Aggregating Evidence from Hospital Departments to Improve Medical Records Search

Nut Limsopatham[1], Craig Macdonald[2], and Iadh Ounis[2]

School of Computing Science
University of Glasgow
G12 8QQ, Glasgow, UK
nutli@dcs.gla.ac.uk[1], firstname.lastname@glasgow.ac.uk[2]

Abstract. Searching medical records is challenging due to their inherent implicit knowledge – such knowledge may be known by medical practitioners, but it is hidden from an information retrieval (IR) system. For example, it is intuitive for a medical practitioner to assert that patients with heart disease are likely to have records from the hospital's cardiology department. In this paper, we propose to capture such implicit knowledge, by grouping aggregates of medical records from individual hospital departments, which we refer to as department-level evidence, to enhance a medical records search engine. Specifically, we propose two approaches to build the department-level evidence based on a federated search and a voting paradigm, respectively. In addition, we introduce an extended voting technique that can leverage this department-level evidence while ranking. We evaluate the retrieval effectiveness of our approaches using the TREC 2011 Medical Records track. Our results show that modelling department-level evidence in medical records search improves retrieval effectiveness. In particular, our proposed voting-based technique obtains results comparable to the best submitted TREC 2011 systems without requiring any of the external resources that are exploited in those systems.

1 Introduction

Government-led initiatives worldwide have digitised the medical records of patient visits to healthcare providers, resulting in electronic medical records (EMRs) [1, 2]. These initiatives have generated a large volume of EMRs, which could aid healthcare practitioners in identifying effective treatments for patients showing particular symptoms [3, 4]. For example, when a doctor compiles a list of possible treatments for patients with skin cancer, it would be advantageous to be able to search for patients who were admitted to a hospital with that disease. However, the accuracy of such a search system is crucial, since the consequences of an error can be an incorrect assessment of the efficacy of a treatment in a population or an inappropriate recommendation for a patient [4].

To foster research on the searching of medical records, the Text REtrieval Conference (TREC) initiated the Medical Records track [5] in 2011 to facilitate the evaluation of EMRs search tools. In particular, the TREC Medical Records track uses the NLP Repository corpus of medical records provided by the University of Pittsburgh[1]. This corpus provides anonymised medical histories of patients throughout their visits to a

[1] http://www.dbmi.pitt.edu/nlpfront

```
<report>
<type>ECHO</type>
<subtype>TEE</subtype>
<admit_diagnosis> 414.12</admit_diagnosis>
...
<report_text>
... (report text here) ...
</report_text>
</report>
```

Fig. 1. An example of a transesophageal echocardiography medical record, from the cardiology department.

hospital, including their detailed EMRs from various hospital departments. As illustrated in Figure 1, the EMRs are semi-structured documents containing hospital and medical information of a patient issued during his visits to the hospital, such as the issuing department information (*type* and *subtype* tags), codes identifying admission diagnosis – in the form of International Classification of Diseases codes (*admit_diagnosis* tag), and a textual description of the patient made by the clinician (*report_text* tag).

Previous works have shown that an effective retrieval does not only depend on the occurrences of query terms in the medical records [6, 7]. In particular, one of the major challenges is the implicit knowledge known among healthcare practitioners, but hidden from an information retrieval (IR) system. For example, when searching for patients suffering from heart disease, experienced healthcare practitioners would go directly to the medical records from the cardiology department, since the medical records of the patients with a heart disease are more likely to be issued from the cardiology department than from other hospital departments. In this paper, we propose to explicitly make available to an IR system some of this implicit knowledge, by exploiting insights gained from aggregates of medical records. We argue that department-level evidence built from aggregates of medical records from particular departments can be used to capture some useful evidence for an IR system, assuming this information is available. Indeed, given that a particular hospital department specialises in a specific group of medical conditions (e.g. the cardiology department specialises in heart diseases), the speciality or expertise of a given department can be inferred by examining all the medical records from that department in aggregate. For each query, we propose to weight the importance of each hospital department for the query by considering the medical records created by the department. In particular, we leverage this evidence to prioritise the medical records that were created by the departments whose expertise is relevant to the query. In this paper, we form the department-level evidence from the list of medical records that share the same *type* and *subtype* tags shown in Figure 1. This department-level evidence is used to give higher importance to medical records from the hospital departments that specialise in the medical condition(s) stated in a query. For example, for a query about heart disease, higher importance is given to medical records from the cardiology department. We hypothesise that the modelling and use of the department-level evidence by an EMRs search system will lead to enhanced retrieval performance.

