Query Translation for Cross-lingual Search in the Academic Search Engine PubPsych

Cristina España-Bonet[1][0000−0001−5414−4710]✉, Juliane
Stiller[2][0000−0001−8184−6187], Roland Ramthun[3][0000−0003−3300−6575], Josef van
Genabith[1], and Vivien Petras[2][0000−0002−8113−1509]

[1] Universität des Saarlandes & DFKI, Saarbrücken, Germany
`cristinae@dfki.de`
[2] Berlin School of Library and Information Science, Humboldt-Universität zu Berlin,
Germany
[3] Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID),
Trier, Germany

Abstract. We describe a lexical resource-based process for query translation of a domain-specific and multilingual academic search engine in psychology, PubPsych. PubPsych queries are diverse in language with a high amount of informational queries and technical terminology. We present an approach for translating queries into English, German, French, and Spanish. We build a quadrilingual lexicon with aligned terms in the four languages using MeSH, Wikipedia and Apertium as our main resources. Our results show that using the quadlexicon together with some simple translation rules, we can automatically translate 85% of translatable tokens in PubPsych queries with mean adequacy over all the translatable text of 1.4 when measured on a 3-point scale [0,1,2].

Keywords: academic search, psychology domain, logfile analysis, query languages, query translation, translation quality, digital library

1 Introduction

Academic search refers to the domain of information retrieval (IR), which concerns itself with searching scientific data, mostly researching output in the form of publications [18]. Academic search is an inherent multilingual challenge. While English is considered the lingua franca of science [1], a significant portion of the world's scientific output is published in other languages [2]. The research presented here is part of the project CLuBS[4] (Cross-Lingual Bibliographic Search) with the aim to improve multilingual access to relevant material in the domain of psychology. The project works with the established academic search engine PubPsych[5] in order to test its approaches on real data and users.

PubPsych is a portal for searching a large and continuously updated database of psychological literature, treatments, tests and research data. It aggregates

---

[4] http://www.clubs-project.eu
[5] https://www.pubpsych.eu

bibliographic metadata from various sources, mainly in English, French, German and Spanish. If the language of the source document is not one of the four, metadata in at least one of these languages is provided. Already in 2008, a survey of psychology researchers found that native language information (besides English) was considered helpful for access [29]. A survey conducted after the launch of PubPsych in 2015 confirmed this and revealed that users were satisfied with the portal, but wished for the possibility to perform multilingual search [32].

The CLuBS project develops, implements and evaluates different approaches to enable cross-lingual (CL) search such as the automatic translation of queries and the translation of metadata content, in both cases with the help of specialized multilingual dictionaries. We use MeSH, Wikipedia and Apertium as our main resources for building these dictionaries. This paper presents the results of the first query translation approach adopted by the project with the objective to provide CL search in the languages English, German, French and Spanish for the PubPsych portal. The guiding research question of the analysis is: can PubPsych queries be translated into the four target languages by mapping them to purpose-built lexical resources? Subquestions are: What proportion of queries can we cover with this approach? How good is the translation quality?

In order to answer these questions we structured the paper as follows. Section 2 discusses related work on academic search engines. In Section 3, we conduct a query analysis using PubPsych query logs from 2014 to 2016 to identify the prevailing query languages and their type. Section 4 describes the methodology and the resources we created for the translation of queries. We evaluate the coverage and the adequacy of the translations in Section 5. Finally, Section 6 concludes with a summary and an outlook on future work.

2 Related Work

Academic search engines (also termed bibliographic information systems or bibliographic digital libraries) are one of the oldest applications in IR (see, for example, MEDLINE[6] [21]) and were used as the first standard test collection for the Cranfield experiments [9]. The interest in these collections reawakened when digital libraries became a prominent research topic as bibliographic information systems moved to web-based user interfaces [8, 26].

Academic search is studied from the user perspective with qualitative methods [25] and in recent years with large-scale logfile analyses of general academic search engines such as Elsevier's ScienceDirect [16, 18] or domain-specific analyses for portals in chemistry [10], computer science [17, 20], history [33], medicine [15, 23, 34], science and technology in general [24], or the social sciences [14]. For the psychological domain, we found one comparative analysis [33].

