Experiments with Cross-Language Information Retrieval on a Health Portal for Psychology and Psychotherapy

Andrea ANDRENUCCI[1]

Dept. of Computer and Systems Sciences, Stockholm University, Kista, Sweden

Abstract. Few studies have been performed within cross-language information retrieval (CLIR) in the field of psychology and psychotherapy. The aim of this paper is to to analyze and assess the quality of available query translation methods for CLIR on a health portal for psychology. A test base of 100 user queries, 50 Multi Word Units (WUs) and 50 Single WUs, was used. Swedish was the source language and English the target language. Query translation methods based on machine translation (MT) and dictionary look-up were utilized in order to submit query translations to two search engines: Google Site Search and Quick Ask. Standard IR evaluation measures and a qualitative analysis were utilized to assess the results. The lexicon extracted with word alignment of the portal's parallel corpus provided better statistical results among dictionary look-ups. Google Translate provided more linguistically correct translations overall and also delivered better retrieval results in MT.

Keywords. Cross-language information retrieval, Health Portals, E-Health

1. Introduction

Current research about medical information on the web and on medical portals reveals two important aspects about health search patterns: 1) Users of medical portals in general value the possibility of formulating their information needs in their own native language [1] and 2) Not all consumers of online medical information are domain experts or work in health care: a large majority consists of laypeople (i.e. non-experts) with different information needs [2]. A study of search engine logs by Stanton et al. [3] showed that non-medical experts tend also to submit queries in the medical domain using simpler colloquial words, so called "circumlocutions" (i.e. the act of talking around), since they do not know the correct medical terms for the information they seek. It is thus important to allow laypeople to search for online medical information with their own words, in their own native language. The information retrieval process where the user query and the documents/answers retrieved by the search engine are in different languages is called cross-language information retrieval (CLIR). There are three main types of CLIR methods: approaches based on a) bilingual dictionary search, b) machine translation (MT) and c) parallel corpora [4].

There are several studies on CLIR-research in the medical domain (e.g. [5],[6],[7]). However, to our knowledge, few studies have been performed in the field of

---

[1] Corresponding Author: andrea@dsv.su.se

psychology and psychotherapy and no previous study elicited the differences between Single Word Units (SWUs) and Multi Word Units (MWUs). The goal of this study is to analyze and assess the quality of available query translation methods (bilingual dictionary search and MT) in this domain for a source language (Swedish) and a target language (English). Specifically, the following research questions are posed: Do domain-specific dictionaries enhance the quality of retrieved documents? Which query translation methods work for which contexts/words (SWUs/MWUs)? Does the translation quality affect retrieval results? To address these questions, a test base from a portal with multilingual Frequently Asked Questions (FAQs) about mental health is used (Web4health[2]). Three dictionaries and two MT approaches are employed and compared when applied on two different search engines (Google Site Search[3], i.e. a custom search engine for specific websites, and Quick Ask [8]).

2. Methods

Data and CLIR approach: The Web4health medical portal consists of a set of semantically classified FAQs that provide answers in matters of psychological and psychotherapeutic advice. Users submit questions in natural language, which are then matched against pre-stored question/answer pairs, where the question part has a template created to match many different variations of the same question. Two CLIR approaches were used for query translation: machine translation (MT) and dictionary look-up, generating in total five different translations for each query. The queries translated with the different approaches were then submitted to Google Site Search and Quick Ask.

 Machine Translation: Two commonly available MT systems were employed: 1) Google Translate[4] - using the publicly available Google Translator API and 2) Systran[5] - using a non-domain specific API that was provided to the research group during an earlier EU-project.

 Dictionary look-up translations: Three dictionaries were utilized for dictionary look-up translations: 3) a generic bilingual lexicon available online (Lexin[6], used as the lexicon baseline: *lexicon baseline,* approx. 50.000 entries), 4) a domain-specific dictionary developed through word alignment (*word alignment lexicon*), and 5) a domain-specific lexicon based on the English words from the FAQ templates (*template based lexicon)*. The *word alignment lexicon* (approx. 14.000 entries) was extracted with word alignment of Web4health's parallel corpus (as described in previous research, [9]) and part-of-speech tagged as well as syntactically parsed using Uplug [10]. The *template based lexicon* (approx. 16.000 entries) was created utilizing a backward translation approach, i.e. translating from the target language to the source language; all the English words and word stems from Web4health question templates were submitted to two generic Swedish dictionaries, Lexin and Tyda[7], gathering all

---

English to Swedish translations as well as their part-of-speech. The Walter's synonymlexikon[8], was used to extend the dictionary with Swedish synonyms.

