.e. the criterion “comprehensiveness” of this research), consistency and applicability (i.e. “adoptability”), but their evaluation results cannot be directly compared to this research, since the data collection was based on collecting “yes/no” answers to evaluation questions and did not utilize point scales. The evaluation produced positive results for 8 out 9 patterns.
5.3 Lessons learned

In summary, the lessons learned in the knowledge evaluation phase are:

- The formative part of the evaluation through the support of expertise in the fields of knowledge patterns, HCI, language technology and e-health gave an important contribution to the improvement of the content and the vocabulary of the knowledge patterns. The feedback from the knowledge pattern experts was important to achieve a better understanding of the sections of the templates, a more effective and informative name-giving to the patterns and their classification. The workshops with domain experts were important for merging together patterns with similar content and discarding redundant patterns. Their feedback was very helpful to improve and update the knowledge in the sections covering the rationale and the solution. The domain experts also suggested enhancing, when possible, the solution sections with example screenshots in order to better illustrate the solution.

- The summative evaluation indicates that overall the knowledge patterns address a problem that is relevant to the knowledge domain and the knowledge embedded is useful to solve the problems. The description of the problem and the solution were considered coherent and without contradictions in most patterns. The language used in the patterns was precise in its description of the problems and the proposed solutions.

- NLP patterns need some further work in the solution parts to be finalized. This is left as future work.

- The comments of the evaluators gave indications for further improvement and future directions of the knowledge patterns: 1) the addition of more details and graphic examples in solutions that were considered vague. The better mental health developers understand the proposed solution, the higher is the probability that they can develop interfaces and search facilities that improve the information search experience and the relevance of the retrieved results. This improvement may eventually lead also to a better health condition of the information seekers; 2) the addition to the proposed solutions of more alternatives that also consider financing issues and cost-benefit outcomes; 3) mental health experts’ advice in Sweden is subject to revision of governmental authorities such as National Board of Health, which may cause conflicts in cases where patients require that the content of their questions not be revealed to “outsiders”. This issue could also be relevant for countries with similar regulations; 4) some patterns were considered to be adoptable even in other health domains than mental health, which increases their utility range.
5.4 Research quality

The knowledge patterns have been evaluated with a sequential mixed method based on both formative and summative evaluation. Semi-structured participatory workshops with knowledge pattern experts and specific domains experts were utilized in the formative evaluation. Online surveys with both close questions and open-ended comments were assessed by a non-randomized set of international domain experts.

There is no clear consensus in the research community on what can be considered an acceptable response rate. Some authors consider 20% an acceptable response rate (e.g. Visser et al. 1996), while others accept only 50% as the minimum size (Babbie 1990). Holbrook et al. (2003) showed that lower response rates generally imply lower demographic representativeness, but in this research, the participants’ expertise and domain knowledge were more important than demographic representativeness. Visser et al. (1996) showed that surveys with lower response rate (e.g. near 20%) had more accurate feedback and more dedicated answers by the respondents, compared with surveys with higher response rates (e.g. 60% or higher). So we consider the response rate of this research to be acceptable.

Our respondents were all internationally experienced professionals from both industry and academia, and the higher the respondents are in an organizational hierarchy, the more difficult it is to involve them in answering online surveys (Anseel et al. 2010). The sample size is subjected to time and budgetary constraints (Isaac & Michael 1995) and the acceptable sample size is always an arbitrary judgement (Hill 1998). So even in this case there is no clear consensus in the research community. For some researchers (e.g. Isaac & Michael 1995) smaller respondent groups are a better fit for research with complex or extensive qualitative information. In this case, the risk of missing relevant information is lower compared to using larger sample groups (Hill 1998). There is no doubt that a larger number of evaluators would provide more relevant material for the summative evaluation, however, the respondents that answered our survey belonged to a non-randomized set of domain experts that more likely were motivated and engaged in this research. Their qualitative feedback in the text comments was vital to understanding the quantitative results and to identify deficiencies and future improvements of the patterns. Hence we consider that the small sample size did not affect the validity of the results.

The knowledge patterns synthesize the results of the three foundations which might be limited by the techniques and tools utilized. Technology may evolve and change over time and so would some of the results if new techniques or tools are applied. Even if technology may evolve, there is a component in the
knowledge that will endure: the diversity of people. People with different backgrounds and health conditions experience health portal access and information search differently. It is thus important to adapt health portals to the diverse characteristics of the information seekers. We managed to include this knowledge in the patterns thanks to the sample groups consisting of people of different ages, backgrounds and health conditions.
Chapter 6 Conclusions

This chapter summarizes the thesis and its theoretical and practical contributions. The chapter ends with recommendations for future work.

6.1 Summary and contributions

The main goal of this thesis was to design, develop and evaluate an artefact for the development and implementation of online health portals aimed at information seekers without medical expertise. The artefact consists of a set of fourteen knowledge patterns. To the best of our knowledge, these are the first knowledge patterns that explicitly cover health portals in the field of psychology/psychotherapy and that cover three research foundations:

1) User studies of portal interaction, i.e. interviews and observations about how users experience health information online and personalized search.
2) Knowledge integration of language technology approaches.
3) Experiments with language technology applications, in the field of cross-language information retrieval/question-answering.

The knowledge patterns synthesize the results of research studies and experiments which resulted in scientific publications. The results were used as input to create the knowledge patterns. The knowledge patterns were then evaluated with a sequential mixed method based on both formative and summative evaluation.

The evaluation results of the knowledge patterns were positive and encouraging. All patterns were considered relevant and coherent in their description of the problem. The language used in the patterns was considered precise in the description of the problem and its solution. Three out of four pattern groups were considered a substantial contribution for the area and the patterns were also considered adoptable to a large extent.
6.1.1 Theoretical contributions

This research has provided important directions for further advances in the online health portal area and has contributed to the further development of the disciplines involved in the research foundations.