The contributions of this paper are threefold: (1) We propose to group medical records from the same hospital department to make some of the implicit knowledge found in EMRs explicit to the IR system. In particular, we propose two approaches to build the department-level evidence to represent the department's medical expertise.

The first technique, inspired by work in federated search, models the department-level evidence using the CORI database selection algorithm [8, 9]. Specifically, we propose to model aggregates of medical records from particular departments as different database resources, and use the database scores to estimate given departments' expertise towards the query. The second approach builds upon a voting paradigm [10], which estimates the department's expertise based on the relevance scores of its corresponding medical records; (2) To rank patients for a given query, we introduce an extended voting technique that takes into account the department-level evidence, thereby allowing a search system to focus on medical records issued from particular hospital departments; (3) We thoroughly evaluate our proposed approaches in the context of the TREC 2011 Medical Records track. Our results show the potential of exploiting the department-level evidence to enhance retrieval effectiveness. Moreover, we show that the proposed approach to leverage department-level evidence built using a voting technique leads to an effective result comparable with the best performing systems in TREC 2011 without requiring any external resources.

The remainder of this paper is organised as follows. Section 2 discusses related work and positions our paper in the literature. Section 3 introduces our proposed voting technique that could leverage department-level evidence while ranking patients. Sections 4 and 5 discuss our approaches to build the department-level evidence within a voting paradigm and a federated search, respectively. Section 6 discusses our experimental setup. We empirically evaluate our proposed approaches in Section 7. Finally, Section 8 provides concluding remarks.

2 Related Work

Traditional IR approaches use terms in documents to represent the *aboutness* of the documents. However, attempts have been made to effectively exploit the structure of documents while ranking [11, 12]. For example, Robertson et al. [11] extended the BM25 weighting model to combine scores from weighted fields of documents. Similarly, Plachouras and Ounis [13] introduced randomness models based on multinomial distributions to consider the structure of documents for retrieval. In contrast, our work does not propose a ranking function for term weighting in structured documents. Instead, we focus on exploiting the inherent inter-document structure that medical records exhibit, due to the fact that they are authored by different hospital departments.

To rank documents, search engines traditionally use only the terms occurring within a document, or terms in the anchor text of the document's incoming hyperlinks, in the context of Web search, to rank documents. However, recent works [14, 15] suggested that by aggregating evidence across all of the documents within a host or domain, the impact of incomplete document-level evidence can be reduced. In particular, Metzler et al. [15] aggregated the anchor text for all documents within a host, to permit enriched textual representations for all of the documents within that host. Later, Broder et al. [14] created both host- and document-level indices, from which scores were combined to improve effectiveness in Web search. Sharing the same paradigm as these prior works but operating in a different domain-specific application, we propose to aggregate medical records from the same hospital departments, which are identified by particular tags in the structured medical records, to create a useful representation of department-level evidence that can be used to capture some of the implicit knowledge found in EMRs.

Medical records search in the context of the TREC 2011 Medical Records track [5] aims to find patients having a medical history relevant to the query, based upon these patients' medical records. In particular, a medical records search system ranks patients with respect to the relevance of their medical records towards a query. In this paper, we propose to handle medical records search using well-established approaches previously developed for expert search [16], since both tasks share the same paradigm where the goal is to rank people (i.e. patients or expert persons) based on the relevance of their associated documents. Indeed, in expert search, the aim of the task is to rank experts based on the relevance of the documents they have written, or which mention them [10, 16]. The most effective approaches in expert search use ranked documents to rank expert persons (e.g. Voting Model [10] and Model 2 [17]). Specifically, the Voting Model sees expert search as a voting process, where the ranking of documents (denoted $R(Q)$) defines votes for expert persons to be retrieved. Each document retrieved in $R(Q)$ is said to vote for the relevance of its associated candidate expert using a voting technique (e.g. CombMAX, CombMNZ, expCombMNZ). Indeed, each voting technique firstly ranks documents based on their relevance towards a query using a traditional weighting model (e.g. BM25, DPH from the Divergence from Randomness framework [18]), and then aggregates the votes from documents to experts, to create a ranked list of expert persons related to the query [10]. The voting techniques devised for expert search can also be applied in the medical records search. In our case, the ranking of documents $R(Q)$ is a ranking of medical records which are associated with patients instead of expert persons. Building upon the Voting Model, we propose to apply a voting technique to exploit the expertise of a hospital department from its aggregate of medical records. Moreover, we introduce an extended voting technique that takes into account the department-level evidence when ranking patients.

