Effects of Language and Terminology on the Usage of Health Query Suggestions

Carla Teixeira Lopes and Cristina Ribeiro

DEI, Faculdade de Engenharia, Universidade do Porto and INESC TEC,
Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
`{ctl,mcr}@fe.up.pt`

Abstract. Searching for health information is one of the most popular activities on the Web. In this domain, users frequently encounter difficulties in query formulation, either because they lack knowledge of the proper medical terms or because they misspell them. To overcome these difficulties and attempt to retrieve higher-quality content, we developed a query suggestion system that provides alternative queries combining the users' native language and English language with lay and medico-scientific terminology. To assess how the language and terminology impact the use of suggestions, we conducted a user study with 40 subjects considering their English proficiency, health literacy and topic familiarity. Results show that suggestions are used most often at the beginning of search sessions. English suggestions tend to be preferred to the ones formulated in the users' native language, at all levels of English proficiency. Medico-scientific suggestions tend to be preferred to lay suggestions at higher levels of health literacy.

Keywords: health information retrieval · query suggestion · English proficiency · health literacy · topic familiarity

1 Introduction

Searching for health information is the third most popular online activity after email and using a search engine, being performed by 80% of U.S. Internet users [2]. This domain poses specific challenges to health consumers, who frequently encounter additional difficulties in finding the correct terms to include in their queries [16, 6] because they lack knowledge of the proper medical terms [17, 14]. The misspelling of medical terms is another common problem [5, 11]. For these reasons, support in query formulation may contribute to an improved retrieval experience. Considering this is a domain in which the quality of the retrieved content is crucial, and considering that quality depends on the language used to conduct the search [7], support for query translation may also be useful.

Our goal is to improve the health search experience of users, particularly users for whom English is not their primary language. The importance of query formulation support in health searches, the lack of such support and the findings of previous studies motivated the development of a system that, based on an

initial user query, suggests 4 different queries combining two languages (English and Portuguese) and two bodies of terminology (lay and medico-scientific). To the best of our knowledge, no previous works have explored cross-language query suggestions in the health domain.

To assess users' receptivity to query suggestions in a language that not their native and to suggestions using different types of terminologies, we conducted a user study and, based on clicks, analysed the effective use of the proposed suggestions. It is important to note that, although it has been proved that some of these suggestions contribute to improve the retrieval performance [7, 8], this will only be the case if users have the willingness to take the recommendations.

Previous studies have concluded that search assistance should be personalized to achieve its maximal outcome [4]. Yet, little attention has been paid to how people perform query reformulations across different user groups. Therefore, we have considered users' English proficiency, health literacy and topic familiarity.

2   Related Work

In consumer health information retrieval, there is an awareness that several difficulties can emerge due to the terminology gap between medical experts and lay people [18]. To overcome these difficulties in query formulation, some authors have proposed query expansion approaches. The Health Information Query Assistant proposed by Zeng et al [16] suggests terms based on their semantic distance from the original query. To compute this distance, the authors use co-occurrences in medical literature and log data as well as the semantic relations in medical vocabularies. A user study with 213 subjects randomized into 2 groups, one receiving suggestions and the other not receiving them, showed that recommendations resulted in higher rates of successful queries, i.e., queries with at least one relevant result among the top 10, but not in higher rates of satisfaction. Two proposed search engines for health information retrieval — iMed [9] and MedSearch [10] — provide suggestions of medical phrases to assist users in refining their queries. The phrases are extracted and ranked based on MeSH (Medical Subject Headings), the collection of crawled webpages, and the query.

Zarro and Lin [15] presented a search system that uses social tagging and MeSH to provide users with peer and professional terms. To evaluate the impact of these suggestions, the authors conducted a user study with 10 lay subjects and 10 expert subjects. Both groups preferred MeSH terms because their quality was considered superior to the quality of social tags. Also in the health domain, Fattahi et al [1] proposed a query expansion method that uses non-topical terms (terms that occur before or after topical terms to represent a specific aspect of the theme, such as 'about' in 'about breast cancer') and semi-topical terms (terms that do not occur alone, such as 'risk of' in 'risk of breast cancer') in conjunction with topical terms (terms that represent the subject content of documents, such as 'breast cancer'). The authors found that web searches can be enhanced by the combination of these three types of terms.