These studies found significant differences between previously tested search environments —mostly for newspaper or web documents— and the academic search domain. Not only do publications or bibliographic metadata records have

_____

[6] https://www.nlm.nih.gov/bsd/medline.html

a particular document structure but searcher information needs and their representations in queries are different as well. The query content is adapted to the collections and document structure. While studying CiteSeer queries, Khabsa et al. [17] found searches for particular document components such as author, title, or keywords. Queries also contain more technical terms and many more entities [18], which can be mapped to controlled document keywords (such as the MeSH keywords in MEDLINE) for improved search [30].

Multilingual IR is an important research topic for academic search [11]. Qualitative studies show that switching between languages is difficult for searchers [3, 22, 31]. In general, the translation of queries for achieving multilinguality is performed either via dictionary-based methods or by using information extracted from parallel or comparable corpora. The first approaches show ambiguity problems due to possible multiple entries in a dictionary; the second ones limit the coverage to the domains of the corpora. As in most applications, a combination of techniques shows the best performance [28]. Few systems include multilingual functionality though. Because of their similarities with our case study, we refer here to the general domain CL retrieval systems based on machine translation Mulinex [7], MultiLexExplorer [19] and UTACLIR [13], all of them dictionary or knowledge-based engines with possible support for sense disambiguation.

3 Query Characteristics and Languages in PubPsych

First we studied query logs of PubPsych to learn more about query characteristics, the potential for multilingual retrieval and how we could best adapt our strategy for query translation. We analyzed query logs ranging from 1 January 2014 to 31 December 2016 covering 154,495 sessions with 553,799 queries, of which 378,500 were unique queries.[7]

Multilingual users and content. As the retrieval of records in PubPsych is solely based on the bibliographic metadata, it is the language of the metadata which defines whether a record is retrievable in a given language. As for most records the metadata language is not available as separate information, it had to be inferred from the provider's language and their indexing practices. One third of the around 1 million PubPsych records came from PSYNDEX[8], which translates titles in German and/or English for records, which are not in one of these languages. Similarly, Medline, contributed approximately 25% of Pub-Psych records, always providing an English abstract and title and keywords in English, French and German. Retrievable content in Spanish accounted for just approximately 5%. An estimation of the content that cannot be accessed with an English query, because it has no English metadata for any field, amounts to 20%. This uneven distribution of metadata language produces result sets of varying

---

[7] A median of 2 queries was issued over all sessions. PubPsych's mean query length (3.6 tokens for simple, 4.9 for advanced search) was comparable to other reported numbers (e.g. PubMed 3.5 [15], Citeseer 4.8 [17], ScienceDirect 3.8 [18]).

[8] German-speaking countries' database for psychology: https://www.psyndex.de

Annotations	Category description	Example
de, en, fr, es, other	unique language	"Brain"
mixed	at least two languages	"schreiben older adults"
ambi	identical in at least two languages	"Psychologie"
none	not a specific language (e.g. ISSN)	"1869-7712"
unclear	query language and content unclear	"ey Lim"

Table 1: Query language categories

sizes, depending on the query language. Therefore, PubPsych would benefit from CLIR solutions which enhance access to content and improve search results for users. One of these solutions is query translation.

Query languages and types. In order to analyze the language and the type, we took a random sample with replacement of 500 queries from our corpus. Two independent raters manually determined the query language according to the schema in Table 1.

In determining the language, raters reached an interannotator agreement (IAA) of $\pi = 0.89$ (Scott's Pi) with a percentage agreement of 92.2%. The assignment of a query to a language category follows a binomial distribution, allowing us to calculate both a point-estimator and a 95% confidence interval for the real values. In our set, we find $37.4^{+4.3}_{-4.1}$ queries for *de*, $33.4^{+4.2}_{-4.0}$ for *en*, $4.0^{+2.1}_{-1.4}$ for *es*, $3.8^{+2.1}_{-1.4}$ for *fr*, $0.4^{+1.0}_{-0.3}$ for *other*, $1.2^{+1.4}_{-0.6}$ for *mixed*, $4.0^{+2.1}_{-1.4}$ for *ambi*, $0.6^{+1.1}_{-0.4}$ for *unclear*, and $15.2^{+3.4}_{-2.9}$ for *none*. English and German queries were prevalent, with each accounting for about one third of the total queries, followed by queries with non-linguistic content (e.g. ISSN) with a share of around 15%. French, Spanish and ambivalent queries were issued with each around 5% share, while other query language categories were less common. The fact that more than half of the sessions originated from Germany suggests that German users are also searching in English or other languages. The relatively high amount of non-language-specific queries corresponds to the amount of queries dealing exclusively with named entities, which often cannot be attributed to a language.

