University of Glasgow at Medical Records Track 2011: Experiments with Terrier

Nut Limsopatham, Craig Macdonald, Iadh Ounis,
Graham McDonald, and Matt-Mouley Bouamrane

terrier@dcs.gla.ac.uk

University of Glasgow
Glasgow, UK

ABSTRACT

In our participation in the TREC 2011 Medical Records track, we investigate (1) novel voting-based approaches for identifying relevant patient visits from an aggregate of relevant medical records, (2) the effective handling of negated language in records and queries, and (3) the adoption of medical-domain ontologies for improving the representation of queries, all within the context of our Terrier information retrieval platform.

1. INTRODUCTION

In our participation in the first Medical Records track [18], we aim to enrich the Terrier platform[1] [15] with approaches for searching medical records. In particular, we build upon the effective Divergence from Randomness (DFR) framework [1] and the Voting Model [12] to find patient visits with the relevant medical condition(s) expressed in a query. Indeed, our participation has three major objectives:

- First, we adapt our Voting Model, which has been shown to be effective for aggregate ranking tasks [13, 14], to the task of identifying relevant patient visits. We model medical record retrieval as an aggregate ranking task [13], such that medical records related to the query topic vote for the relevance of their associated patient visits.

- Second, we investigate the effective handling of negated language in the medical records and queries. Indeed, one of the unique characteristics of medical records is the frequent use of negated language [4, 8, 9]. We propose an approach integrating a negation detection technique for differentiating negated text from normal text in both the records and queries. This ensures that records containing query terms in negative contexts are only retrieved for queries addressing negated meanings.

- Third, in order to cope with a vocabulary mismatch resulting from different word preferences among the healthcare providers who wrote the medical records, we explore approaches to exploit multiple medical ontologies (e.g. SNOMED[2]) to identify medical concepts in records and queries, process the concepts as normal tokens, and expand the concepts in a query with concepts that are nearby within the used ontologies.

This paper is organised as follows. In Section 2, the medical search task, as well as the approaches to rank patient visits and handle negation are discussed. Section 3 introduces approaches to employ controlled-terminology dictionaries and ontologies to improve retrieval effectiveness. Runs and results are presented in Section 4, and the conclusions are discussed in Section 5.

2. THE MEDICAL SEARCH TASK

The objective of the medical search task is to identify cohorts, which are groups of patients having the same medical conditions, for comparative effectiveness research [18]. For example, a study on comparing the effectiveness of medical interventions on middle-age women with breast cancer needs to find a cohort of patients with those medical conditions before conducting medical experiments. The queries for this task define inclusion criteria (e.g. personal profiles, or medical conditions) to describe the type of patients required for medical studies. The task is then to find patients with certain medical conditions stated in a query from a corpus containing about 101k de-identified medical records from the University of Pittsburgh NLP Repository[3]. This collection consists of one month of structured reports from multiple hospitals and includes nine types of reports from different departments in those hospitals. The types of reports are Radiology Reports, History and Physicals, Consultation Reports, Emergency Department Reports, Progress Notes, Discharge Summaries, Operative Reports, Surgical Pathology Reports, and Cardiology Reports. Figure 1 shows an example of a medical record. By using the mapping table provided with the repository, each medical record or document can be mapped to one of 17,265 patient visits. A patient visit is an individual stay at a hospital by a patient, and may contain many different medical records. Relating multiple patient visits to a single patient is made impossible as a result of the de-identification process. Therefore, the task uses *a patient visit* as a representative of a patient.

2.1 The Voting Model

Our participation in the medical records search task builds upon our Voting Model [12]. Instead of directly combining

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

[2] `http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html`

[3] `http://www.dbmi.pitt.edu/nlp/report-repository`

```
<report>
<checksum>
20060205PGN-V7c6v8UJPyhL-848-1663275873
</checksum>
<subtype>CCM ATTEND</subtype>
<type>PGN</type>
<chief_complaint>COPD</chief_complaint>
<admit_diagnosis>496</admit_diagnosis>
<discharge_diagnosis>
496,785.51,518.5,998.11,511.9,996.84,
998.12,414.01,416.8,401.9,482.83,
</discharge_diagnosis>
<year>2007</year>
<downlaod_time>2009-10-05</downlaod_time>
<update_time/>
<deid>v.6.22.08.0</deid>
<report_text>
... (report text here) ...
</report_text>
</report>
```

Figure 1: Example of a medical record in the corpus.

medical records of a particular visit into a single 'document' representing the visit and ranking these visit documents directly with respect to the query Q, we propose to model patient visit retrieval as an aggregate ranking task by adapting the Voting Model to rank visits based on their associated medical records. In particular, we retrieve a set of medical records with respect to query Q, which we denote as $R(Q)$. Every time a medical record is retrieved in this result set, this is considered to be a vote for the visit associated to the record to have a relevant medical condition to Q. Finally, the votes for each visit are aggregated to form a ranking of patient visits taking into account the relevance score of the voting records. We apply the expCombSUM voting technique [12] to rank visits, which considers the sum of the exponential of the relevance scores of the medical records associated to each visit. In expCombSUM, the score of a visit V to a query Q is given by:

$$score\_visit(V,Q) = \sum_{r \in R(Q) \cap profile(V)} w(r) \cdot e^{score(r,Q)} \quad (1)$$

where $R(Q) \cap profile(V)$ is the set of medical records associated to the visit V that are also in the ranking $R(Q)$, $score(r,Q)$ is the relevance score of medical record r for query Q, and $w(r)$ is a function that permits the higher weighting of important records (e.g. medical records from a particular department may be more important than the others) . An example of the function is shown in Section 3.3.