**User studies of portal interaction and contributions to e-health, HCI and UM.** The research has provided information on how users without medical expertise interact with health portals in the field of psychology/psychotherapy and how the mediated information is accepted by the information seekers. The findings indicate that users with mental health problems experience health portals and information search differently from information seekers without health problems. For example, information with unsympathetic or non-empathetic tone is rejected more easily by people suffering from mental distress. The research has also investigated the efficacy of adapting the information delivery to health portal users’ characteristics and information needs. The findings indicate that encapsulating information about the users in user profiles enhances the information retrieval quality.

People suffering from mental distress have lower cognitive workload capacity (Tollenaar et al. 2008, Trammell et al. 2014). Our findings indicate that this category of users tends to omit detailed information in their queries and that user profiles may help them find relevant documents/answers that they would have missed otherwise, reducing their cognitive workload. Our findings also indicate that health information seekers prefer submitting their profile directly and are sceptical about letting the portal infer their preferences and interests. The scepticism of users with mental health problems is driven by emotional aspects and the scepticism of users without mental health problems is more influenced by technical matters.

**Knowledge integration of language technology approaches and contributions to e-health, QA and CLIR.** This research has described the main approaches in QA and CLIR, defining the context of application in the health domain that better fit the approaches. Among QA approaches, formal semantics-based NLP provides reliable answers with justification for the answers retrieved. This indicates that it fits systems that support medical professionals in decision support situations. The findings of our research also indicate that template-based QA fits health portals better: 1) it supports multilingual content and the retrieval of answers in different formats (e.g. images, audio-files and FAQ-objects); 2) it is easier to maintain and update since every template is a self-contained piece of knowledge that can be updated individually; 3) it solves word sense disambiguation issues without computationally heavy calculations. Among CLIR approaches, our findings indicate that the dictionary
look-up approach is the better choice for applications that have to process queries consisting of SWUs, for example systems that have to support a keyword-based search of medical articles. Since MWUs might not be appropriately identified in dictionary-based look-up, the MT-approach is recommended for systems that have to process MWUs or longer queries in natural language.

**Experiments with language technology applications and contributions to e-health, CLIR and NLP**: The main contribution of the experiments is the analysis of query translation methods from Swedish to English and how they affect the retrieval results of health information. The findings indicate that the domain-specific dictionaries improve the quality of IR and the quality of the translations for both SWUs and MWUs, compared to the generic lexicon. Their implementation helped to find the translations that were more appropriate for the psychological context, so the usage of such domain-specific lexicons is recommended. The dictionary extracted with word alignment provided better statistical results both with SWUs and MWUs among the dictionary look-ups utilized. The quality of the translations of MWUs and compound words was poor in the baseline lexicon, which affected the IR quality with lower precision results. Our experiments also indicate that Google Translate achieved better retrieval among MT approaches and managed word-sense disambiguation better, producing translations that better fit the psychological/psychotherapeutic context. Google Translate also provided linguistically more correct translations overall, in particular with compound words and MWUs, which are very common in the Swedish language and in the health domain. Hence it is advisable to utilize Google Translate (or similar techniques) for Swedish to English CLIR in this domain.

The experiments performed in this research also provide indications about how the pre-processing of parallel corpora affect the quality of the extracted word alignments performed with NLP tools. Our findings indicate that corpora consisting of word inflections provide better statistical results than lemmas overall. POS tagging and syntactic information improve the quality of the extracted word alignments, in particular for low frequency units (both SWUs and MWUs with frequencies 1-2 or 3-4). So if the texts consist of low frequency words or if the targets of the extraction are low frequency words, it is advisable to utilize corpora with inflected words, annotated with POS tagging and syntactic parsing. Our results indicate that the morphological and syntactic information does not improve the quality of alignments of units with higher frequencies (7-9, >=10). Our findings also indicate that alignments of proper nouns with almost identical strings are generally correct in all frequency-rates, and for both lemmas and inflected words (e.g, *Trifluoperazion* - *Trifluoperazione*).
6.1.2 Practical contributions and societal consequences

Health care is moving from a traditional doctor-patient interaction at medical arrangements towards e-health services, in order to stimulate patient empowerment and engagement (Eklund 2014, Ricciardi et al. 2013), to help people suffering from physical and/or mental impairments (Farrell et al. 2004, Zeng & Parmanto 2004), or to support people living in rural areas (Christensen & Hickie 2010a). This trend is particularly evident in mental health, where people tend more and more to seek help online rather than through traditional arrangements (Webb et al. 2008). Online health portals are considered essential tools for providing e-health services and for improving mental health among citizens; their importance is expected to grow as the availability of the Internet widens worldwide. The evaluation results of the knowledge patterns indicate that the knowledge patterns can be used as instruments for developing online mental health portals and that they can provide further directions for the development of this area. Developing health portals with multilingual support and search services tailored to the background of information seekers has the potential of reaching a larger group of citizens globally, including minorities. Helping more people to access relevant health information will hopefully lead to a reduction of the “medical divide”\(^51\) (Eklund 2014) among health consumers, reducing inequalities in health care and society as a whole. It will hopefully also help health portal users, developers and medical experts to understand the complexity of building mental health portals. This research is a step forward towards the future direction of modern health care, where citizens utilize online e-health resources for control and management of their own health care, including prevention, self-diagnosis and treatment (Eriksson 2012).

6.2 Future work

This thesis has focused on the design, development and evaluation of knowledge patterns for online health portals development. A natural step forward for the research would be a demonstration with real-life cases, i.e. utilizing the knowledge patterns as guidelines in the development of mental health portals, or even portals in other health domains, since several patterns could potentially be applied in other domains than mental health. A study with a real-case scenario would likely provide more feedback concerning their usefulness, relevance and adoptability and also add further improvements to this work.

\(^{51}\) Modern health care requires health consumers to be more active in the health care process. However, only consumers with access to medical information and know-how are able to question or demand treatments, which causes inequalities in health care.
The evaluation described in Article VII identified some deficiencies in the patterns:

- Further work is suggested to finalize the NLP patterns, in particular in the solution sections.
- The evaluators suggested the inclusion of interesting topics that were not discussed in the patterns - for instance - the financing of portal services such as human expert consultations. It would be interesting to investigate further whether these services could be subsidized by the local authorities of different countries. Patterns with financing guidelines could add more value to our pattern set and would help mental health portal developers in their planning work.
- It would be useful to discuss and define more explicitly the differences between empathetic and technical text, in terms of text quality and text composition.
- Another further development step could be to extend the knowledge patterns with results of language technology experiments performed on other source languages than Swedish, for instance, other Germanic languages such as Danish, German or Dutch.