Another area of research relevant to this work is federated or distributed information retrieval (IR) [8, 19]. Federated IR has been studied to deal with situations where information is distributed across multiple uncooperative search *databases* and a search system aims to rank documents from these databases based on their relevance to a query. In particular, federated search is concerned with three majors problems: resource description, resource selection, and results merging [8]. Firstly, resource description focuses on representing the contents of each resource. Secondly, resource selection aims to make a decision on which resources to be searched, given an information need and a collection of resources. Finally, results merging integrates the ranking results returned by each resource into a final rank list. Approaches such as, CORI [8], ReDDE [20], and CRCS [21] have been proposed to handle such problems. In this work, we apply the federated IR paradigm in a different way. Specifically, we adapt the CORI database selection algorithm to build department-level evidence as an estimate of a particular department's expertise for a given topic represented by a query.

Next, we will present our extended voting technique, called *expCombMNZw*, that takes into account department-level evidence while ranking patients in Section 3, and propose two approaches to build department-level evidence from aggregates of medical records issued by particular hospital departments, within a voting paradigm and a federated search in Sections 4 and 5, respectively.

3 An Extended Voting Technique for Department-Level Evidence

First, we introduce our *expCombMNZw* voting technique, which is an extension of the expCombMNZ [10] voting technique from the Voting Model. This voting technique has performed effectively in various aggregate ranking tasks [10]; however, we extend it to be more effective by allowing setting different weights on particular EMRs. Hence, the expCombMNZw can take into account the expertise of the department of each medical record (i.e. department-level evidence) to focus on EMRs from hospital departments having medical expertise relevant to the query when ranking patients.

In particular, we define $profile(p)$ to be the set of EMRs associated to a patient p, while $R(Q)$ is a ranking of all EMRs with respect to query Q. As each patient is represented by an aggregate of the associated medical records, and each medical record retrieved in $R(Q)$ is said to vote for the relevance of its associated patient. Hence, our proposed expCombMNZw voting technique scores a patient p with respect to a query Q as:

$$score\_patient_{expCombMNZw}(p, Q) = \left[|R(Q) \cap profile(p)| \cdot \sum_{d \in R(Q) \cap profile(p)} w(d, Q) \cdot e^{score(d,Q)}\right] \tag{1}$$

where $R(Q) \cap profile(p)$ is the set of medical records associated to the patient p that are also in the ranking $R(Q)$; $|R(Q) \cap profile(p)|$ is the number of EMRs in the set; and $score(d, Q)$ is the relevance score of medical record d for query Q, as obtained from a standard weighting model.

Within Equation (1), we draw attention to the addition of $w(d, Q)$ to expCombMNZ [10], which permits different weights for different EMRs but not so powerful as the relevance score of EMRs ($score(d, Q)$). In particular, we use $w(d, Q)$ to put emphasis on EMRs associated with particular hospital departments that are relevant to query Q, as follows:

$$dep = department(d) \tag{2}$$

$$w(d, Q) = 1 + (\lambda \cdot score\_department(dep, Q)) \tag{3}$$

where $department(d)$ returns the department dep that issues medical record d, and λ is a parameter controlling the importance of department-level evidence weighting ($\lambda \geq 0$, where $\lambda = 0$ disables the department-level evidence). The relevance of a department dep towards a query Q, $score\_department(dep, Q)$, allows the expCombMNZw to focus on medical records from particular hospital departments whose department-level evidence relevant to the query Q.