Although not in the specific area of health information retrieval, we identified only one work involving the proposal of query suggestions in a language different from the original query's language, namely, a study performed by Gao et al [3]. The authors proposed a method to translate generalist queries using query logs and then estimate the cross-lingual query similarity using information such as word translation relations and word co-occurrence statistics. The evaluation was performed on French-English and Chinese-English tasks. They found that these suggestions, when used in combination with pseudo-relevance feedback, improved the effectiveness of cross-language information retrieval.

Since 2014, the Conference and Labs of the Evaluation Forum (CLEF) eHealth lab began to propose a multilingual information user-centred health retrieval task, incorporating queries in several languages in its dataset.

3 Suggestion Tool

We designed and developed a prototype for a suggestion tool that can be integrated into IR systems. Given a health query, our tool suggests alternative queries in two languages, Portuguese (PT) and English (EN), using medico-scientific (MS) and lay terminology. In Figure 1, we present the architecture of the suggestion tool, which will be further detailed in the following paragraphs.
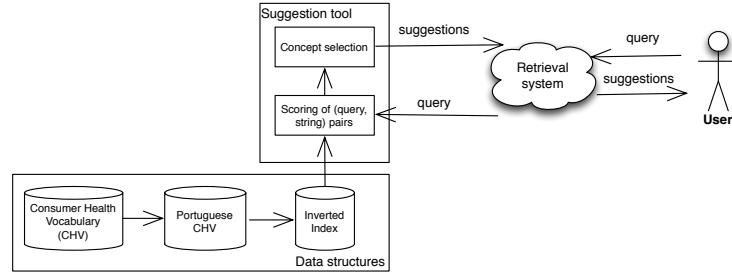

Fig. 1. Architecture of the suggestion tool.

We created an inverted index using the Portuguese translation of the Consumer Health Vocabulary (CHV), an open-access and collaborative vocabulary that maps technical terms to consumer friendly language [12]. In the index, each stemmed term is associated with an inverse string frequency (isf_t) and a postings list, i.e., a list of the strings in which the stemmed term appears. The computation of the inverse string frequency is similar to the computation of the inverse document frequency that is traditionally performed in IR, that is, $isf_t = log(N/sf_t)$, where sf_t is the number of strings in which the term appears and N is the total number of strings. Because strings are typically small, the probability of finding multiple occurrences of the same term in a string is very

small. For this reason, we decided to weight each term based only on its isf_t, ignoring its frequency in each string ($tf_{t,s}$). To determine the vocabulary of terms, namely, the list of terms in our inverted index, the strings were tokenized and stop words were removed. In the terms, all letters were converted to lower case and the accents were removed, and the terms were also stemmed.

The score assigned to each (query, string) pair is defined by $score(q,s) = \sum_{t \in q} isf_t$. Because the length of strings and queries has a very small variance, we found that the additional computational power required to normalize the above score formula would not be justified by the gains thus achieved.

In this stage of the prototype development, to limit the number of suggestions, we decided to select only the string with the maximum score for each input query. For this string, we identify the associated concept and then return its CHV and Unified Medical Language System (UMLS) preferred names in English and Portuguese. If a suggestion is identical to the query or to any other suggestion, it will not be presented, i.e., the output of the system will contain only unique suggestions different from the query. This results in a maximum of 4 suggestions for each query.

The CHV vocabulary was translated into Portuguese using the Google Translator API (Application Programming Interface). We manually evaluated 1% of the total number of translated strings and concluded that 84.2% (95% CI: [82.3%, 85.9%]) of the translations were correct, a very satisfactory outcome.

Our retrieval system used the Bing Search API to obtain web results for users' queries. To increase the usability of the interface with regard to learning, we decided to keep the interfaces very simple and similar to those used in the most popular search engines. All the suggestions are presented in a single line above the list of retrieved documents. As an example, a set of suggestions could be: "colectomia", "remoção do cólon", "colectomy" and "colon removal".

All types of suggestions might be useful after any type of query. Imagine, for example, an initial query in English. Portuguese query suggestions might be useful for an user that is not proficient in the English language and prefers to retrieve documents in his native language. On the other hand, if he is proficient in English he might just want to switch terminology, keeping the same language.