The raters also determined whether a query is *informational, navigational* or *transactional* following Broder's definition [6] which was refined by Li et al. [18] for academic search engines. An informational query seeks for topics, specific authors, places or other themes. The definition of a navigational query for academic search involves publication identifiers (such as ISSN, DOIs or the use of the title field operator) to retrieve particular articles or publications. Transactional queries only play a marginal role as they seek the original resource by looking for downloadable content like PDF files. We found no transactional queries and a low number of navigational queries resulting in a Scotts's Pi[9] of 0.27 for the IAA despite a high percentage agreement of 89.4%. Table 2 compares our results to other studies in different domains. Informational queries are the most frequently observed query type across all domains but are particularly

---

[9] We also calculated Krippendorff's alpha and Cohen's Kappa with the same results.

Query Type	PubPsych (Psychology)	CiteSeer [17] (Comp. Sci.)	ScienceDirect [18] (General academic)[10]	Library [5][10]	Web search [6][10]
Informational	88.4%	87.5%	92.3%	47.85%	48.0%
Navigational	11.6%	12.5%	7.6%	50.25%	20.0%
Transactional	0.0%	n.d.	0.5% (downloads)	n.d.	30.0%

Table 2: Query types and their distribution in different domains

high for academic search engines. With regard to query translation, informational queries need translation whereas most navigational queries do not (e.g. DOI, ISSN searches) and should not be translated. Given that the 88.4% of PubPsych queries are informational with domain specific vocabulary and there are no transactional queries, we choose to use an approach of mapping lexical resources in four languages to translate our queries.

4 Approach to Query Translation

Mapping queries to a multilingual vocabulary. One approach to translate queries is to map them to a multilingual controlled vocabulary of the domain under consideration. Controlled vocabularies contain technical terminology for a domain and are used to index content in retrieval systems. A multilingual version contains precise translations of such technical terms and is therefore a useful resource. Observing a high amount of technical terminology in our queries, we tested how many queries could be translated by matching them to multilingual thesauri and implement some simple rules to maximize the matching.

This mapping approach does not require the explicit detection of the language of a query, which is considered to be a hard task [27]. We matched tokens from the query against the keys in our quadlexicon in German, English, Spanish and French, and extracted the three non-matching languages as translations. In case of multiple matches for a key, we assumed the most frequent language derived from the query language distribution in PubPsych. The following excerpt of the lexicon shows how the first field, the key, does not include any language information, only translations do:[11]

```
bienestar|||en:well-being|||de:wohlbefinden|||fr:bien-etre
bien-etre|||en:well-being|||es:bienestar|||de:wohlbefinden
wohlbefinden|||en:well-being|||es:bienestar|||fr:bien-etre
wohlfuhlen|||en:well-being|||es:bienestar|||fr:bien-etre
well-being|||es:bienestar|||de:wohlbefinden|||fr:bien-etre
```

The complete methodology used to translate a query in this language-independent fashion can be seen in Figure 1. Initially we parsed a query, removed punctuation

---

[10] Percentages do not add up to 100% but are reported as found in the papers. Whereas in [18] and [6] no explanation is given, [5] report on another two categories "non-classifiable" and "other" for 1,95% of their queries.

[11] This method additionally solves the problem of intra-query language shifts, since different tokens in the same query can be matched to different languages.

INPUT: Parsed term
action potentials

in quad-lexicon?