In Sections 4 and 5, we will propose two approaches from a voting paradigm and a federated search to obtain department-level evidence and estimate the relevance score of a department ($score\_department(dep, Q)$) in Equation (3).

4 A Voting Approach for Modelling Department-Level Evidence

Within the voting paradigm [10], we introduce our first approach to represent the inherent implicit knowledge in the form of department-level evidence. In particular, we propose to aggregate the medical records from each hospital department to capture some of the implicit knowledge about the expertise of that department. This implicit knowledge may not be available in a traditional IR system, since such knowledge is not explicitly stated in a single medical record, but could be captured from the aggregates of medical records issued by particular hospital departments using a voting technique. Indeed, we hypothesise that these implicit insights about the hospital departments' expertise are useful for improving the retrieval performance.

Specifically, we build department-level evidence by using the medical records associated to individual departments. Figure 2 shows examples of the structure of medical records from hospital departments. For instance, the department-level evidence of the

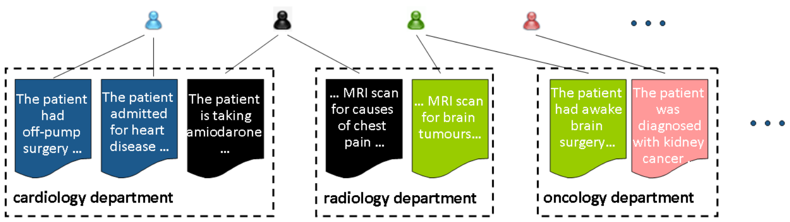

Fig. 2. Examples of medical records from hospital departments.

cardiology department contains all the medical records issued by that department. This permits the IR system a high-level view of each hospital department's expertise that could not be captured in an individual medical record. For example, the expertise of the cardiology department captured in the department-level evidence may encompass evidence of its expertise in heart disease, heart failure, valvular disease, or off-pump surgery. Hence, the IR system can infer that a medical record from the cardiology department has at least a small probability to be about a heart condition.

This department-level evidence is used to estimate the relevance of a hospital department's expertise towards a query Q based on the relevance score of aggregates of medical records from the individual hospital departments. As department-level evidence is represented by aggregates of their associated medical records, a voting technique from the Voting Model [10] can be used to effectively rank departments with respect to a query. We define $profile(dep)$ to be the set of medical records associated to the hospital department dep, while $R(Q)$ is a ranking of all medical records with respect to query Q. Following the Voting Model approach [10], each medical record retrieved in $R(Q)$ is said to vote for the relevance of its associated department. Any approach to rank aggregates could be deployed to rank departments; however, we deploy the expCombMNZ voting technique since it has been shown to perform effectively in various aggregate ranking tasks [10]. In particular, the expCombMNZ calculates the relevance score according to Equation (4):

$$score\_department_{expCombMNZ}(dep, Q) = \left[|R(Q) \cap profile(dep)| \cdot \sum_{d \in R(Q) \cap profile(dep)} e^{score(d,Q)} \right] \tag{4}$$

where $R(Q) \cap profile(dep)$ is the set of medical records associated to the department dep that are also in the ranking $R(Q)$; $|R(Q) \cap profile(dep)|$ is the number of medical records in the set; and $score(d, Q)$ is the relevance score of medical record d for query Q, as obtained from a standard weighting model.

This relevance score of a hospital department is further used by the expCombMNZw voting technique (Equations (1) and (3)) introduced in Section 3 to highly weight medical records from particular hospital departments when estimating the relevance of a patient towards a query.

5 A Federated Search Approach for Modelling Department-Level Evidence

Our second approach to extract department-level evidence is inspired by the work on a federated search of Callan [8]. We propose that federated search techniques could be directly deployed to rank hospital departments when representing them as databases of associated medical records. Specifically, to model the department-level evidence inherent to medical records, we represent each database (i.e. resource) by the occurring terms and their frequencies found in the medical records of the same hospital department. In particular, we build an index (i.e. a database) for the set of medical records from each hospital department. For instance, the database representing the cardiology department contains statistics of terms occurring in EMRs issued from this department. This may allow each database to represent the expertise of the corresponding hospital department. For example, the EMRs of patients having symptoms or treatments related to heart diseases are issued by the cardiology department, as shown in Figure 2.