4   Experiment

We conducted a user study with 40 participants (24 female; 16 male), with a mean age of 23.48 years (standard deviation (sd)=7.66). Portuguese was the native language of every participant. The study had two within-subject independent variables: language and terminology; three quasi-independent variable: English proficiency, health literacy and topic familiarity; and one major dependent variable: clicks on query suggestions.

English proficiency was evaluated using an instrument developed by the European Council that grades English proficiency in the Common European Framework of Reference for Languages (CEFR), a widely accepted European standard for this purpose. To evaluate the users' health literacy, we have used the Medical

Term Recognition Test (METER), an instrument proposed by Rawson et al [13]. The users' familiarity with each topic was self-assessed on a five-level scale.

Each user was assigned a set of 8 tasks, each associated with one of 8 simulated work task situations. To define the simulated situations, we selected 20 persons with no medical expertise and spanning a wide range of ages (from 30 to 68) and education levels (from high school to PhD degrees). These individuals were asked to state the health topic for which they had most recently searched on the Web. From these topics, we randomly selected 8 and created a scenario for each. The situations were described to the users in Portuguese. The following situation, presented as an example, was one of them: "Two weeks ago, someone from your family was diagnosed with shingles. To understand what characterizes this disease, you decided to search on the Web for its causes and symptoms. Find out what causes the disease and identify two common symptoms.".

In each task the user had to formulate 3 queries and assess the relevance of the top 10 results for each query. In the first query, the user had to formulate the query without any help from the system. Users did not have any type of restrictions in query formulation. Users used their preferred language and terminology. Based on the initial query, the system presents suggestions that can, or not, be used for the formulation of the second query. The same happens when the user is moving from the second to the third iteration. The set of 3 iterations constituted a search session. The usefulness of the suggestions was assessed in a post-search questionnaire.

Our experiment was motivated by the following research questions: (1) In which stage of the search process are suggestions used, and considered useful, more often? (2) To which type of suggestions are users most receptive? (3) To which language are users most receptive and how does this change with the users' English proficiency? (4) To which terminology are users most receptive and how does this change with the users' health literacy and topic familiarity?

5   Data Analysis

To evaluate the usage given to suggestions, we considered that users might use them as suggestions, clicking or not on them, and as source of terms they can use in the following queries. Considering this last scenario, for each type of suggestion, we computed the proportion of suggestion's terms that were used in the subsequent query (termsUsed) and the proportion of the suggestion's terms that were used in the following query and were not used in the previous query (newTermsUsed). The former is useful to assess the quality of suggestions' terms and the latter is also useful to assess the utility of the suggestions for the users. Let Q_{it} be the set of unique stemmed terms belonging to the query of the iteration it and S_{it} the set of unique stemmed terms belonging to the suggestion presented in the iteration it, these proportions are computed as follows: $termsUsed_{it} = \frac{|Q_{it} \cap S_{it}|}{|S_{it}|}$ and $newTermsUsed_{it} = \frac{|(Q_{it} \cap S_{it}) \setminus Q_{it-1}|}{|S_{it}|}$.

We used the test of equal proportions with the chi-squared value to compare proportions between groups. To compare the means of termsUsed and

newTermsUsed between groups we used the Student's t-test. When the variances were found to be not homogeneous, we applied the Welch t-test. In comparisons involving more than two groups (e.g.: comparing levels of English proficiency in terms of suggestion usage), we applied the one-way ANOVA and the Tukey's test to assess the location of the differences, whenever significant differences were found. When reporting our results, we use a * to mark significant results at $\alpha = 0.05$ and a ** to mark significant results at $\alpha = 0.01$.

6   Use of Suggestions

As explained in Section 3, all types of suggestions were generated through the same algorithm. Moreover, the translation of the CHV was considered very good with almost 85% of correct translations. Consequently, we don't expect quality differences between types of suggestions that might have affect user behaviour.

During the experiments, in the second and third iterations, the suggestion system did not present suggestions in only 4.7% of the iterations. Note that the system generates suggestions based on the query just inserted by the user, what means the first query has to be formulated without any help. All these initial queries were formulated using the Portuguese language and lay terminology.