Test base: To create the test base we randomly extracted 100 user queries in Swedish from the portal's log: 50 single word units (SWUs), i.e. one word noun phrases (including compound words), and 50 multi word units (MWUs), i.e. verb or noun phrases with two or more words. Relevance assessment of the retrieved results was performed by three human assessors, two information retrieval experts and a domain expert (a psychologist). The result lists were presented to the assessors without any information about the translation method or search engine that gave that result list. The list of retrieved documents consisted of content descriptions of the retrieved FAQs and a link to the FAQ objects containing the complete text.

Evaluation: We utilized both quantitative and qualitative methods for evaluation. To assess the retrieval results, two standard evaluation measures were used: Mean Average Precision (MAP) [11] and Precision at a cut-off of 5 documents (P@5). We also performed a qualitative analysis of the quality of the user queries translations and evaluated how differences in the translations affected the search results. The qualitative analysis was performed as a complement to the statistical results.

3. Results

In Table 1, MAP and P@5 results are presented (highest results in bold, min-max in parenthesis). Table 2 shows some user queries translation examples obtained with different translation approaches.

Machine translation: Google Translate had higher retrieval results than Systran with both search engines and the overall quality of Google translations was higher. Google could manage word sense disambiguation better, finding the translations that better fit the medical/psychotherapeutic context (e.g. *näringstillskott* - nutritional supplements vs Systran's industry boosts, or *vätskedrivande medel* – diuretics vs liquids).

Dictionary lookup: The domain specific dictionaries gave higher statistical results compared to the baseline. The *word alignment lexicon* gave higher precision results for both SWUs and MWUs, with both search engines. The qualitative analysis of the translations and the search results confirmed that creating domain-specific dictionaries was important for finding the word that better fit the psychological context, (i.e. the Swedish words *medel* – funds vs medicines, *besvär* – complaints vs disorders, *leder* – links vs joints, *ond* – evil vs pain), improving the quality of the search results. Word alignment was very helpful for compound word translations (included in SWUs).

Table 1. Statistical results for Mean Average Precision (MAP) and Precision at a cut-off of 5 (P@5).

Translation Method/ Search Engine	Map SWUs	Map MWUs	P@5 SWUs	P@5 MWUs
Lexin/ Quick Ask	0,15 (0,13-0,20)	0,27 (0,17-0,34)	0,34 (0,29-0,44)	0,34 (0,20-0,43)
Lexin/ Google Search	0,17 (0,14-0,20)	0,13 (0,10-0,17)	0,33 (0,29-0,20)	0,11 (0,04-0,16)
Word Alignment /Quick Ask	0,27 (0,23-0,32)	0,37 (0,30-0,43)	0,46 (0,37-0,57)	**0,46** (0,36-0,55)
Word Alignment /Google Search	**0,31** (0,26-0,38)	**0,38** (0,32-0,43)	**0,49** (0,28-0,68)	0,36 (0,25-0,44)
Template Based /Quick Ask	0,18 (0,16-0,21)	0,36 (0,25-0,45)	0,36 (0,32-0,44)	0,44 (0,28-0,55)

---

[8] http://www.synonymer.se

Template Based /Google Search	0,20 (0,18-0,23)	0,13 (0,07-0,18)	0,36 (0,29-0,44)	0,17 (0,09-0,24)
Systran /Quick Ask	0,15 (0,13-0,19)	0,28 (0,16-0,36)	0,35 (0,29-0,44)	0,37 (0,22-0,46)
Systran /Google Search	0,16 (0,14-0,20)	0,12 (0,06-0,17)	0,32 (0,29-0,37)	0,17 (0,08-0,24)
Google Translate /Quick Ask	0,23 (0,19-0,28)	**0,39** (0,33-0,45)	0,45 (0,41-0,54)	**0,49** (0,38-0,58)
Google Translate /Google Search	**0,30** (0,26-0,34)	0,23 (0,12-0,32)	**0,55** (0,51-0,64)	0,31 (0,18-0,40)

The baseline and the *template based lexicon* provided none or poor English translations for compound words (e.g. *skilsmässorådgivning*, *amfetaminmissbruk*), which affected the search results (sleep drive vs insomnia, slim treaty vs slim ideals, mind complex of problems vs memory impairments).

Table 2. Translation examples. **N/A** means that no English translation was provided

Query in source language	Lexin	Word Alignment	Template Based	Systran	Google Translate
näringstillskott	sustenance charge	nutritional supplements	vitamine	industry boosts	nutritional supplements
smalhetsideal	narrow agitation ideal	slim ideals	slim treaty	narrow ideal	ideals of slimness
minskad minnesförmåga	decrease **N/A**	descreased memory	depressed mind	slimmed-down memory ability	reduced memory capacity
minnesproblematik	**N/A**	memory impairments	mind complex of problems	**N/A**	memory problems
sömnstörningar	sleep drive	Insomnia	sleep drive	torpor disturbances	sleep disorders
mättnadskänsla	**N/A**	satisfaction	**N/A**	saturation feeling	satiety
spelmissbruk	courtship abuse	pathologic gamblers	courtship abuse	playing misuses	gambling
skilsmässorådgivning	**N/A**	divorce advice	**N/A**	**N/A**	divorce advice
amfetaminmissbruk	**N/A**	drug abuse	amphetamines abuse	**N/A**	amphetamine abuse
tecken på ätstörningar	character at eat disorder	signs on eating disorders	character of anorexia	sign indicatings eating disorders	signs of eating disorders