In our experiments with language technology applications within CLIR, we did not utilize external resources such as medical ontologies and metathesauri as a complement to the query translation methods utilized. It would be interesting to utilize these resources together with bilingual lexicons and MT software, in order to assess whether they enhance the quality of the translations or the precision results. Furthermore, it would be interesting to compare the translations of Google Translate with other MT software, such as Microsoft Bing, and measure the information retrieval results with other search engines such as Yahoo and Microsoft Bing Search.

The portal interaction studies were performed with informants utilizing personal computers. It would be of interest to expand the studies to smartphones or tablets, which nowadays are standard devices in everyday life.
References


Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS quarterly, 67-93*


Appendices

Knowledge Pattern 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Choice of question-answering (QA) approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Language technology Patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>Knowledge integration of language technology approaches</td>
</tr>
</tbody>
</table>

**Problem**

Choosing the most appropriate automated question-answering (QA) approach to implement on a mental health portal is not an easy task for portal designers. A wrong choice may burden the budget, affect the quality of the answers and not lead information seekers to relevant information.

**Criteria and Context**

If portal designers want to implement QA functionality on a mental health portal, they will have to make decisions about which QA approach to choose. Developers need to choose the approach that is most suitable for the domain depending on the following parameters: types of questions supported, size of the domain, domain and language portability, and type of data processed (structured data such as relational databases vs unstructured data such as text).

**Rationale and Solution**

Finding the right QA approach to implement is not an easy task for portal designers, since there are several parameters to consider. In our research (Andrenucci and Sneiders 2005) we have integrated knowledge from previous research and defined three QA approaches, considering parameters such as type of questions supported, quality...
of the answers, portability, size of the domain and type of data processed.

In our research we have defined three QA approaches,
1) formal-semantics Natural Language Processing (NLP): text input is converted into formal representations of meaning such as predicate logic and semantic networks
2) Information Retrieval (IR) enhanced by shallow/surface NLP, i.e. “lighter” NLP techniques that do not represent or analyse the “meaning” of the input
3) Template-based QA: the understanding of the input is precompiled in a collection of manually created question templates.

Template-based QA is the most viable approach commercially (Andrenucci and Sneiders 2005), (Sneiders 2009) and it is the approach that fits health portals better: It fits advice giving/counselling, it supports multilingual content and answers in different formats (FAQs, multimedia files, static text or data from a database). Its knowledge is easy to maintain and update ( unlike formal-semantics NLP), since every template is a self-contained piece of knowledge.
It also fits restricted specific domains ( unlike IR) such as mental health and the accuracy of the answers is higher than IR (Dalianis et al. 2010).

<table>
<thead>
<tr>
<th>Forces and Influences</th>
<th>There is a trade-off between the quality of the answers and developmental costs based on knowledge, time and resources.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences</td>
<td>Implementing template-based QA will help portal designers to save time and resources, and will provide the best service for the users.</td>
</tr>
</tbody>
</table>
Implementing template-based QA in larger open domains or transferring it to new domains will require more time and resources than IR with shallow NLP.

### Papers


### References


<table>
<thead>
<tr>
<th>Name</th>
<th>Choice of query translation approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Language technology Patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>Knowledge integration of language technology approaches</td>
</tr>
<tr>
<td>Problem</td>
<td>Choosing the most appropriate approach for translating queries on a mental health portal is not an easy task for portal designers. A wrong choice may burden the budget, affect the quality of the answers and not lead information seekers to relevant information.</td>
</tr>
<tr>
<td>Criteria/Context</td>
<td>Health portals aimed for an international audience may not contain documents/information in the native language of the information seekers. When information seekers submit queries/questions in their own native language and the documents/answers retrieved by a search engine are in different languages, we have cross-language information retrieval (CLIR). If health portal designers want to implement CLIR, they will have to make decisions about which CLIR-approaches to implement, based on the time and resources available, query translation and retrieval quality, the types of queries supported (e.g. one word queries or multi-word queries) and available expertise in language technology.</td>
</tr>
<tr>
<td>Rationale and Solution</td>
<td>According to current research there are three main approaches where query translation is performed: Machine Translation, Dictionary look-up and parallel corpora (Kishida 2005, Gey et al. 2005). Query translation and retrieval quality</td>
</tr>
</tbody>
</table>
In our studies, we found that Google Translate achieved better retrieval results among **Machine Translation (MT) approaches**, as well as more linguistically correct translations overall. The domain-specific lexicon based on word alignment (i.e. extracted from parallel corpora) achieved better retrieval results among **dictionary look-ups** and its retrieval results were comparable to Google Translate. The lexicon extracted with word alignment proved to be very helpful with **compound words**, so this approach is to recommend among **domain-specific lexicons**, considering that compound words are **very common** in the Swedish language and in the health domain.

**Types of queries supported**
Research studies (Tran et al 2004, Oard 1998) showed that dictionary-based look-up works better with **short queries (single term queries)**, while MT tends to provide better retrieval results for **longer queries (two terms or more)**. Our research has partly confirmed these results. However, Google Translate managed to produce high retrieval results even for single term queries consisting of compound words.

If the target of a portal is information retrieval based on **short queries (single term queries)** then dictionary look-up, enhanced with domain-specific lexicons, should be implemented. If the target is to process **multi-term phrases** than MT should be implemented.

**Time and resources available/expertise in language technology**
Extracting a lexicon from parallel corpora is a time and resource consuming process. It is advisable to utilize Google Translate from Swedish to English CLIR as Machine Translation software if a *quick and qualitative solution* is necessary. Unless it is possible to invest time and resources on a bilingual lexicon, i.e. an enduring linguistic resource useful even for future work. Extracting word alignment lexicons from parallel corpora also requires language technology knowledge of how to apply word alignment tools.