Almost 55% of the sessions presenting suggestions had one or more clicks in suggestions. Of the 40 participants in the study, 5 did not click on any suggestion during their tasks. Suggestions were used in 86.9% of the iterations and the participants tended to find the suggestions useful more often in the initial iterations than in the final ones (Table 1). This is also supported by the proportions presented in the two last lines of Table 1, showing that initial iterations have more suggestions' clicks and use suggestions as a source of terms more often.

Table 1. Use of suggestions by iteration. "Proportion of iterations where..."

	Iteration		
	2	3	2 versus 3
suggestions were not presented	1.9%	7.5%	$\chi^2(1) = 4.5$, p=0.02*
suggestions were not considered useful	10.0%	16.3%	$\chi^2(1) = 3.7$, p=0.07
users used terms from 1 suggestion	30.6%	37.5%	$\chi^2(1) = 1.39$, p=0.12
users used terms from several suggestions	57.5%	38.8%	$\chi^2(1) = 10.5$, p=6e-4**
suggestions were clicked	41.4%	24.3%	$\chi^2(1) = 9.3$, p=0.002**

After iterations with suggestions, participants used, in average, 1.34 terms from the suggestions. Considering only the new terms, that is, the terms that did not belong to the previous query, this value falls to 0.66. In the analysis by iteration (Table 2), we see that the number of suggestions' terms, new or not, used in the subsequent query is higher in earlier iterations. This is in line with what was described above.

Table 2. Means of terms used by iteration. One sided significant differences.

	It2	It3	It2 vs It3
termsUsed	1.89	0.76	t(302.2)=8.7, p=2.2e-16**
newTermsUsed	0.82	0.49	t(301)=3.1, p=0.002**

Users employ all terms from suggestions in 56.7% of the iterations where suggestions were presented. The above tendency is still true, that is, complete suggestions are used more in the initial iterations as shown in Table 1.

If, instead of entire suggestions, we consider clicks, the proportion of iterations with clicked suggestions falls to 33.1% and is also significantly higher in the initial iterations as shown in Table 1. Still regarding clicks, we found that 54.7% of the sessions had at least one click and 8.8% of the sessions had two clicks. We found that a large proportion of users (87.5%) have clicked at least once in the proposed suggestions. A lower proportion of users (27.5%) have clicked on suggestions in the two iterations where they were presented.

As shown in Table 3, users extract the larger number of terms from Portuguese/medico-scientific suggestions. In terms of significant differences, we found that the mean number of terms extracted from this type of suggestions is larger than the mean number of terms extracted from English suggestions (Tukey's adjusted p=0 for EN/Lay; Tukey's adjusted p=1.7e-6 for EN/MS). Moreover, as can be seen in Table 4, Portuguese/Lay suggestions are also preferred to both types of English suggestions. Regarding the use of new terms, the English/medico-scientific suggestions are the ones with the greatest contribution to the expansion of terms in users' queries. This difference is explained by users' lack of habit to begin their searches with an English query. English/medico-scientific suggestions are also the ones with a higher proportion of clicks. On the other hand, Portuguese/Lay suggestions are the ones with lower proportion of clicks and lower proportion of new terms, showing that users consider these suggestions the least useful ones. In fact, this type of suggestions has a significantly lower mean of new terms and a significantly lower proportion of clicks with respect to all the other types of suggestions (Table 4).

Table 3. termsUsed and newTermsUsed: mean and standard deviation (SD) by type of suggestion. Proportion of clicks by type of suggestion. Boldface indicates the maximum per column.

	termsUsed		newTermsUsed		Clicks
	Mean	SD	Mean	SD	Proportion
EN/Lay	0.24	0.38	0.16	0.33	14%
EN/MS	0.30	0.41	**0.19**	0.37	**18%**
PT/Lay	0.40	0.40	0.08	0.20	5%
PT/MS	**0.46**	0.43	0.16	0.33	14%

Table 4. Tukey's adjusted p-value for one-sided significant comparisons of the termsUsed and newTermsUsed. Holm adjusted p-value for one-sided significant comparison of proportions of clicks.