Classical federated search [8] includes a typical uncooperative environment of databases, requiring the use of a query-based sampling technique to create a representation of each resource. Instead, we do not apply such a query-based sampling technique, since we only focus on leveraging a resource selection technique to rank hospital departments based on their issued medical records whereby all the required statistics are readily available. Hence, the simulation of an uncooperative environment is not required. Specifically, we apply the CORI database selection algorithm [8] to calculate the relevance scores of databases (i.e. hospital departments) since it has been shown effective on different federated search tasks [8, 19, 20]. In particular, the relevance score (i.e. belief) $p(t_i|dep)$ of the database representing a hospital department dep, according to a query term t_i is calculated by [8]:

$$T = \frac{df}{df + 50 + 150 \cdot \frac{cw}{avg_{cw}}} \tag{5}$$

$$I = \frac{log|DB| + 0.5}{cf} \tag{6}$$

$$p(t_i|dep) = b + (1 - b) \cdot T \cdot I \tag{7}$$

where df is the number of EMRs in the database representing the hospital department dep that contain term t_i, cf is the number of databases that contain t_i, $|DB|$ is the number of the databases in the collection, cw is the number of terms in database representing department dep, avg_{cw} is the average number of terms among the databases in the collection, and b is the default belief, which is set to 0.4 as recommended in [8].

Next, the beliefs based on each term in a query are combined into the final belief that a database representing department dep is relevant to the query (i.e. the relevance score of the department for the query) using belief operators [22]. In particular, during our experiments, we combine beliefs using SUM, OR, and AND operators, as follows:

$$score\_department_{CORI\_SUM}(dep, Q) = \frac{\sum_{t_i \in Q} p(t_i|dep)}{|Q|} \tag{8}$$

$$score\_department_{CORI\_OR}(dep, Q) = 1 - \prod_{t_i \in Q} (1 - p(t_i|dep)) \tag{9}$$

$$score\_department_{CORI\_AND}(dep, Q) = \prod_{t_i \in Q} (p(t_i|dep)) \tag{10}$$

where $p(t_i|dep)$ is the relevance score (i.e. belief) calculated using Equation (7) and $|Q|$ is the number of query terms.

Generally in federated search systems, 5 or 10 databases with the highest belief scores are selected so that documents will be retrieved only from these databases. However, in our case, we focus on using all the databases' relevance scores to estimate the relevance of hospital departments towards a query. In particular, the expCombMNZw (Equations (1) and (3)) proposed in Section 3 leverages these database relevance scores to take into account the expertise of hospital departments while ranking patients.

Unlike the Voting Model that allows the use of sophisticated scores (e.g. document relevance score) to rank aggregates, a federated search [8] takes into account only term and document frequencies when ranking databases since it is designed to efficiently rank a collection of databases in a distributed retrieval environment. Hence, in this work we do not apply a federated search to rank patients.

6 Experimental Setup

As discussed in Section 3, we propose to model department-level evidence in medical records search could leverage the inherent implicit knowledge within medical records, and hence improve retrieval effectiveness. In particular, we hypothesise that department-level evidence gained from aggregates of medical records issued by particular departments could be used as novel evidence to infer the importance of a patient's medical record to a particular query when searching for relevant patients. To validate our hypothesis, we evaluate our proposed approaches in the context of the TREC 2011 Medical Records track test collection [5]. In this track, the task is to identify relevant patient *visits* for each query topic, where a visit contains all of the medical records associated with a patient's visit to a hospital. A *visit* is used to represent a *patient* as a unit of retrieval since relating multiple visits to a particular patient is made impossible by a de-identification process when building the medical records repository [5]. The TREC medical records collection consists of approximately 102k medical records, which can be mapped to 17,265 patient visits. In addition, using the information of the structure of the collection, we define 328 hospital departments[2]. In particular, Table 1 shows statistical information of the collection of 328 hospital departments.