Extract entry
action potentials ||| es:potenciales de acción ||| de:aktionspotentiale ||| fr:potentiels d'action

yes

no

OUTPUT: Quad-term

Split by tokens

Token alternatives

in quad-lexicon?

yes

Extract entry

yes

Singular form

no

more tokens?

in quad-lexicon?

no

Recompose full entry
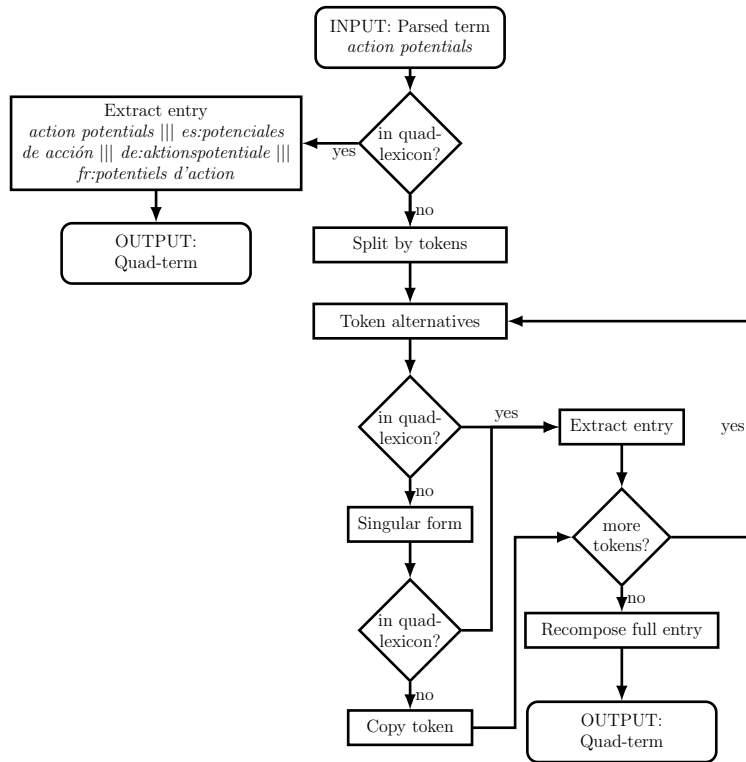
Copy token

no

OUTPUT: Quad-term

Fig. 1: Flowchart for query term translation. If a complete term cannot be matched, a token by token translation is applied

and non-alphanumeric characters, and extracted text not belonging to a specific field (unbound text) and text for translatable fields. Examples of translatable fields[12] included titles, controlled terms and keywords. Examples of untranslatable fields[13] included years, journal titles and authors. The subset of queries with translatable text comprised 536,479 elements. These elements were mainly from informational queries — an exception were searches in the title field where one can seek an exact title (navigational search where translation is debatable) or look for keywords in title (translation is desired). Each translatable term within a query was then translated independently. A term might have more than one token. In case the full term was found in our multilingual lexicon, we extracted the translations. If not, we split the terms into single tokens, and tried to match each token individually. For those tokens still not present in the lexicon, we derived its possible singular form by applying simple rules of regular formation of plurals in the four languages and tried again. Finally, we copied the token as its own translation in case no match was found. Note that copying the source word into the output does not necessarily correspond to an incorrect transla-

---

[12] CM, IT, SH, CT, SW, TI and AB fields in PubPsych
[13] AGE, EV, PLOC, AU, ISBN, ISSN, PU, SEG, CS, JT, DB, PY, LA, DT and ID

	German	English	French	Spanish
MeSH	70,694	175,004	96,333	66,828
WP (titles/categories)		(81,369/38,038)		
Apertium	7,792	5,935	6,020	5,846
Manual	4,262	4,142	4,047	4,081
Total unique (QuadLex)	202,128	304,277	225,607	195,937

Table 3: Aligned terms per language in the quadrilingual lexicon. Mismatches between the numbers of a row hint to the availability of synonyms for a language

tion because in most cases the words not contained in the lexicon were named entities. Equivalently, using the quadrilingual lexicon to translate an entry did not guarantee a correct translation, because, besides of the existing noise, the concatenation of word translations does not need to correspond to a correct term translation. Section 5 reports the evaluation of our approach.

