Improving Health Consumer Search with Contextual Information

Alexander Beloborodov
Ural Federal University
Russian Federation
xander-beloborodov@ya.ru

Lorraine Goeuriot
LIG - Université Grenoble Alpes
France
lorraine.goeuriot@imag.fr

ABSTRACT

Online search is a widely used way for patients to find information about their condition. Studies have shown that search engines can fail at finding appropriate results [1], and that their use can lead to anxiety [8]. We claim that personalization can help retrieving documents that are better suited to searchers' needs, and therefore relieve the anxiety. This paper reports a preliminary study on health personalized search, using medical reports as a personalization mean.

Keywords

Personalized IR, health consumer search, query expansion

1. INTRODUCTION

The amount of online medical information is constantly increasing. Surveys conducted by the Pew Research Center show that in 2012 in the US, 72% of the internet users have looked online for health information within the past year [1] Getting the right information for users is crucial, as it can have a direct outcome on their health [8]. One of the challenges faced by layperson in the medical domain relies in the understanding of documents. In the US for instance, patients leaving the hospital are given a report describing their condition, called a discharge summary. While given to the patients, this document is not specifically written to be understood by the patient. We focus in this paper on patients health search and claim that personalizing search results can help finding better results. We report a preliminary investigation on how information related to the patient can provide valuable contextual knowledge to improve search results. Given patients reports, we explore how these can be used in order to refine health search.

The remaining of this paper is structured as follows. In Section 2 we will describe the related work. Our approach for contextualization medical IR will be presented in Section 3. In Section 4 we will give the results of the given approach, and give conclusions in Section 5.

2. RELATED WORK

We are considering in this paper the personalization as the use of the users context to get a better understand of her/his needs.

[1]http://www.pewinternet.org/fact-sheets/health-fact-sheet/

As explained in [5], the concept of context in IR is multidimensional. They distinguish: device context (brought from the user's device), spatio-temporal context, and user context as the highest level dimensions. User context is itself divided into personal context and social context. Treating contextual information obviously depends on the dimension it relates to, which implies that contextual information is a very wide topic.

Personalized IR relies on classical IR approaches, the user's context can be integrated at various levels: before retrieval, by modifying the query; or during or after retrieval, by reranking for instance. Sieg et al (2004) are reformulating queries in order to better fit the user profile [6]. Speretta et al (2005) re-rank the top documents issued by a query with information extracted from the user profile [7]. Teufel et al (2001) [10] focus on biomedical articles search, that would be relevant to specific patients cases. They use the patient medical reports to re-rank the results of a classical IR tool.

In this preliminary study, we investigate two methods to conduct personalized health search using the patient's medical profile: simple expansion of the query and run fusion.

3. APPROACH

Our approach aim is to bring patient's contextual information into the usual medical search workflow. To do so, we consider the use case where the patient profile is given by a medical report.

Query expansion.

Searching information which regards her/his medical condition, layperson intuitively keeps in mind some knowledge about her-/himself like her/his age, gender, previous illnesses and so on. Classic retrieval systems normally are not aware of such personal information during the retrieval process, so that the search results might be unsatisfactory. Our main idea is to provide a retrieval system with a layperson context to make the search results more suitable for that concrete user. We use the query expansion (QE) technique to supplement user queries with her/his medical condition information. Normally such kind of information are not available in public access. In our experiments we used medical reports provided by the CLEF eHealth 2014 IR benchmark (which is described in section 4.1).

The key point of our approach is adding a few types of contextual information (for example age, gender, chief complaint) presented by various fields in a user medical profile to an initial user query. Due to computational complexity we experimented with adding only one or two fields. Each term of an additional part of a query was equipped with a weight. We tuned the weights independently for each type of information to be added which is discussed in details in section 4.3.

Re-ranking.

We cannot use all information about patient we have, to expand her/his query, because it leads to excessive query inflation and corresponding drop in performance. One possible solution is using the re-ranking technique. This method joins the result list we want to improve with few other result lists, then recalculates the resulted documents' retrieval scores, and, finally, reorder the mixture according to the new scores. For example, we could expand a layperson query in two ways: (1) with her/his age[2] and gender; (2) with her/his chief complaint, then send the two expanded queries into the retrieval system and get result lists accordingly. Finally we could apply the re-ranking procedure to the two result lists and get one result list which is influenced not only by the patient's age and gender but also by her/his chief complaint. Actually a number of lists to fuse are not limited so that we could bring even more context data into health search than could be brought by the QE only or the 2-lists fusing.