We index the medical records using the Terrier retrieval platform [23][3], applying Porter's English stemmer and removing stopwords. In all experiments, the DPH document weighting model [18] is used to rank medical records (i.e. $score(d, Q)$). DPH is a parameter-free document weighting model from the Divergence from Randomness (DFR) framework, hence no parameters need to be trained [18]. In addition, the number of medical records in $R(Q)$ to vote for the relevance of departments and patient visits (the representations of patients) is limited to 5,000, as suggested in a prior work [24]. We evaluate our approaches to model department-level evidence while ranking using patient visits the 34 topics from the TREC 2011 Medical Records track; however, with such a small number of topics, the use of a statistical test validation is precluded [25].

[2] We define the department of a medical record automatically using its *type* and *subtype* tags; however, this may allow sub-units of a department to be considered as departments.

[3] `http://terrier.org`

Table 1. Statistics of hospital departments in the collection.

Number of databases (i.e. hospital departments)	328
Minimum number of medical records per database	1
Maximum number of medical records per database	19,769
Average number of medical records in the databases	307.52
Standard deviation of the number of medical records in the databases	1397.44
Minimum number of terms per database	79
Maximum number of terms per database	2,723,596
Average number of terms in the databases	91,609.29
Standard deviation of the number of terms in the databases	332,880.76

We compare the effectiveness of our proposed approaches to exploit the department-level evidence with baselines that do not consider the department-level evidence, in terms of bpref measure [26]. The official measure of the TREC 2011 Medical Records track is bpref, since the absolute number of judged visits per topic is relatively small [5]. In particular, bpref is designed for evaluating environments with incomplete relevance data and penalises a system which ranks a judged non-relevant document above a judged relevant document [26].

7 Experimental Results

To validate our hypothesis that our approaches to leverage the department-level evidence could improve retrieval performance, we compare the bpref retrieval performance of our expCombMNZw voting technique proposed in Section 3 with the baseline applying expCombMNZ [10]. Figure 3 shows the bpref retrieval performance of our proposed approaches to model the department-level evidence within our expCombMNZw voting technique, as we vary λ. The baseline, applying the expCombMNZ alone without considering department-level evidence (i.e. $\lambda = 0$) is shown as a horizontal line. 'expCombMNZw – Voting-based' exploits the department-level evidence built using a voting-based approach, introduced in Section 4. While 'expCombMNZw – CORI_AND', 'expCombMNZw – CORI_OR', and 'expCombMNZw – CORI_SUM' deploy the CORI database selection approach to create department-level evidence, with AND, OR and SUM operators, proposed in Section 5.

From Figure 3, we observe that our approaches to modelling department-level evidence while ranking patient visits could outperform the baseline. Specifically, 'expCombMNZw – Voting-based' and 'expCombMNZw – CORI_AND' both outperform the baseline. In particular, 'expCombMNZw – Voting-based' with $\lambda = 9.5$ outperforms the baseline for 15/34 topics. In addition, we find that these approaches are robust, as they could outperform the baseline for a wide range of λ. However, the approaches to leverage department-level evidence built using the SUM and OR operators ('expCombMNZw – CORI_SUM' and 'expCombMNZw – CORI_OR') are not as effective. This suggests that as department-level evidence is useful indicator when they are strongly appropriate for all of the medical conditions expressed in the query.

7.1 Query Expansion

Next, as query expansion (QE) techniques have been shown to be effective for the task of ranking patients [5, 24], we further evaluate the effectiveness of department-level evidence after applying a QE technique (namely DFR Bo1 [27]) to expand queries when

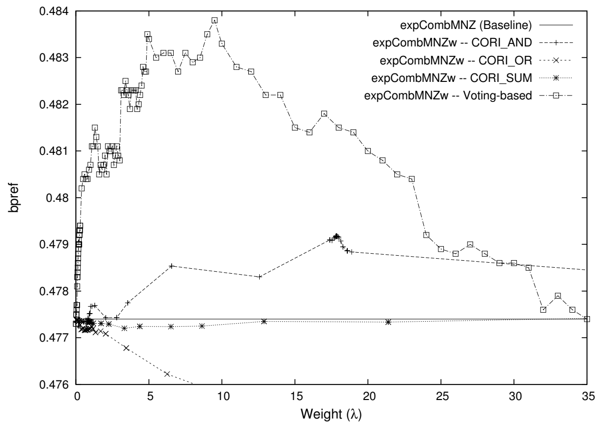

Fig. 3. bpref performance comparing the baseline and our approaches to model department-level evidence, varying λ.

calculating scores for departments and visits. Indeed, QE is applied to reformulate the query based on occurrences of terms in the top retrieved EMRs [28]. We select the top-10 terms from top-3 ranked medical records to expand the query, as suggested in [27].