Quadrilingual lexicon. We created an aligned dictionary in English, Spanish, French and German for our experiments. The lexicon covered specially the domain of psychology but also contained general sources to increase the coverage:

MeSH: A common thesaurus in the bio-medical domain is the Medical Subject Headings[14] (MeSH) by the U.S. National Library of Medicine. We developed the MeSHMerger software[15] to create a multilingual MeSH version from different translations of its 2017 edition. Each of the 54,901 concepts in MeSH can have a varying number of terms, permutations of terms and synonyms attached to it in each language version. Extracting the available entries in the four languages, we obtained 175,004 concept related terms in English, 96,333 in French, 70,694 in German and 66,828 in Spanish.

Multilingual Wikipedia Entries: To increase the amount of psychological term translations, we extracted multilingual in-domain titles from Wikipedia related to psychology and health with the WikiTailor tool[16][4]. WikiTailor extracts domain articles by exploring the categories graph starting from the category describing the domain (psychology and health in our case) and identifying a subset of related categories and their associated articles[17]. Such associated articles were gathered independently in English, German, French and Spanish versions of Wikipedia. Afterwards, we expanded the set of articles to include all the articles identified as in-domain articles in at least one of the languages with the equivalent article in the other three languages in case they exist. With this methodology we obtained a multilingual lexicon of article names with 81,369 entries.

---

[14] https://www.nlm.nih.gov/mesh
[15] https://github.com/clubs-project/MeSHMerger
[16] https://github.com/cristinae/WikiTailor
[17] We used models WT0.5-100 or WT0.5-500 depending on the language. Refer to WikiTailor manual for more details http://cristinae.github.io/WikiTailor

In a similar way, we extracted aligned category names from Wikipedia, but this time selecting all of them and not only those related to psychology. The category alignment in Wikipedia is quite clean and this allowed us to increase the coverage introducing little noise. 38,038 entries were obtained in this case.

Apertium Dictionaries: Apertium [12] is a free/open-source ruled-based translation engine that uses bilingual dictionaries for lexical transfer. We used three of their dictionaries[18] (*en-de*, *en-es* and *es-fr*) to extract a quadrilingual dictionary with the overlapping entries. Notice that different from Wikipedia entries, dictionary entries were not sense disambiguated and the union of the four languages could have introduced additional noise to the translation. The motivation was to widen the coverage of out-of-domain vocabulary also used in queries frequently. Table 3 shows the number of entries of this multilingual dictionary in comparison with the other sources.

Post-edited Automatic Translations: Finally, we selected a set of highly frequent tokens from PubPsych controlled terms which were not already covered by the previous resources. This mainly affected $\sim$4,000 specialized in-domain terms and expressions, which we translated with the automatic translation engine DeepL[19] and manually post-edited to improve mistranslations. Table 3 shows the exact number of entries depending on the source language in the row "Manual".

The complete multilingual lexicon, *QuadLex*, is the union of the resources in the four languages with 927,949 unique elements. Since internally terms are sent to the search engine lowercased and without diacritics, we prepared a normalized version of the lexicon with these characteristics in order to translate the queries in a realistic setting, that is, after the normalization and parsing of the query takes place. That gave a total of 927,764 entries, with 680,567 being unique. The number of elements per source language is shown in Table 3.

5   Evaluation

We evaluated our methodology to translate queries by two means. First, we studied the coverage of the quadrilingual lexicons and second, their translation quality.

Table 4 shows the coverage of the plain MeSH and extended QuadLex multilingual lexicons on the set of 536,479 queries with translatable terms. When we translated terms as a whole with the MeSH lexicon, coverage was only 7.7% of the terms. We expect this subset to have high quality translations. The inclusion of the out-of-domain resources increased the coverage to 14.9%, but the true improvement resulted from the translation at token level. In this case, we could cover up to 85.0% of the translatable tokens in PubPsych queries.