PT/Lay vs:	PT/MS	EN/Lay	EN/MS
termsUsed ($>$)	-	5e-6	0.01
newTermsUsed ($<$)	0.002	0.002	1e-5
clicks ($<$)	0.009	0.009	8e-6

6.1 Analysis by Language

In terms of CEFR classes, 16 users had *basic* English proficiency, 17 were *independent*, and 7 were *proficient* users. In Table 5, we present an analysis of the users' preferred language, in general and by English proficiency. If we consider all the used terms, we can see that *basic* and *independent* English proficiency users prefer to use terms from Portuguese suggestions. In the *proficient* group, users tend to prefer English suggestions but this is not a significant difference. If we only consider the newly introduced terms or clicks, users preferred English suggestions. In the *independent* group, the users also tended to prefer English suggestions but not significantly.

In addition to testing the differences between languages, we also tested the differences between levels of English proficiency. We found that *proficient* users clicked on English suggestions more often than *independent* users (Tukey's adjusted p = 0.012*). *Proficient* users are also associated with a higher mean of *termsUsed* than *independent* (Tukey's adjusted p = 0.001**) and *basic* users (Tukey's adjusted p = 0.006**). Excluding the previously used terms (*newTermsUsed*), *proficient* users employ more terms from English suggestions than *independent* users (Tukey's adjusted p = 0.01*).

6.2 Analysis by Terminology

The distribution of the users among the health literacy (HL) classes was as follows: *low* (7 users), *marginal* (28 users) and *functional* (5 users). In Table 6 we can observe that, in general, the users preferred medico-scientific suggestions to lay suggestions. However, if we consider the users' health literacy, we see that this statement only holds for the *marginal* and *functional* groups.

In addition to the differences presented in Table 6, we also found that the use of terms from medico-scientific suggestions significantly increases with the health literacy of the users. In fact, the *low* health literacy group uses fewer terms than the *marginal* (Tukey's adjusted p = 0.01*) and *functional* literacy group (Tukey's adjusted p = 5e-04**). In addition, the *marginal* health literacy group uses fewer terms than the *functional* one (Tukey's adjusted p = 0.05*). Regarding clicks, the *functional* group clicked more often on medico-scientific suggestions

Table 5. Means, proportions and one-sided significant differences of termsUsed, new-TermsUsed and clicks by language and English proficiency (EP). Boldface identifies the row's maximum.

	EN	PT	English vs Portuguese	
			test value	p value
termsUsed				
All	0.27	**0.43**	t(1214.7)=-6.9	3.3e-12**
Basic EP users	0.25	**0.45**	t(483.5)=-5.38	6e-06**
Independent EP users	0.24	**0.44**	t(511.4)=-6.0	2.4e-09**
newTermsUsed				
All	**0.18**	0.12	t(1153.7)=3.2	5e-04**
Basic EP users	**0.18**	0.12	t(455.8)=2.1	0.01*
English proficient users	**0.25**	0.10	t(171.7)=3.3	5e-04**
clicks				
All	**15.9%**	9.7%	$\chi^2(1) = 10.1$	7.6e-04**
Basic EP users	**16.8%**	8.2%	$\chi^2(1) = 0.15$	0.003**
English proficient users	**23.6%**	13.2%	$\chi^2(1) = 1.5$	0.04*

than did the *low* (Tukey's adjusted p =0.013**) and *marginal* groups (Tukey's adjusted p = 0.008**).

The users' familiarity with each topic depends on the theme of the task. The pairs "user, topic" were distributed as follows: *not familiar* (86 pairs), *familiar* (114 pairs) and *extremely familiar* (120 pairs). In Table 6, we can see that, with one exception, users significantly prefer medico-scientific suggestions, disregarding their familiarity with the topic. The only exception occurs among users *extremely familiar* with the topic where we could not find a significant difference between the terminologies. When comparing the several levels of familiarity, we found that users *extremely familiar* with a topic use more newTerms from lay suggestions than *non-familiar* users (Tukey's adjusted p = 0.05*).

7 Discussion

We found that suggestions were used more often at the beginning of the search sessions. In the initial iterations, users not only click more often but they also use more terms from suggestions. Almost 55% of the sessions with suggestions had at least one click and almost 87% had a query with, at least, one of the suggestions' terms. Of the 40 participants in the study, 5 did not click on any suggestion during their tasks. These findings indicate a good acceptance rate of the suggestions, similar to that found by Jansen and McNeese [4]. In the initial iterations, users tend to found suggestions useful more often, but this difference is not significant.