<table>
<thead>
<tr>
<th>Forces and Influences</th>
</tr>
</thead>
<tbody>
<tr>
<td>The word alignment lexicon was extracted from texts written by medical experts. The texts were written for readers without medical expertise.</td>
</tr>
<tr>
<td>Access to domain-specific Swedish-English corpora is needed in order to extract word alignment lexicons.</td>
</tr>
<tr>
<td>The user queries utilized as input data in our studies were written by users without medical expertise. Studies with user queries submitted by medical professionals should be conducted in order to assess the quality of the approaches with clinical queries.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementing the most suitable technique for the application context/needs will save time and resources, and will provide the best service for the users.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
</tbody>
</table>
Knowledge Pattern 3

<table>
<thead>
<tr>
<th>Name</th>
<th>Dynamic user profiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Language technology patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
</tbody>
</table>

**Problem**
User profiles enhance the quality of information retrieval (Hawkins et al. 2008, Vasilyeva et al. 2005). Portal designers will have to make decisions about how to implement a user profile that facilitates search services for information seekers and makes the portal credible.

**Criteria/Context**
Mental health information seekers have different interests, knowledge skills and search goals (Borzekowski et al. 2009), but portal technologies are rather limited and generally do not implement personalized search facilities (Moon & Burstein 2005). In order to fill this gap portal designers need to tailor the retrieval of information to users’ interests and search goals with the help of user profiles.

**Forces and Influences**
Users may be unwilling to reveal personal information such as interests or search goals for privacy reasons.

Encapsulating sensitive information requires higher knowledge of information security for encrypting data.

**Rationale and Solution**
Previous research (Hawkins et al. 2008, Vasilyeva et al. 2005) and our user studies (Andrenucci 2006, Andrenucci & Forsell 2004) have shown that the “one size fits all” approach among search and navigation facilities on health portals need to be changed. Encapsulating information about the users in user profiles enhances the information retrieval quality (Gauch et al. 2007, Andrenucci 2006) and the retrieved information is better accepted by users (Mittal et al 94). User profiles also reduce the cognitive workload of information seekers (DiMarco et al 2005, Andrenucci 111)
Users with mental health problems in particular see a portal as someone who “cares”, trying to reach a more personal contact with it (Andrenucci 2005, Andrenucci & Forsell 2004). Those categories of users also tend to omit detailed information in their queries (Andrenucci 2006, Andrenucci & Forsell 2004).

We advise the implementation of user profiles that encapsulate user categories of interests, search goals, health condition and their knowledge of the domain. We advise to encapsulate at least three categories of interest matching the categories in the knowledge base, and the search goals based on a medical taxonomy based on five well-known terms (Bhavnani et al. 2003, Bader & Theofanos 2003): Terminology, Risks/Effects, Prevention, Diagnosis, Treatment and Prognosis plus one extra term: Causes, since it is recurrent search goal among search queries that we have studied on a mental health portal (Andrenucci 2005).

Interests may change and users may submit requests outside the range of their original topics of interests. The profile of a user should dynamically evolve in the retrieval processes.

**Consequences**

Better service for health information seekers and more satisfied consumers. Retrieval of information that is better fit for the seekers search goals and backgrounds. Better credibility of the portal.

**Papers**

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borzekowski D. et al. 2009 Use of the Internet and other media for health information among clinic outpatients with serious mental illness. Psychiatric Services 2009 60:9, 1265-1268</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Vasilyeva E., Pechenizkiy M. &amp; Puuronen S. Towards the Framework of Adaptive User Interfaces for eHealth Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS’05)</td>
</tr>
</tbody>
</table>
**Knowledge Pattern 4**

<table>
<thead>
<tr>
<th>Name</th>
<th>Choice of direct or indirect user profiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>User role patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
</tbody>
</table>

**Problem**


If portal designers decide to utilize user profiling, they will need to choose **the approach** that is **better accepted** by the information seekers, or it might affect negatively the quality of the information retrieved and the credibility of the portal.

**Criteria/Context**

Mental health information seekers have different interests, knowledge skills and search goals (Borzekowski et al. 2009).

Since health portals’ technologies are rather limited and generally do not allow personalized search facilities (Moon & Burstein 2005), portal designers need to tailor the retrieval of information to users’ interests and search goals with the help of user profiles.

**Rationale and Solution**

User profiles can be created explicitly or implicitly (Kass & Finin, 1998).

Explicit knowledge is acquired letting the user **directly** state choices, preferences that can be immediately used in the profile, while **implicit knowledge is inferred indirectly** by observing the interaction between the user and the system.

Our studies (Andrenucci 2006) have shown that in general information seekers prefer stating their choices and preferences **explicitly**. They are sceptical towards letting a system monitoring their behaviour and inferring their interests.

The reasons are both emotional (e.g. fear of losing control and being monitored) and technical (avoidance of misunderstandings).
We recommend to implement *explicit* profiling, i.e. the approach where users choose directly their categories of interests and search goals, since this is the choice that is best accepted by users. The creation of the profile is not subjected to false assumptions or misunderstandings, unlike profiling approaches based on monitoring the input or the behaviour of the users.

**Forces and Influences**

Some users who do not know which kind of information they need to search - or cannot explicitly state their topics of interests/search goals - may prefer letting the system infer this information.

**Consequences**

Better service for information seekers and possibly better help for people suffering from mental diseases. Better credibility of the portal.

It is important that the indirect profiling approach is transparent, i.e. it informs how the profile is being inferred, so that users can re-edit it if they feel that it is wrong, or the portal will lose credibility and the users will be confused.

Users can find the direct approach tedious and time consuming which might hinder them from providing all necessary information (Waern 2004). This might affect the quality of the information retrieved.

**Papers**


## References


Vasilyeva E., Pechenizkiy M. & Puuronen S. Towards the Framework of Adaptive User Interfaces for eHealth Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS’05)

## Knowledge Pattern 5

<table>
<thead>
<tr>
<th>Name</th>
<th>Attention capturing answer headings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Interface patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
</tbody>
</table>

### Problem
Information seekers might miss relevant documents/answers when they scan the list with answer headings. This hinders users from finding and assimilating information that could solve their problems.