We evaluated the quality of translations for 500 queries manually. For this, we used a subset of the annotated queries described in Section 3 and added new manually annotated queries in order to obtain 100 queries in each language

---

[18] http://wiki.apertium.org/wiki/List_of_dictionaries
[19] https://www.deepl.com, work took place as of 25th Jan. and 1st-2nd Feb. 2018

Source	Trad. Terms(%)	Untrad. Terms(%)	Trad. Tokens(%)	Untrad. Tokens(%)
MeSH	167,152 (7.7)	2,010,469 (92.3)	2,225,598 (64.2)	1,240,800 (35.8)
QuadLex	324,033(14.9)	1,853,588 (85.1)	2,945,959 (85.0)	520,439 (15.0)

Table 4: Coverage of MeSH and QuadLex on the set of 536,479 queries with translatable terms

	Rater A			Rater B			Rater C			Mean		
	0	1	2	0	1	2	0	1	2	0	1	2
en2de	20	34	46	0	41	59	13	41	46	11±10	39±4	50±8
en2es	14	36	50	1	43	56	6	37	57	7±7	39±4	54±4
en2fr	17	39	44	0	42	58	8	43	49	8±8	41±2	50±7
de2en	36	31	33	21	40	39	23	40	37	27±8	37±5	36±3
de2es	39	31	30	24	37	39	25	41	34	29±8	36±5	34±4
de2fr	38	32	30	23	38	39	25	41	34	28±8	37±5	34±5
es2en	25	46	29	3	44	53	8	39	53	12±12	43±4	45±14
es2de	24	51	25	3	46	51	14	45	41	14±11	47±3	39±13
es2fr	21	43	36	4	47	49	12	37	51	12±9	42±5	45±8
fr2en	32	41	27	1	46	53	7	48	45	13±16	45±4	42±13
fr2de	44	33	23	5	42	53	14	45	41	21±20	40±6	39±15
fr2es	31	34	35	4	48	48	8	41	51	14±15	41±7	45±9
none2en	10	4	86	2	11	87	7	11	82	6±4	9±4	85±3
none2de	10	2	88	3	13	84	11	8	81	8±4	8±6	84±4
none2es	11	0	89	4	11	85	8	9	83	8±4	7±6	86±3
none2fr	13	1	86	5	10	85	12	7	81	10±4	6±5	84±3
Mean	24±11	29±17	47±25	6±8	35±14	59±14	13±6	33±15	54±18	**14±8**	**33±15**	**54±20**

Table 5: Number of queries evaluated with a score for adequacy 0/1/2, per rater and in average, in our test set of 500 queries with 100 items per source language

(*de*, *en*, *fr* and *es*) plus 100 queries without a definite language identification (*mixed/unclear/ambi/none*). The average length for the 100 English queries was 3.1 words, 2.2 for German, 3.5 for Spanish, 3.6 for French and only 1.7 words when no language was assigned. We evaluated the translations according to adequacy defined as how much of the meaning expressed in the source query was also expressed in the translation. Since we did not have gold translations, adequacy was defined with respect to the source and not to a gold. Given the short length of the queries, we measured the adequacy by a three-point scale: *0-None of the meaning was transfered, 1-Part of the meaning was transfered, 2-All meaning was transfered.*

Three evaluators performed the task and Table 5 shows the raw results together with average values and uncertainties given by the standard deviation. The IAA for the different languages is presented in Table 6 using Fleiss' kappa statistic for more than two raters. The main divergence among annotators was due to the fact that one of them (A) considered the meaning of non-content

source	2de	2en	2fr	2es
de	n/a	0.616	0.658	0.598
en	0.442	n/a	0.455	0.521
fr	0.243	0.268	n/a	0.384
es	0.422	0.354	0.472	n/a
none	0.494	0.458	0.513	0.440

Table 6: IAA (Fleiss' kappa) of three raters for different language pairs

words to rate for adequacy while the other two did not. Despite this discrepancy, some conclusions could be gathered. We obtained an average adequacy of 1.4 on the [0,1,2]-scale, meaning that most of the queries had at least some of their terms properly translated. 54%±20% of the queries had the maximum adequacy score when looking at the mean over languages, while only 14%±8% of the queries got completely incorrect translations; the remaining 33%±15% were partially well translated. The behavior per language was quite similar with two clear exceptions: *(i)* the translation of German queries had a lower quality (mean adequacy 1.1) mainly because the compound nature of German increases the number of untranslated tokens with respect to the other languages, and *(ii)* queries with undetermined language had a very high adequacy (1.8) because they are shorter and, in most of the cases, leaving the source token untranslated resulted in a good translation.