The suggestions formulated in the Portuguese language and lay terminology had a smaller proportion of clicks than other types of suggestions. The average number of new terms extracted from these suggestions is also lower. This indicates that the users found these suggestions to be the least useful, which is not

Table 6. Means, proportions and one-sided significant differences of termsUsed and newTermsUsed and clicks by terminology, health literacy (HL) and topic familiarity. Boldface identifies the row's maximum.

	Lay	MS	Lay vs Medico-scientific	
			test value	p value
termsUsed				
All	0.14	**0.46**	$t(275.1)=-7.0$	9e-10**
Marginal HL	0.14	**0.47**	$t(204.6)=-6.3$	7e-08**
Functional HL	0.06	**0.70**	$t(27.1)=-5.8$	1.7e-06**
Not familiar	0.12	**0.51**	$t(136.4) = -6.3$	1.8e-09**
Familiar	0.16	**0.48**	$t(55.3)=-3.0$	0.001**
Extremely familiar	0.16	**0.36**	$t(80.8)=-2.3$	0.01*
newTermsUsed				
All	0.09	**0.32**	$t(247.2)=-5.5$	9.5e-06**
Marginal HL	0.07	**0.32**	$t(176)=-5.0$	6e-05**
Functional HL	0.06	**0.48**	$t(24.5)=-3.3$	0.001**
Not familiar	0.00	**0.31**	$t(27) = -5.3$	7e-04**
Familiar	0.15	**0.35**	$t(57)=-2.1$	0.02*
Extremely familiar	0.14	**0.30**	$t(79.9)=-1.8$	0.03*
clicks				
All	5.3%	**24.3%**	$\chi^2(1) = 20.4$	3e-06**
Marginal HL	4.5%	**21.4%**	$\chi^2(1) = 12.8$	2e-04**
Functional HL	5.6%	**55.5%**	$\chi^2(1) = 8.4$	0.002**
Not familiar	2.6%	**22.4%**	$\chi^2(1) = 11.8$	3e-04**
Familiar	9.4%	**37.5%**	$\chi^2(1) = 5.6$	0.009**

surprising considering that queries formulated by lay people without assistance will, most probably, use their native language and lay terminology. In this experiment, all the queries in the first iteration were formulated with Portuguese language and lay terminology.

Users from the lowest and highest levels of English proficiency clicked more often on English suggestions and extract a larger number of new terms from English suggestions than from Portuguese ones. These findings regarding *basic* proficiency users surprised us because we expected that suggestions in a language in which they were not proficient would not attract them. This might have occurred because these suggestions had a great degree of novelty, which may have aroused their curiosity. Ignoring the terms used in previous queries, both types of Portuguese suggestions are preferred to both types of English suggestions. Yet, this only happens in the lower levels of English proficiency, advanced proficiency users tend to prefer terms from English suggestions. As expected, the preference for English suggestions is more notorious in proficient users than in *independent* users, in clicks and in number of extracted terms.

In general, the users preferred medico-scientific suggestions to lay suggestions. However, if we consider the users' health literacy, we see that this is true only in the *marginal* and *functional* groups. The *functional* group exhibited a larger proportion of clicks on medico-scientific suggestions compared with the *low* and

marginal groups. Topic familiarity was not found to be a discriminating factor in terms of lay versus medico-scientific terminology use.

8   Conclusion

In this study we assessed users' acceptance to health query suggestions proposed in two languages, Portuguese and English, using two types of terminologies, lay and medico-scientific. In this analysis we also considered if and how users' English proficiency, health literacy and topic familiarity affect their preference. The usage analysis takes into account the utility of the suggestions as new whole queries and as sources of terms.

Suggestions were found to be used more often and, when contributing to query expansion, in a larger quantity, in the initial stages of a search session. In general, suggestions had a good acceptance by the users and the novelty aspect seems to be important in the choice of suggestions to use. Excluding the scenario in which terms from suggestions are used, useful to assess the quality of suggestions' terms but not so useful to assess their utility to the users, English suggestions tend to be preferred to the ones in Portuguese in all levels of English proficiency, a significant preference in the *basic* and *proficient* users. On the other hand, medico-scientific suggestions tend to be preferred to lay ones in the higher levels of health literacy and the extraction of new terms from these suggestions increases with health literacy.