In the current stage of research we experimented with the 2-lists fusing only. To fuse two result lists of size 1000 together we exploited the method Comb-SUM introduced in [9]. It computes the new score of each document d as a weighted sum of the normalized document scores using parameter α:

$$score_{SUM}(d) = \alpha * score_n(d, L_1) + (1-\alpha) * score_n(d, L_2) \quad (1)$$

where $score_n(d, L_i)$ is the Min-Max normalized retrieval score for the document d in the result list L_i. The weight α denotes the relative importance of the first list to fuse over the second one, assuming that $0 \leq \alpha \leq 1$. Then the fusion process result is the top 1000 documents re-ranked according to the $score_{SUM}(d)$. We tuned the α parameter to optimize the NDCG@10 score.

4. EXPERIMENTS

In order to validate our approach, we conducted some experiments using the CLEF eHealth Information Retrieval task benchmark. We describe in this section the evaluation dataset, the settings and results of our experiments.

4.1 Evaluation Dataset

The CLEF eHealth 2014 IR task dataset [2], composed of a document collection, topics, medical reports from which the topics are built, and manual relevance assessment for each topic.

The document collection is a large crawl of about one million web documents. This collection consists of web pages covering a broad range of health topics, targeted at both the general public and healthcare professionals.

The medical reports are used as scenarios to build the topics. They are discharge summaries (i.e. reports given to patients when they are discharged from the hospital), and are extracted from a subset of the MIMIC II dataset [3]. The discharge summaries are semi-structured reports, containing information about the patient condition when leaving the hospital (e.g. demographics, medical history, current treatment, etc.). An example is given in Figure 1.

[2]As explained in Section 4.2, the age is mapped to age categories.

```
Admission Date:  [**2014-03-28**]
Discharge Date:   [**2014-04-08**]
Date of Birth:  [**1930-09-21**]
Sex:   F
Service: CARDIOTHORACIC
Allergies:
Patient recorded as having No Known Allergies to Drugs

Attending:[**Attending Info 565**]
Chief Complaint: Chest pain
Major Surgical or Invasive Procedure:
Coronary artery bypass graft 4.
History of Present Illness:
83 year-old woman, patient of Dr. [**First Name4
(NamePattern1) **] [**Last Name (NamePattern1) 5005**],
Dr. [**First Name (STitle) 5804**] [**Name (STitle)
2275**], with increased SOB with activity, left shoulder
blade/back pain at rest, + MIBI, referred for cardiac
cath. This pleasant 83 year-old patient notes becoming
SOB when walking up hills or inclines about one year
ago. This SOB has progressively worsened and she is now
SOB when walking [**01-19**] city block (flat surface).
[...]

Past Medical History:
arthritis; carpal tunnel; shingles right arm 2000;
needs right knee replacement; left knee replacement
in [**2010**]; thyroidectomy 1978; cholecystectomy
[**1981**]; hysterectomy 2001; h/o LGIB 2000-2001
after taking baby ASA; 81 QOD
[...]
```

Figure 1: Example of a discharge summary

The topics are enriched queries, expressing the patients information need. They have been constructed manually by medical professionals from a given discharge summary, and more specifically from the main diagnosis given in the report. A total of 50 topics are distributed by the task, in a standard TREC format, comprising a topic *title* (text of the query), a *description* (longer description of what the query means), a *narrative* (expected content of the relevant documents) and a patient *profile* (relevant information on the patient identified in the discharge summary), as shown in the Figure 2.

4.2 Experiments Settings

The document collection is indexed with Terrier search engine [4], using Porter Stemmer. As a weighting model for document retrieval we use language model with Bayesian smoothing using Dirichlet priors with parameter $\mu = 2400$. This configuration achieves the best relevance score among all other weighting models available in Terrier.

The 50 topics provided by the CLEF eHealth benchmark and corresponding patients' discharge summaries have been used in the experiments. Using the discharge summaries structure we automatically extracted the patient's `age`, `gender`, chief complaint (`compl`), past surgical procedures (`proc`) to expand their queries. Some types of the information like

```
<query>
<title>
  thrombocytopenia treatment corticosteroids length
</title>
<desc>
  How long should be the corticosteroids treatment
  to cure thrombocytopenia?
</desc>
<narr>
  Documents should contain information about
  treatments of thrombocytopenia, and especially
  corticosteroids. It should describe the treatment,
  its duration and how the disease is cured using
  it.
  <scenario>
    The patient has a short-term disease, or has
    been hospitalised after an accident (little to
    no knowledge of the disorder, short-term
    treatment)
  </scenario>
  <profile> Professional female </profile>
</narr>
</query>
```

Figure 2: Example of a topic

age or gender cannot be used in the query expansion without a preprocessing. For instance, we could know that a patient's age is 55. We cannot just use the term '55' to expand her/his query directly, we need to clarify for the retrieval system that we are adding the age information. So that we assigned different age categories expressed by words to different age intervals. In the example above we would use the words *adult, middleaged* for the age of 55. Another example: if a patient's age is between 4 and 10 we use the terms *child, kid*.