6   Conclusion

In the query analysis, we observed queries that are typical for academic search engines especially regarding the occurrence of domain-specific terminology. About 12% of the queries were navigational and contained components that should not be translated, but the remaining 88% were informational and need to be translated for CL search.

We have shown how precise term mappings could be successfully applied with the help of multilingual thesauri without the need to derive this data from general parallel corpora. We built a quadrilingual lexicon with aligned terms in German, English, French and Spanish using as main resources MeSH, Wikipedia and Apertium. After parsing the queries and extracting the translatable terms, we mapped them to the lexicon in a language-independent way to extract translations in the four languages. The design of the mapping approach was intended to maximize the retrieval recall and not translation quality by falling back to a word by word translation and generalizing for gender and number when the complete mapping was not found. With this approach, we covered 85% of the tokens in PubPsych queries and obtained an average adequacy of 1.4 on a [0,1,2]-scale over the full set.

The main advantage of our approach, besides simplicity, is that thesauri assure the correct translation of the domain-specific terminology and we only introduce ambiguities for the general domain vocabulary. The main disadvantage

is that it does not reach a 100% of coverage as data-based systems can achieve. Our future work involves using multilingual word embeddings to increase the coverage and solve this limitation of dictionary-based methods in combination with the usage of the MeSH quadlexicon to still assure the high quality translation of the domain-specific terminology. By analyzing the translations we found several systematic errors such as those seen when trying to translate non-content words. We also detected a number of untranslated terms for German due to its compound nature. In the future, we will approach these issues to further improve the translation quality and CL retrieval performance.

References

1. Alastrué, R.P., Pérez-Llantada, C.: English as a Scientific and Research Language: Debates and Discourses. de Gruyter (2015)
2. Amano, T., González-Varo, J.P., Sutherland, W.J.: Languages Are Still a Major Barrier to Global Science. PLoS biology 14(12), e2000933 (2016)
3. Aula, A., Kellar, M.: Multilingual Search Strategies. In: Conference on Human Factors in Computing Systems (CHI). pp. 3865–3870. ACM (2009)
4. Barrón-Cedeño, A., España-Bonet, C., Boldoba, J., Màrquez, L.: A Factory of Comparable Corpora from Wikipedia. In: Proceedings of the 8th Workshop on Building and Using Comparable Corpora (BUCC). pp. 3–13 (July 2015)
5. Behnert, C.: Evaluation Methods within the LibRank Project. Working Paper, LibRank (2016)
6. Broder, A.: A taxonomy of web search. In: ACM Sigir forum. vol. 36, pp. 3–10. ACM (2002)
7. Capstick, J., Diagne, A.K., Erbach, G., Uszkoreit, H.: MULINEX: Multilingual Web Search and Navigation. In: Proceedings of the 14th Twente Workshop on Language Technology (TWLT 14) (1998)
8. Chowdhury, G.: Introduction to Modern Information Retrieval. Facet (2010)
9. Cleverdon, C.: The Cranfield tests on index language devices. In: Aslib proceedings. vol. 19, pp. 173–194. MCB UP Ltd (1967)
10. Davis, P.M.: Information-seeking Behavior of Chemists: A Transaction Log Analysis of Referral URLs. J. Amer. Soc. Inf. Sci. and Tech. 55(4), 326–332 (Feb 2004)
11. Diekema, A.R.: Multilinguality in the digital library: A review. The Electronic Library 30(2), 165–181 (2012)
12. Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Sánchez-Martínez, F., Ramírez-Sánchez, G., Tyers, F.M.: Apertium: A Free/Open-source Platform for Rule-based Machine Translation. Machine Translation 25(2), 127–144 (2011)
13. Hedlund, T., Airio, E., Keskustalo, H., Lehtokangas, R., Pirkola, A., Järvelin, K.: Dictionary-based cross-language information retrieval: Learning experiences from CLEF 2000–2002. Information Retrieval 7(1-2), 99–119 (2004)
14. Hienert, D.: User Interests in German Social Science Literature Search: A Large Scale Log Analysis. In: Conference on Human Information Interaction & Retrieval (CHIIR). pp. 7–16. ACM (2017)