Dictionary lookup vs machine translation: Google Translate had linguistically more correct translations than dictionary look-ups, including even compound words and MWUs (e.g. *minskad minnesförmåga* - reduced memory capacity, *mättnadskänsla* – satiety or *spelmissbruk* - gambling). Word alignment sometimes translated adjectives or nouns instead of verbs or vice versa; this did not affect the search quality though, since the context and the meaning of the word was semantically correct (e.g. *självskada* - self-injurious vs Google's correct translation: self-harm). Furthermore, conjunctions and prepositions were problematic (needs **for** relatives vs Google's needs **of** relatives, signs **on** eating disorders vs signs **of** eating disorders) but that did not either affect the information retrieval quality.

Google Search vs Quick Ask: Quick Ask provided higher precision with MWUs than Google Site Search. One possible reason for this is that Quick Ask templates were manually produced by psychologists and therapists and were tailored to match natural language questions within the psychological domain - i.e. Quick Ask resembles more a question answering system than a search engine and it is better tailored for MWUs or longer phrases.

4. Discussion

To our knowledge this is the first study in health portals for psychology that addresses CLIR for laypeople and considers both SWUs and MWUs. Our main contribution is the analysis of the query translation methods from Swedish to English and how they affect retrieval results, which could be informative when developing linguistic resources for this domain. The domain-specific dictionaries enhanced the IR quality, so the usage of such lexicons (on top of generic lexicons) is recommended. In particular the word alignment lexicon provided better statistical results both with SWUs and MWUs among dictionary look-ups. It 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 vocabulary.

In MT Google Translate provided more linguistically correct translations overall (in particular for MWUs and compound words) and better retrieval results (as word alignment). Since extracting a lexicon from parallel corpora is time and resources consuming process, it is advisable to utilize Google Translate from Swedish to English CLIR, unless it is possible to invest time and resources on a bilingual lexicon, i.e. an enduring linguistic resource useful even for future work.

The statistical results of our systems achieved lower precision results compared to the best performing systems in Cross-Language Evaluation Forum (CLEF) eHealth [5], but are comparable to the results in [6], even though no results in these studies involved Swedish as source language. The MAP results are comparable to best performing CLEF systems, which indicates a high recall. As future work it would interesting to add external resources such as medical ontologies and metathesauri as a complement to the baseline and the translation methods described in this paper.

References

[1] M. Koivunen et al., A preliminary usability evaluation of Web-based portal application for patients with schizophrenia. *Journal of Psychiatric and Mental Health Nursing* **14** (2007), 462–469.

[2] L. Wang et al., Using Internet search engines to obtain medical information: a comparative study. *Journal of Medical Internet Research* **14:3** (2012).

[3] I. Stanton, S. Ieong and N. Mishra, Circumlocation in Diagnostic Medical Queries. In *SIGIR'14 Proceedings*. ACM Press.

[4] F. Gey, N. Kando, C. Peters. Cross-language information retrieval: the way ahead. *J of Information Processing and Management: an International Journal* **41** (2005), 415-431.

[5] L. Goeuriot et al. ShARe/CLEF eHealth Evaluation Lab 2014, Task 3: User-Centred health information retrieval. In *Proceedings of Conference and Labs Evaluation Forum (CLEF 2014)*.

[6] P. Pecina et al. Adaptation of machine translation for multilingual information retrieval in the medical domain. *Artificial Intelligence in Medicine* **61** (2014), 165-185

[7] TD. Tran et al. Experiments in cross-language information retrieval using a mixing translation module. In *Proceedings of Medinfo 2004*.

[8] E. Sneiders. Automated Question Answering Using Question Templates that Cover the Conceptual Model of the Database. In *Proceedings of Natural Language Processing and Information Systems, (NLDB'2002)*, 2002, Stockholm, Sweden, pp.235-239

[9] A. Andrenucci. Creating a Bilingual Psychology Lexicon for Cross Lingual Question Answering, A Follow-up Study. *In Proceedings of the 6th International Conference on Web Systems and Technologies, (WEBIST-2010)*, 2003, Valencia, Spain.

[10] J. Tiedemann. Combining Clues for Word Alignment. In *Proceedings of 10th Conference of the European Chapter of the ACL, (EACL'03)*, 2003.

[11] EM. Voorhees and DK. Harman. TREC: experiment and evaluation in information retrieval, *Digital libraries and electronic publishing series* **63** (2005).