### Criteria and Context
Users who submit queries to a search engine on a mental health portal adopt the following pattern in the information retrieval process (Andrenucci 2006):

1. They first submit the question/query,
2. They scan the list with answer headings, containing the title and a short description (preamble) of the answer/document
3. They select the answers/documents whose headings seem most relevant.

A drawback of this selection process is that users may sort out relevant answers/documents if the list with the answer headings is not easily interpretable for information seekers.

### Rationale and Solution
Our user studies of portal interaction (Andrenucci 2006, Andrenucci and Forsell 2004) showed that participants found it difficult to see how relevant an answer was when its title or short description did not explicitly mention the terms submitted in the query. For instance, an informant, interested in preventing depression, retrieved, among other entries, an entry entitled “Exercise and weight regulation”, whose content described the advantages of a more active life, e.g. prevention of mental problems. S/he could not see the coupling between the query and the relevance of the answer (“…Strange I asked about depression problems and I got information about exercise…”).

Adapting the text in the headings (preamble),explicitly coupling the content of the retrieved document to the user’s information need or search interests, would help users to access relevant information that might solve their health problems.

**How?**

Answers or document headings (preambles) should *explicitly mention the users’ query terms, search goals and topics of interests* so that they explicitly grab the users’ attention. The adaptation should be applied also for the links pointing to documents with related information.
The picture below gives an example of the results for the query “depression prevention” from web4health.info.

And the picture below is an example of text adaptation that could be done just including the query terms submitted by the user.
### Forces and influences
The short term memory capacity and focusing capacity differ among users. People suffering from mental distress have lower cognitive workload capacity (Tollenaar et al. 2008, Trammell et al. 2014). Users with low short term memory capacity may benefit from words that recall their information needs (query terms, search goals and topics of interests) and reduce their cognitive workload.

### Consequences
Information seekers suffering from mental distress - or people with low cognitive workload capacity in general - may find relevant documents/answers that they would have missed otherwise. This might help to improve their health condition.

For users who have in-depth domain knowledge, this type of adaptation may be unnecessary. Domain-experienced users may be able to select answers/documents which do not explicitly mention the topics in the queries. They can more easily assess the relevance of the information provided.

### Papers


### References

### Knowledge Pattern 6

<table>
<thead>
<tr>
<th>Name</th>
<th>Empathetic text and answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Functionality patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>Information with unsympathetic/non-empathetic attitude and higher grade of technicality is rejected more easily by people suffering from mental distress (Andrenucci &amp; Forsell 2004, Andrenucci 2006, De Carolis 1996). This might hinder users from finding and assimilating information that could solve their problems.</td>
</tr>
<tr>
<td><strong>Criteria and Context</strong></td>
<td>Users suffering from mental health problems and users without mental health problems have different search goals, are interested in different levels of technical details and appreciate different tones/styles in the text (Andrenucci &amp; Forsell, 2004, Andrenucci 2006). A mental health portal should adapt the content to these different types of information seekers.</td>
</tr>
<tr>
<td><strong>Rationale and Solution</strong></td>
<td>In order to deliver and adapt the texts/answers for different information seekers, it is necessary to encapsulate users’ search goals and mental condition in a user profile. Information about the mental condition should clarify if the information seekers suffer from mental health problems or not. For information seekers who are suffering from mental distress and are interested in finding a solution to their problem, the texts should be written with an empathetic tone and with a lesser degree of technicality, since details are less important for this category of users.</td>
</tr>
<tr>
<td><strong>Forces and Influences</strong></td>
<td>Information seekers suffering from mental health problems are interested in empathetic texts written in an empathetic tone, but other users may prioritize detailed and technical texts.</td>
</tr>
<tr>
<td><strong>Consequences</strong></td>
<td>Providing texts with an empathetic tone and a lesser degree of technicality may help people suffering from mental distress to find and assimilate information that might improve their health condition. Texts with an empathetic tone and a lesser degree of technicality might be uninteresting for information seekers with in-depth knowledge or users without health problems, due to their priority for medical content with a higher level of details.</td>
</tr>
</tbody>
</table>
Papers


References
### Knowledge Pattern 7

<table>
<thead>
<tr>
<th>Name</th>
<th>Mental health expert advisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Functionality patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health consumers suffering from diseases value the possibility to contact a human expert advisor when accessing a health portal. If this feature is not available, it might affect the credibility of the portal.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria/Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information seekers who search for information in order to solve their problems appreciate the possibility to receive individual counselling on a mental health portal (Andrenucci 2006).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rationale and Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our studies and interviews have shown that different user groups have different priorities when it comes to the functionalities of online health portals. People suffering from mental health problems highly value the possibility to submit questions to a human expert. Studies of health portal users (e.g. Silence et al. 2006) confirm this priority, in particular in crisis situations (Farrell et al. 2004, Koivunen et al. 2007).</td>
</tr>
</tbody>
</table>

If a mental health portal not only has an educational role but also aims at helping patients or people suffering from mental health problems, it should offer the possibility to consult a human expert that provides answers to individual problems.

If possible, implement “ask a health expert” functionality, where psychology and psychotherapy experts provide individual counselling.
<table>
<thead>
<tr>
<th><strong>Forces and Influences</strong></th>
<th>Requires the active involvement of a psychologist/psychotherapist that answers the questions submitted by the users.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consequences</strong></td>
<td>Better service for information seekers and possible help for people suffering from mental diseases.</td>
</tr>
</tbody>
</table>
Knowledge Pattern 8