In Figure 4, the baseline approach, where QE is applied to rank medical records before aggregated using the expCombMNZ without considering the department-level evidence, is shown as a horizontal line. We observe that our proposed expCombMNZw voting approach that leverages department-level evidence obtained using the voting technique improves the retrieval performance over the baseline. Specifically, bpref is improved from 0.5218 to 0.5305. Bo1 QE and our approach to leverage department-level evidence obtained using a voting technique combine effectively, as they bring different levels of evidence to the search system. Indeed, Bo1 QE helps to better estimate the importance of department-level evidence, which results in the highest improvement of retrieval effectiveness. Moreover, our proposed voting-based approach to build and leverage department-level evidence is robust, with large ranges of λ showing improvements. However, we find that the database selection-based approaches do not combine effectively with QE, which aligns with the previous work of Ogilvie and Callan [19].

Overall, we find that our approaches to build department-level evidence from aggregates of medical records bring useful evidence to a search system. This department-level evidence can be modelled to focus on medical records that are more likely to be related to the query while ranking patient visits using our expCombMNZw approach. Indeed, our best bpref of 0.5305 is comparable to the third ranked participating group in the TREC 2011 Medical records track, without the use of any domain-specific ontologies or any of the other external resources (e.g. MeSH[4] and UMLS[5]) deployed by the first three groups. Moreover, this setting achieves an R-precision of 0.4305 and pre-

[4] http://www.nlm.nih.gov/mesh/

[5] http://www.nlm.nih.gov/research/umls/

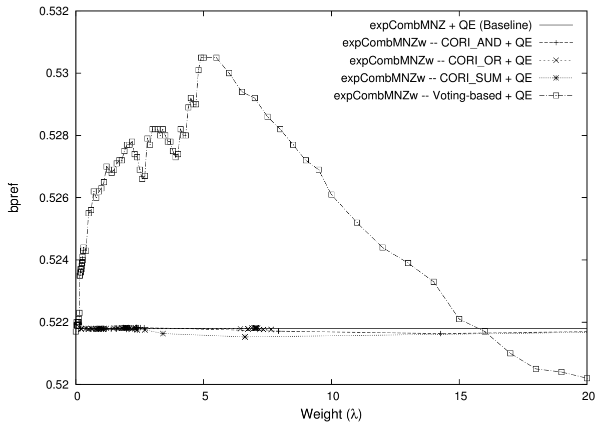

Fig. 4. bpref performance comparing the baseline and our approaches to model department-level evidence, when applying QE, varying λ.

cision@10 of 0.600, which are comparable with the performances of the second ranked participating group from a total of 29 participating groups.

8 Conclusions

We have highlighted the issue of implicit knowledge in medical records search, where the knowledge of healthcare practitioners is hidden from a search system, and proposed a potential alleviation by using the knowledge gained from aggregates of medical records associated to hospital departments (i.e. department-level evidence). In particular, we proposed the extended expCombMNZw voting technique that considers department-level evidence to better weight individual medical record while ranking patients. In addition, we proposed two approaches to build department-level evidence from medical records associated to particular hospital departments, based on a voting paradigm and a federated search, respectively. Our results show the potential of our approaches to leverage department-level evidence, especially our approach to obtain department-level evidence using a voting technique. In particular, the proposed approach can outperform an effective voting approach on the TREC 2011 Medical Records track test collection, and can produce a performance comparable with the top participating TREC groups, without resorting to any external resources, such as ontologies, as used in those systems. In addition, our approach is general, in that it could be used to capture implicit knowledge using different types of corpus structures (e.g. aggregates of medical records having the same diagnosis code). In the future, we plan to investigate how topic modelling (e.g. LDA [29]) can be used to capture the medical evidence obtained from aggregates of medical records issued by specific hospital departments.