Along with the fields extracted from patients' medical reports we used the topic description field (`desc`) as one more source for the query expansion to check the hypothesis that a short query often does not explain clearly a seeker information need especially in the medical domain. We suppose that the query description by design may clarify the user intention. We did not make any processing of fields `desc`, `compl`, `proc` before query expansion (example of a query expanded with `age` and `desc` might be seen in figure 3).

4.3 Results

To choose the weights maximizing the relevance scores we ran the retrieval process multiple times assigning different weighting values to each type of the contextual information. For example, suppose a patient is a female, she is 55 years old, her/his query is *gastrointestinal bleed*. Then the expanded query might be looking like `gastrointestinal bleed adult^0.1 middleaged^0.1 female^0.2 woman^0.2`. Here one can see that the different weights 0.1 and 0.2 are applying to different types of the patient contextual information: age and gender accordingly. Enumerating the weights from 0.1 to 0.9 with the step 0.05 we optimized the metrics P@10 and NDCG@10 so that we got the best weight assignments for each pair of metric and expansion method.

Table 1 contains the P@10 and NDCG@10 scores achieved after the weights tuning. Each row in the table shows scores averaged over all 50 queries provided by the CLEF eHealth task, where each query is expanded by an appropriate method. *Baseline* row shows the retrieval score for initial queries, i.e. queries without any expansion. For saving space reasons the results which outperform the baseline are shown only.

QE Data ID	w1	w2	P@10	w1	w2	NDCG@10
`age-desc`	0.2	0.15	**0.790**	0.3	0.3	**0.733**
`desc-gender`	0.5	0.15	**0.790**	0.3	0.15	**0.730**
`age-gender`	0.1	0.2	**0.788**	0.1	0.2	**0.722**
`desc`	0.3		**0.778**	0.3		**0.721**
`age`	0.15		**0.782**	0.3		**0.716**
`compl-desc`	0.15	0.4	**0.770**	0.35	0.6	**0.715**
`gender`	0.1		**0.780**	0.1		**0.714**
`desc-proc`	0.3	0.2	**0.782**	0.3	0.2	**0.709**
`age-compl`	0.15	0.1	0.764	0.25	0.1	**0.709**
`age-proc`	0.2	0.15	**0.786**	0.25	0.15	**0.707**
baseline			*0.766*			*0.704*
best team			*0.756*			*0.744*

Table 1: Query expansion results (only the items outperforming the *baseline* are shown). The rows are ordered by decreasing NDCG@10. Best team indicates the results obtained by the best team at CLEF eHealth 2014

As can be seen from the table 1, queries expanded with patient's age and query description fields give the best score according to the both metrics. Figure 3 shows the example of a query improved.

One can observe that out of 10 expansion methods which are better than the baseline, 5 methods involve the query description. Also 5 and 3 methods involve the patient's age and gender accordingly which implies that these types of personal information are useful in the health search.

Query	NDCG@10
chronic duodenal ulcer	0.6706
chronic duodenal ulcer senior^0.3 older^0.3 How^0.3 common^0.3 is^0.3 it^0.3 that^0.3 the^0.3 ulcer^0.3 starts^0.3 to^0.3 bleed^0.3 again^0.3	0.9537

Figure 3: Example of a query which NDCG@10 score has been improved after expansion with both the patient's age and the query description

The `proc` and `compl` performance is reasonable: our manual investigation shows that only about 50 percents of discharge summaries contain meaningful information in these fields.

In the second part of the experiments we tried to improve the query expansion relevance scores shown in the table 1 by fusing together different QE result lists. A QE result list here means the retrieval system output, i.e. top 1000 pages resulted from the expanded query with their retrieval scores. To perform the re-ranking we fuse together two QE result lists into one list by the technique described in the section 3. In the α parameter tuning stage we focused on the NDCG@10 metric maximization.