<table>
<thead>
<tr>
<th>Name</th>
<th>Anonymous information access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Functionality patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
<tr>
<td>Problem</td>
<td>Information seekers do not likely access sensitive information or reveal personal information if anonymity or confidentiality is not preserved. This might hinder solution of health problems.</td>
</tr>
<tr>
<td>Context/Criteria</td>
<td>Some diseases in mental health are often stigmatized. Information seekers who search for information in order to solve their mental health problems might be afraid of stigmatization.</td>
</tr>
<tr>
<td>Rationale and Solution</td>
<td>Our user studies showed that anonymity and confidentiality were considered an important factor when disclosing personal information (Andrenucci 2005), in particular for information seekers suffering from mental health problems. Studies of health portal users (e.g. Christensen &amp; Hickie 2010) and medical guidelines (HON Code of Conduct for medical and health Web sites) confirm this priority. We advise to allow anonymous information submission and anonymous information access, i.e. identities, IP-numbers and other parameters that could lead to identification should not be visible for users, web masters or psychotherapists/psychologists accessing the portal. We also advise confidential handling of the information queries submitted by the users.</td>
</tr>
<tr>
<td>Forces and Influences</td>
<td>Anonymous information access requires technical solutions so that sensitive information is not openly accessible.</td>
</tr>
<tr>
<td></td>
<td>Anonymity could be a drawback in cases of offensive, defamatory or racist content.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Consequences</strong></td>
<td>Users are prepared to reveal more information about their personal problems and issues, thus facilitating the information retrieval process. This information can be encapsulated in the user profile and will hopefully help people suffering from mental diseases to improve their health condition.</td>
</tr>
</tbody>
</table>
## Knowledge Pattern 9

<table>
<thead>
<tr>
<th>Name</th>
<th>Search and navigation paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Interface patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
<tr>
<td>Problem</td>
<td>Information seekers search and navigate information in different ways (Brusilovsky 2007). If navigation and search variety is not offered on mental health portals, some users might be hindered from finding and assimilating information that could solve their problems.</td>
</tr>
<tr>
<td>Criteria/Context</td>
<td>When interacting with information portals, users implement several browsing and searching patterns, depending on their age, search goals, type of questions/queries and health condition.</td>
</tr>
</tbody>
</table>

### Rationale and Solution

**Search Paths: Menu-based interfaces and Natural Language (NL) search panels**

In our studies, we have utilized both menus-based search interfaces (see figure 1) and NL search panels (see figure 2). In the menu-based interface, users utilized drop-down menus to choose from a list of specific diseases, to select up to three categories of interest from the knowledge domain, and to pick the goal with their search. Through the NL search panel, users submitted queries/questions in natural language, without clicking and selecting pre-defined keywords.
Menu-based interfaces and age, search goals, types of requests, health condition

Related research studies (e.g. Stronge et al. 2006) and our studies (Andrenucci 2005) have shown that senior information seekers prefer menu-based search interfaces. The menu-based interfaces were considered less flexible than Natural Language (NL) based search panels, but they reduced the cognitive workload of users since they allowed choosing from a pre-defined set of topics. Menu-based interfaces fit information requests that could be summarized to relatively short (1-3 words) generic questions/queries, based on keywords matching the categories in the knowledge base (Andrenucci 2005). These types of requests were particular common among information seekers who did not undergo therapy and considered themselves as “healthy”.

It also fit users who needed an overview of the topics available on the portal (“I like the menu better. You see directly what is available”).
NL search panels and search goals, types of request, health condition

NL search panels were considered more flexible and fit users with more advanced requests ("submitting questions in NL was the better alternative, I could specify my needs better") or users who utilized the panel to give vent for their feelings, work off worries or needed to be more honest with intimate matters, i.e. users suffering from mental diseases that utilized the search panel as counselling tool ("with NL you can steer the dialogue with the system better and ask what you exactly want", “I like the fact that you can submit questions in NL, you can write exactly what is on your mind”).

Solution

We advise implementing combined menu-based interfaces and NL search panels in order to support users with different search preferences. One possibility is to provide 1) menu-based search interfaces as the recommended search tool for senior information seekers or “healthy” information seekers interested in generic information and 2) NL search panels as recommended search tool for users suffering from mental health problems in need of counselling and expressing their feelings. Information concerning user health condition, age and search goals can be encapsulated in the user profile.

Navigation paths through the portal

Our user studies showed that users appreciate the possibility to choose between different navigation paths through the documents (Andrenucci 2005, Andrenucci & Forsell 2004): either jumping back and forth between the list with answer headings and the text bodies (breadth-first navigation, figure 3) or following the links to related entries at the bottom of each (depth-first navigation, figure 4). The second path was chosen when a highly relevant text was found. We did not notice any particular difference between healthy users and users undergoing psychotherapy concerning navigation preferences.

Solution

We advise the implementation of both breadth-first navigation and depth-first navigation in order to offer several navigation paths to the information seekers.
Figure 3. Breadth-first navigation, i.e. navigating back and forth from document headings and their text bodies.
Forces and Influences

Implementing several search and navigation paths might require larger setup work for interaction design.

Users who do not know which kind of information they require - or cannot explicitly state their topics of interests/search goals in natural language - may prefer utilizing the menu-based interface.

Consequences

Variety of navigation paths leads to better search possibilities for information seekers, which leads to better acceptance of the portal among users.