References

1. E. Siegel, D. Channin. Integrating the healthcare enterprise: A primer. *RadioGraphics*, 21(5), 2001.
2. E. Tambouris, M. Willimas, C. Makropoulos. Co-operative health information networks in Europe: experience from Greece and Scotland. *Intl. J. Med. Inform.* 64(1), 2000.

3. W. Hersh. Health care information technology: progress and barriers. *J. of the American Medical Association*, 292(18), 2004.
4. W. Hersh. Information retrieval: A health and biomedical perspective (3rd ed.). *New York : Springer*, 2009.
5. E. Voorhees, R. Tong. Overview of the TREC 2011 Medical Records Track. In *Proc. of TREC*, 2011.
6. H. Jain, C. Thao, Z. Huimin. Enhancing electronic medical record retrieval through semantic query expansion. *Info. Systms & E-Business Mngmnt*, 2010.
7. N. Limsopatham, C. Macdonald, R. McCreadie, I. Ounis. Exploiting Term Dependence while Handling Negation in Medical Search. In *Proc. of SIGIR*, 2012.
8. J. Callan. Distributed Information Retrieval, *Advances in Information Retrieval*, Kluwer Academic Publisers, 2000.
9. J. French, A. Powell, J. Callan, C. Viles, T. Emmitt, K. Prey, Y. Mou. Comparing the performance of database selection algorithms. In *Proc. of SIGIR*, 1999.
10. C. Macdonald, I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proc. of CIKM*, 2006.
11. S. Robertson, H. Zaragoza, M. Taylor. Simple BM25 extension to multiple weighted fields. In *Proc. of CIKM*, 2004.
12. A. Trotman. Choosing document structure weights. *IPM*, 41(2):243–264, 2005.
13. V. Plachouras, I. Ounis. Multinomial randomness models for retrieval with document fields. In *Proc. of ECIR*, 2007.
14. A. Broder, E. Gabrilovich, V. Josifovski, G. Mavromatis, D. Metzler, J. Wang. Exploiting site-level information to improve web search. In *Proc. of CIKM*, 2010.
15. D. Metzler, J. Novak, H. Cui, S. Reddy. Building enriched document representations using aggregated anchor text. In *Proc. of SIGIR*, 2009.
16. K. Balog, P. Thomas, N. Craswell, I. Soboroff, P. Bailey. Overview of the TREC 2008 Enterprise Track. In *Proc. of TREC*, 2008.
17. K. Balog, L. Azzopardi, M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proc. of SIGIR*, 2006.
18. G. Amati, E. Ambrosi, M. Bianchi, C. Gaibisso, G. Gambosi. FUB, IASI-CNR and University of Tor Vergata at TREC 2007 Blog Track. In *Proc. of TREC*, 2007.
19. P. Ogilvie, J. Callan. The effectiveness of query expansion for distributed information retrieval. In *Proc. of CIKM*, 2001.
20. L. Si, J. Callan. Using sampled data and regression to merge search engine results In *Proc. of SIGIR*, 2002.
21. M. Shokouhi. Central-rank-based collection selection in uncooperative distributed information retrieval. In *Proc. of ECIR*, 2007.
22. H. Turtle, B. Croft. Efficient probabilistic inference for text retrieval. In *Proc. of RIAO*, 1991.
23. I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, C. Lioma. Terrier: A High Performance and Scalable Information Retrieval Platform. In *Proc. of OSIR at SIGIR*, 2006.
24. N. Limsopatham, C. Macdonald, I. Ounis, G. McDonald, M.M. Bouamrane. University of Glasgow at Medical Records Track 2011: Experiments with Terrier. In *Proc. of TREC*, 2011.
25. E. Voorhees, C. Buckley. The Effect of Topic Set Size on Retrieval Experiment Error. In *Proc. of SIGIR*, 2002.
26. C. Buckley, E. Voorhees. Retrieval Evaluation with Incomplete Information. In *Proc. of SIGIR*, 2004.
27. G. Amati. Probabilistic Models for Information Retrieval based on Divergence from Randomness. *PhD thesis*. University of Glasgow, 2003.
28. C. Macdonald, I. Ounis. Using Relevance Feedback in Expert Search In *Proc. of ECIR*, 2007.
29. D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation In *the Journal of Machine Learning Research*, 3, p.993-1022, 3/1/2003.