Having the result list of the `age-desc` QE method which shows the best performance over all other QE methods we aimed at improving its relevance score by bringing into the retrieval process an additional patient contextual information by means of the other QE methods. Implementing this strategy we mixed the `age-desc` result list having the weight α with each other QE result list having the weight $1-\alpha$ separately. In each case we tuned the α parameter from 0 to 1 with step 0.01 to maximize NDCG@10 score. Table 2 shows the five best results.

QE Data ID	α	$1-\alpha$	NDCG@10
`gender`	0.99	0.01	0.7332
`proc`	0.97	0.03	0.7328
`gender-proc`	0.99	0.01	0.7328
`compl`	0.98	0.02	0.7325
`compl-gender`	0.98	0.02	0.7323

Table 2: The `age-desc` QE result list being fused with the other ones (only 5 best results are shown to save space). The rows are ordered by decreasing NDCG@10

There are no fusions performing better than the `age-desc` QE method. It can be seen from the table than the α value is always in close proximity to 1.0 which means that all QE result lists being fused with the `age-desc` one don't have real impact on the result.

The motivation behind the fusion technique is to provide the best QE result list with new relevant documents which might be found in the results influenced by other aspects of user profile. It seems that such an approach is too coarse to improve the QE method because it potentially brings too many irrelevant documents into the final result list. From this perspective we probably should fuse together only well performing QE result lists to get high performance score. So, this issue has to be analyzed in more detailed manner and additional investigation is required.

5. CONCLUSION

This paper describes a first step towards health search personalization through the means of medical records. In particular, we investigated how health information retrieval could be improved by adding contextual information. Intuitively, it seems obvious that the medical profile of a patient is of great importance while assessing the relevance of an information for this patients' queries. Moreover, such a system could be used by medical institutions to provide their patients with adapted and understandable material on their condition. We show that simple contextualization approaches can bring an improvement. Further analysis of the results will be required to understand this phenomenon, and determine which aspects of the profile are of importance. The re-ranking approach which was implemented by means of QE result lists fusion technique should be rethought. Additional experiments comparing the fusion technique with other re-ranking methods should be conducted.

6. ACKNOWLEDGEMENT

This research was supported in part by an ELIAS exchange grant, provided by the European Science Foundation, and in part by the Russian Foundation for Basic Research, project #14-07-00589 "Data Analysis and User Modelling in Narrow-Domain Social Media". The authors also appreciate Philippe Mulhem and Nawal Ould Amer for the fruitful discussion and for their great assistance with the re-ranking technique.

7. REFERENCES

[1] Zuccon G., Koopman B., Palotti J. : Diagnose This If You Can - On the effectiveness of search engines in finding medical self-diagnosis information. ECIR 2015

[2] Goeuriot L., Kelly L., Li W., Palotti J., Pecina P., Zuccon G., Hanbury A., Jones G., Müller H. : ShARe/CLEF eHealth Evaluation Lab 2014, Task 3: User-centred health information retrieval. In Proc. of CLEF 2014 online working notes. 2014.

[3] Saeed M., Villarroel M., Reisner A., Clifford G., Lehman L.-W., Moody G., Heldt T., Kyaw T., Moody B., Mark R. : Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. Critical Care Medicine, 39(5), 952–960, 2011.

[4] Macdonald C., McCreadie R., Santos R., Ounis I. : From puppy to maturity: Experiences in developing Terrier. In Proceedings of the Workshop in Open Source in Information Retrieval at SIGIR, 60–63. 2012.

[5] Tamine-Lechani L., Boughanem M., Daoud M. : Evaluation of Contextual Information Retrieval Effectiveness: Overview of Issues and Research. In Knowledge Information Systems, 24:1–34, 2010.

[6] Sieg A., Mobasher B., Burke R. : User's information context: integrating user profiles and concept hierarchies. In Proceedings of the 2004 meeting of the international federation of classification societies.

[7] Speretta M., Gauch S. : Personalized search based on user search histories. In Proceedings of the IEEE/WIC/ACM international conference on web intelligence, 622–628, 2005.

[8] White R., Horvitz E. : Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search. In ACM Transactions on Information Systems, 27(4), 2009.

[9] Shaw J., Fox E. : Combination of multiple searches. In the TREC-2, 243-252, 1994

[10] Teufel S., Hatzivassiloglou V., McKeown K., Jordan D., Dunn K., Sigelman S., Kushniruk A. : Personalizing retrieval of journal articles for patient care.In Proc AMIA Symp. 2001, 696-700.