Papers


<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>
## Knowledge Pattern 10

<table>
<thead>
<tr>
<th>Name</th>
<th>Reliable medical sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Functionality patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User Studies of portal interaction</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>There is a huge amount of medical information on the World Wide Web. It is important for portal designers to publish health information that information seekers can rely on (Sillence et al. 2006), or it will affect the credibility of the portal.</td>
</tr>
<tr>
<td><strong>Criteria/Context</strong></td>
<td>Health information seekers and patients are increasingly being involved in their health care (Eklund 2014). They are expected to gain empowerment through the acquisition of medical knowledge on the web. This raises the expectations on the quality of the information available on health portals.</td>
</tr>
<tr>
<td><strong>Rationale and Solution</strong></td>
<td>Related research studies (e.g. Sillence et al. 2006) and our studies (Andrenucci &amp; Forsell 2004, Andrenucci 2006) have shown that information seekers prioritize health information originating from well established, reliable domain experts. The HON Code of Conduct for medical and health websites confirms this priority.</td>
</tr>
<tr>
<td></td>
<td>Utilize the knowledge of (or content produced by) professional domain experts as the knowledge base of health portals.</td>
</tr>
<tr>
<td><strong>Forces and Influences</strong></td>
<td>Requires the active involvement of psychologists/psychotherapists, or at least the usage of material written by people with professional medical knowledge. This could lead to additional costs.</td>
</tr>
<tr>
<td><strong>Consequences</strong></td>
<td>Better credibility of the portal and better acceptance of the information among information seekers. Possibly better help for people suffering from mental health diseases.</td>
</tr>
<tr>
<td></td>
<td>HON Code</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.hon.ch/HONcode/Conduct.html">http://www.hon.ch/HONcode/Conduct.html</a></td>
</tr>
<tr>
<td><strong>Knowledge Pattern 11</strong></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Name</strong></td>
<td>Choice of search engines</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td>Language technology patterns</td>
</tr>
<tr>
<td><strong>Founded on</strong></td>
<td>Experiments with language technology applications</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>Deciding which search engine to utilize on a portal for mental health is an important step for designers. A wrong choice may 1) burden the time plan and the budget 2) affect the information retrieval quality.</td>
</tr>
<tr>
<td><strong>Criteria/Context</strong></td>
<td>If a search engine is implemented on a health portal, portal designers will have to make decisions about which search engine to implement. It is necessary to choose the search engine that is most suitable for the mental health domain and the types of information processed.</td>
</tr>
<tr>
<td><strong>Rationale and Solution</strong></td>
<td>In our experiments (Andrenucci 2016) we utilized two search engines: Google Site Search and Quick Ask. Google Site Search(^2) is a custom search engine for specific websites. It is more portable and easier to implement, since knowledge engineering is not required. Quick Ask (Sneider 2002) is a search engine developed at Stockholm University, department of Computer and Systems Sciences. It requires time and human resources for set-up work. The set-up work is based on knowledge engineering of the domain, where patterns covering different variations of frequently asked queries are compiled. In our experiments (Andrenucci 2016) we found that Quick-Ask provided <em>higher precision</em> with Multi-Word Units.</td>
</tr>
</tbody>
</table>

(i.e. phrases with at least two words) than Google Site Search. Quick Ask fits MWUs and natural language since it resembles more like a question-answering system than a search engine. Considering that multi-word units are very common in the health domain, it is advisable to implement Quick Ask as the search engine on mental health portals.

However an assessment of the costs needed to tailor and set-up Quick-Ask for the portal should be done, before choosing this solution. If there is not enough time or resources, then quicker solutions should be considered.

Google-Site Search worked better with Single-Word Units SWUs (i.e. one-word phrases), so portals supporting single keywords search should implement this search engine, or if a solution based on a quicker implementation is needed.

### Forces and Influences

Our studies utilized only two search engines: Quick Ask and Google Site Search. Further research should be conducted even with other search engines.

There is a trade-off between the quality of the retrieved answers/documents and developmental and implemental costs of search engines or question-answering systems (Sneiders 2010).

### Consequences

Implementing the most suitable technique for the needs and budget will help portal designers to save time and resources.

### Papers

References


### Knowledge Pattern 12

<table>
<thead>
<tr>
<th>Name</th>
<th>Vague queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Language technology patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User Studies of portal interaction</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>Information seekers that submit too vague or too abstract queries cannot find relevant documents/answers. If the queries are not completed with further information, users might be hindered from finding and assimilating information that could solve their problems.</td>
</tr>
<tr>
<td><strong>Criteria/Context</strong></td>
<td>People suffering from mental distress have lower cognitive workload capacity (Tollenaar et al. 2008, Trammell et al. 2014), which affects the quality of the queries they submit. Sometimes the submitted queries are vague, and the mental health patients utilize the search panel only to express their feelings (Andrenucci &amp; Forsell 2004, Andrenucci 2005).</td>
</tr>
<tr>
<td><strong>Rationale and Solution</strong></td>
<td>Mental health information seekers submit search queries with different levels of specificity. Our studies (Andrenucci &amp; Forsell 2004, Andrenucci 2005) have shown that users (generally patients undergoing psychotherapy) tend in some cases to submit questions to express their feelings and the query content is so abstract (e.g. “I don’t know what to do”, “I feel bad”) that no matching documents are retrieved. We advise to complete fuzzy user queries or user queries without matching documents with the categories of interests and search goals, i.e. the information encapsulated in the user profile. Below is an example of such a solution. The first picture shows the results for the queries “There is nothing I can do” – “no answers found”.</td>
</tr>
</tbody>
</table>

![Example of solution](image-url)
The second picture shows the results for the same query enhanced with the topic “depression” and the search goal “treatment” from the user profile. Two documents were found thanks to the addition of this information.

<table>
<thead>
<tr>
<th>Forces and Influences</th>
<th>Information seekers who do not know which kind of information they require - or cannot explicitly state their topics of interests/search goals in natural language (NL) - may prefer utilizing menu-based search interfaces instead of NL search panels.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences</td>
<td>Allowing the system to complete fuzzy user queries with the categories and search goals that are encapsulated in the user profile reduces the cognitive workload of distressed information seekers and helps them retrieving information that might solve their problems.</td>
</tr>
<tr>
<td></td>
<td>Users have to reveal personal information about their search interests and search goals that will be encapsulated in a user profile. If the information included in the user profile is not sufficiently informative or is not correctly/honestly compiled, it may lead to retrieval of irrelevant information or in some cases no retrieval at all.</td>
</tr>
</tbody>
</table>
Knowledge Pattern 13

<table>
<thead>
<tr>
<th>Name</th>
<th>Exhaustive help</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Interface patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>User studies of portal interaction</td>
</tr>
</tbody>
</table>

**Problem**
Information seekers encounter difficulties when using menu-based interfaces for submitting their user profiles or when searching for information (Andrenucci 2006). This may hinder them from retrieving relevant information that could solve their problems.

**Context/Criteria**
Health information seekers and patients are being involved more and more in health care (Eklund 2014). They are expected to gain empowerment through the acquisition of medical knowledge on the web. Information seekers utilize available search interfaces in the process of gathering knowledge.

**Rationale and Solution**
The structure of the menu-based search interfaces tend to be quite inflexible (Stone et al. 2005) and allows quite short and static queries (Andrenucci 2005). Menus are effective since they offer clues for user recognition rather than forcing users to recall words from memory (Stone D. et al. 2005). Menus are also useful when users need support in their decision making process. However these advantages are not applicable if the users do not understand how the interfaces work or do not understand their content (Stone et al. 2005).