15. Islamaj Dogan, R., Murray, G.C., Névéol, A., Lu, Z.: Understanding PubMed® user search behavior through log analysis. Database 2009 (2009)
16. Ke, H.R., Kwakkelaar, R., Tai, Y.M., Chen, L.C.: Exploring behavior of E-journal users in science and technology: Transaction log analysis of Elsevier's ScienceDirect OnSite in Taiwan. Library & Information Science Research 24(3), 265 – 291 (2002)
17. Khabsa, M., Wu, Z., Giles, C.L.: Towards better understanding of academic search. In: Joint Conference on Digital Library (JCDL). pp. 111–114. ACM (2016)
18. Li, X., Schijvenaars, B.J., de Rijke, M.: Investigating queries and search failures in academic search. Information Processing and Management 53(3), 666 – 683 (2017)
19. Luca, E.W.D., Hauke, S., Nürnberger, A., Schlechtweg, S.: MultiLexExplorer: Combining Multilingual Web Search with Multilingual Lexical Resources. In: Proc. of Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken Dialogue Systems. pp. 17–21 (2006)
20. Mahoui, M., Cunningham, S.J.: Search behavior in a research-oriented digital library. In: European Conference on Research and Advanced Technology on Digital Libraries (ECDL). pp. 13–24. Springer (2001)
21. McCarn, D.B., Leiter, J.: On-line services in medicine and beyond. Science 181(4097), 318–324 (1973)
22. Nzomo, P., Ajiferuke, I., Vaughan, L., McKenzie, P.: Multilingual information retrieval & use: Perceptions and practices amongst bi/multilingual academic users. The Journal of Academic Librarianship 42(5), 495 – 502 (2016)
23. Palotti, J.a., Hanbury, A., Müller, H., Kahn, Jr., C.E.: How users search and what they search for in the medical domain. Inf. Ret. 19(1-2), 189–224 (2016)
24. Park, M., Lee, T.S.: A longitudinal study of information needs and search behaviors in science and technology: A query analysis. The Electr. Lib. 34(1), 83–98 (2016)
25. Pontis, S., Blandford, A., Greifeneder, E., Attalla, H., Neal, D.: Keeping up to date: An academic researcher's information journey. J. Amer. Soc. Inf. Sci. and Tech. 68(1), 22–35 (2017)
26. Ritchie, A., Teufel, S., Robertson, S.: Creating a test collection for citation-based ir experiments. In: HLT-NAACL '06. pp. 391–398. ACL (2006)
27. Stiller, J., Gäde, M., Petras, V.: Ambiguity of Queries and the Challenges for Query Language Detection. In: CLEF (2010)
28. Ture, F., Boschee, E.: Learning to Translate: A Query-Specific Combination Approach for Cross-Lingual Information Retrieval. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 589–599. Association for Computational Linguistics, Doha, Qatar (2014)
29. Uhl, M.: Survey on european psychology publication issues. Psychology Science Quarterly 51(1), 19 (2009)
30. Vanopstal, K., Buysschaert, J., Laureys, G., Stichele, R.V.: Lost in PubMed. Factors influencing the success of medical information retrieval. Expert Systems with Applications 40(10), 4106 – 4114 (2013)
31. Vassilakaki, E., Garoufallou, E., Johnson, F., Hartley, R.J.: An exploration of users' needs for multilingual information retrieval and access. In: International Conference on Metadata and Semantics Research (MTSR). pp. 249–258. Springer (2015)
32. Waeldin, S.: Results from the PubPsych launch survey: Short report. ZPID Science Information Online 15(2), 3 (2015)
33. Yi, K., Beheshti, J., Cole, C., Leide, J.E., Large, A.: User Search Behavior of Domain-specific Information Retrieval Systems: An Analysis of the Query Logs from PsycINFO and ABC-Clio's Historical Abstracts-America: History and Life: Research Articles. J. Amer. Soc. Inf. Sci. and Tech. 57(9), 1208–1220 (2006)

34. Yoo, I., Mosa, A.S.M.: Analysis of Pubmed user sessions using a full-day Pubmed query log: A comparison of experienced and nonexperienced Pubmed users. JMIR Medical Informatics 3(3) (2015)