The good acceptance of the suggestions is an indicator of their quality, as perceived by users during query formulation. Nonetheless, as future work, we will analyse the impact of suggestions in retrieval effectiveness through users' relevance assessments.

References

1. Fattahi, R., Wilson, C.S., Cole, F.: An alternative approach to natural language query expansion in search engines: Text analysis of non-topical terms in web documents. Inf. Process. Manage. 44(4), 1503–1516 (Jul 2008), `http://dx.doi.org/10.1016/j.ipm.2007.09.009`
2. Fox, S.: Health topics. Tech. rep., Pew Internet & American Life Project (2011)
3. Gao, W., Niu, C., Nie, J.Y., Zhou, M., Wong, K.F., Hon, H.W.: Exploiting query logs for cross-lingual query suggestions. ACM Trans. Inf. Syst. 28(2), 1–33 (2010), `http://dx.doi.org/10.1145/1740592.1740594`

4. Jansen, B.J., McNeese, M.D.: Evaluating the effectiveness of and patterns of interactions with automated searching assistance. J. Am. Soc. Inf. Sci. 56(14), 1480–1503 (2005), `http://dx.doi.org/10.1002/asi.20242`
5. Kogan, S., Zeng, Q., Ash, N., Greenes, R.A.: Problems and challenges in patient information retrieval: a descriptive study. In: Proceedings AMIA Symposium. pp. 329–333 (2001), `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243602/`
6. Kriewel, S., Fuhr, N.: Evaluation of an adaptive search suggestion system. In: 32nd European Conference on Information Retrieval Research (ECIR 2010). pp. 544–555. Springer (2010)
7. Lopes, C.T., Ribeiro, C.: Measuring the value of health query translation: An analysis by user language proficiency. Journal of the American Society for Information Science and Technology 64(5), 951–963 (May 2013), `http://dx.doi.org/10.1002/asi.22812`
8. Lopes, C.T., Ribeiro, C.: Effects of Terminology on Health Queries: An Analysis by User's Health Literacy and Topic Familiarity, vol. 39, chap. 10, pp. 145–184. Emerald Group Publishing Limited (2015), `http://www.emeraldinsight.com/doi/abs/10.1108/S0065-283020150000039013`
9. Luo, G., Tang, C.: On iterative intelligent medical search. In: SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. pp. 3–10. ACM, New York, NY, USA (2008), `http://dx.doi.org/10.1145/1390334.1390338`
10. Luo, G., Tang, C., Yang, H., Wei, X.: MedSearch: a specialized search engine for medical information retrieval. In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge mining. pp. 143–152. ACM, New York, NY, USA (2008), `http://dx.doi.org/10.1145/1458082.1458104`
11. McCray, A.T., Tse, T.: Understanding search failures in consumer health information systems. In: AMIA Annual Symposium Proceedings. pp. 430–434 (2003), `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479930/`
12. NLM: 2012AA consumer health vocabulary source information (2012), `http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/CHV/index.html`
13. Rawson, K.A., Gunstad, J., Hughes, J., Spitznagel, M.B.B., Potter, V., Waechter, D., Rosneck, J.: The METER: a brief, self-administered measure of health literacy. Journal of general internal medicine 25(1), 67–71 (Jan 2010), `http://dx.doi.org/10.1007/s11606-009-1158-7`
14. Toms, E.G., Latter, C.: How consumers search for health information. Health informatics journal 13(3), 223–235 (Sep 2007), `http://dx.doi.org/10.1177/1460458207079901`
15. Zarro, M., Lin, X.: Using social tags and controlled vocabularies as filters for searching and browsing: A health science experiment. In: The Fifth Workshop on Human-Computer Interaction and Information Retrieval (Oct 2011)
16. Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., Dibble, E.: Assisting consumer health information retrieval with query recommendations. Journal of the American Medical Informatics Association : JAMIA 13(1), 80–90 (2006), `http://dx.doi.org/10.1197/jamia.m1820`
17. Zhang, Y.: Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In: Proceedings of the 1st ACM International Health Informatics Symposium. pp. 210–219 (2010)
18. Zielstorff, R.: Controlled vocabularies for consumer health. Journal of Biomedical Informatics 36(4-5), 326–333 (Oct 2003), `http://dx.doi.org/10.1016/j.jbi.2003.09.015`