Our menu-based interfaces consisted of drop-down menus based on a structured taxonomy (Andrenucci 2005), where generic domain categories were hierarchically divided into more specific sub-categories. The menu-based interfaces were utilized in order to allow information seekers explicitly create their
user profiles and compile queries to submit to the search engine.

According to our user studies (Andrenucci 2006), most users find it difficult to submit information through **menu-based interfaces** without assistance or guidance. In some cases, users needed further assistance concerning the meaning of the domain-specific terms in the taxonomies and in the retrieved documents.

According to research studies among health portal users (Sillence et al. 2005), the number one trust marker for a portal was considered the parameter “the site was easy to use”. In order to enhance user-friendliness, we advise the implementation of exhaustive on-line help modules with tutorials about how to utilize the menu-based interface in the search process, and how to submit user profiles. We also advise the implementation of popular scientific explanations to medical terms in the text, or in pop-up menus that are hyperlinked by the medical terms in the text.

<table>
<thead>
<tr>
<th>Forces and Influences</th>
<th>It requires the help of people with professional knowledge in the field of mental health. It also requires people with pedagogical, explanation skills and knowledge about the use of graphical user interfaces (GUIs).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences</td>
<td>Online help may lead to better usage of the search resources of the portal and a better understanding of mental health information. Information seekers with in-depth domain knowledge may find explanations of medical terms unnecessary and tedious.</td>
</tr>
</tbody>
</table>

## Knowledge Pattern 14

<table>
<thead>
<tr>
<th>Name</th>
<th>Bilingual text processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Language technology patterns</td>
</tr>
<tr>
<td>Founded on</td>
<td>Experiments with language technology applications</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>In order to extract bilingual lexicons using word alignment methods, decisions will have to be made concerning how to pre-process bilingual texts. Wrong choices may affect the quality of the extracted lexicon.</td>
</tr>
<tr>
<td><strong>Context/Criteria</strong></td>
<td>Health portals are aimed for an international audience, but often the language of the documents/information available on the portal is not the native language of the information seekers. When information seekers submit queries/questions in their native language and the documents/answers retrieved by a search engine are in different languages, we have cross-language information retrieval (CLIR). According current research (Kishida k. 2005, Gey et al. 2005) bilingual dictionary look-up is one of the main approaches for performing query translation in CLIR. <strong>A bilingual dictionary</strong> can be extracted from bilingual texts with the help of word alignment (Tiedemann 2003), i.e. identifying translation relationships in bilingual texts (e.g. doktor-doctor, åtstörningar-eating disorders) with the help of natural language processing tools.</td>
</tr>
<tr>
<td><strong>Rationale and Solution</strong></td>
<td>In our studies (Andrenucci 2007, Andrenucci 2010) we utilized two versions of bilingual texts. One version consisted of as-is words (i.e. inflected words) and one version consisted of lemmatized words. Lemmatization is the process of morphological analysis of a word to infer the base form, the lemma of the word.</td>
</tr>
</tbody>
</table>
Each version of the texts was also processed in the following ways: 1) part-of-speech (POS) tagged, i.e. the part of speech of the words was marked up, 2) part-of-speech tagged and syntactically parsed, i.e. the types of phrases (i.e. verb phrases, noun phrases) and their relationships were tagged in the sentences of the texts.

The text alignment tool aligns words with the help of both statistical clues based on redundancy (Och & Ney 2003) and string similarity measures (Melamed 1995). It also utilizes length-based measures to align sentences (Church 1993).

Our studies indicated that part of speech tagging and syntactic information improved the quality of the alignments, for low frequency (1-4 single-word phrases and multi-word phrases). For words with high frequency rates (7-9, >10), the morphological and syntactic information did not improve the quality of the alignments, since statistical clues based on redundancy and string similarity provided enough information for qualitative word alignments. Thus if in the bilingual texts we need to extract mainly words with low frequency rates (1-4), we recommend annotating source and target texts with POS tagging and parsing them syntactically, before extracting alignments at the word level. If we need to extract mainly words with high frequency rates (7-9, >10), the redundancy will provide enough clues in order to extract alignments at the word level, so we do not suggest POS tagging nor syntactic parsing.
**Lemmatization**

Our studies (Andrenucci 2007, Andrenucci 2010) showed that lemmatizing the texts before aligning the words caused less accurate alignments at sentence, paragraph and word level.

The removal of number, definiteness and gender information in Swedish nouns and adjectives, obtained through lemmatization, determined a coarser POS tagging, affecting the alignment clues and worsening the quality of the produced alignments. Thus we do not advise the lemmatization of bilingual texts prior to extracting a lexicon.

Lemmatization can be applied after the word alignment process on the extracted alignments, assembling words with the same base form in the source language or the target language, in order to produce lists with synonyms.

**Forces and Influences**

The bilingual lexicon was extracted with the Uplug toolkit (Tiedemann 2003), a collection of tools for processing parallel texts. The tool aligns texts at the sentence and word level. It utilizes similar word alignment clues (statistical clues, Och & Ney 2003, string similarity, Melamed 1995) and sentence alignment clues (length-based sentence alignments, Church 1993). Lemmatization was performed with the CST Lemmatizer (Jongejan & Haltrup, 2005) a trainable, rule-based tool for languages that utilize inflectional suffixes, such as Swedish and English. A portal designer may decide to utilize other tools in order to pre-process the texts, which may lead to different results than the ones we present here.
Choosing the most suitable pre-processing/annotation approach of the bilingual texts leads to better quality of word alignments with the Uplug toolkit, or tools that utilize similar alignment clues. This ultimately leads to better quality of the extracted lexicons.

<table>
<thead>
<tr>
<th>Papers</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Church K., 1993. Char Alignment, a program for aligning texts at the character level. In ACL’93.</td>
</tr>
<tr>
<td>Melamed D., 1995. Automatic evaluation of uniform filter cascades for inducing N-</td>
</tr>
</tbody>
</table>
best translation lexicons. In 3rd Workshop on Very Large Corpora