2.2 The Record Ranking Model

In order to retrieve medical records that will vote for a patient visit, we apply the DPH [2] hypergeometric parameter-free document weighting model. DPH is a weighting model from the Divergence from Randomness (DFR) [1] framework. DPH calculates the score for a medical record r as follows:

$$\begin{aligned} score_{DPH}(r,Q) = \sum_{t \in Q} & tfq \cdot norm \cdot (tfr \\ & \cdot \log((tfr \cdot avg\_rl/rl) \cdot (N/tfc)) \\ & + 0.5 \cdot \log(2 \cdot \pi \cdot tfr \cdot (1-f))) \end{aligned} \quad (2)$$

where tfq is the frequency of term t in the query Q, tfr is the frequency of term t in the record r, tfc is the frequency of term t in the collection, avg_rl is the average length of medical records in the collection, rl is the length of record r, N is the number of medical records in the collection, $f = tfr/rl$, and $norm = (1-f) * (1-f)/(tf+1)$

2.3 Negation Handling

In order to cope with the frequent use of negated language in medical records, we propose our *NegFlag* approach. NegFlag uses the NegEx algorithm [5] to detect negation at the sentence level of each medical record. For example, in the sentence 'The patient denied *experiencing chest pain on exertion*', the terms in italic have a negative context. In NegFlag, the terms that have a negative context are prefixed with a special character (e.g *n0*). For instance, if 'pain' is in a negative context, it will be represented as 'n0pain'. Table 1 shows an example of how a sentence is transformed using the NegFlag approach. As a result, normal terms and negated terms are differentiated (e.g. 'rash' and 'n0rash'), which enables them to be handled differently during retrieval.

Original Sentence	The patient denied experiencing chest pain on exertion
Negated Terms	experiencing chest pain on exertion
Transformed Terms	the patient denied n0experiencing n0chest n0pain n0on n0exertion

Table 1: Handling negation using the NegFlag approach.

3. RECORD & QUERY EXPANSION

The frequent use of synonyms within medical records can possibly cause a vocabulary mismatch. However, there exist several ontologies for the medical domain, which could be used to identify relationships between associated terms (e.g. synonyms). During retrieval, we make use of medical domain ontologies (e.g. SNOMED) to alleviate the vocabulary mismatch between the terms in a query and medical records, and thereby to improve retrieval performance. Inspired by previous works [6, 7, 19, 20] exploiting domain-specific ontologies to address retrieval tasks in the Genomics Track, we employ medical ontologies in three manners: the enrichment of patient visits (Section 3.1), the expansion of concepts for a given query (Section 3.2), and inferring the importance of report types in medical records and queries (Section 3.3). Lastly, we also make use of age and gender information in medical records and queries to improve retrieval performance (Section 3.4).

3.1 Visit Enrichment with ICD Codes and the Wikipedia Pages

The International Classification of Diseases (ICD) codes[4] in the 'admit_diagnosis' section of each medical record provide general information about the existing medical condition of its patient. For example, in Figure 1 the 'admit_diagnosis' field shows the ICD code of a disease/symptom that the patient was diagnosed with. The description of the

[4] http://www.who.int/classifications/icd/en/

code '496' is 'chronic airway obstruction, not elsewhere classified' and it has a dedicated Wikipedia page[5]. Using these ICD codes, we aim to enrich each patient visit with the description and the Wikipedia pages of the ICD codes of the medical records associated to the visit. In particular, during indexing we add information containing the description and Wikipedia pages to the visit.

3.2 Concept Representation & Expansion

Medical ontologies and controlled-vocabulary dictionaries offer an alternative way to view terms in queries and medical records as concepts. A concept is a group of noun phrases that refer to a particular meaning. For example, according to the MeSH[6] controlled vocabulary, 'mad cow disease' is a name of a particular kind of disease. Our approach makes use of the ontologies to identify concepts in records and queries. Intuitively, searchers in the medical domain look for records about a particular medical concept (i.e. mad cow disease) rather than the records that contain all terms of the concept but where the terms are separated in different parts of the records. Therefore, handling groups of terms as a single concept unit could enhance retrieval effectiveness [3, 16]. However, traditional 'bag-of-words' approaches do not consider the order of the terms of a concept. Therefore, records containing all the terms of the concept but not having the meaning of the concept may be highly ranked. In order to cope with the issue, we treat terms of a concept as a single concept unit during indexing and retrieval. For instance, query 'mad cow disease' is tokenised to 'mad', 'cow', 'disease', and 'concept_mad_cow_disease'. This increases the likelihood that the records containing the concept meaning are ranked highly.

In addition, we also expand the concepts identified in a query with nearby concepts in the ontologies, namely synonyms and hyponyms. In particular, we expand a concept occurring in a query (a trigger concept) with its synonyms (candidate expansion concepts) in the ontologies and controlled-vocabulary dictionaries. Inspired by recent work on acronym weighting [11], we weight an expanded synonym by using the co-occurrence value of the query concept and the synonym in the corpus. Indeed, we use the EMIM (Expected Mutual Information Measure) [17] to measure the level of dependence between concepts, by the distribution of their co-occurrences in the University of Pittsburgh NLP Repository collection. EMIM is calculated as [11]:

$$\mathrm{EMIM}(tr, ce) = \log \frac{\Pr(tr, ce)}{\Pr(tr)\Pr(ce)} \tag{3}$$

where tr is a trigger concept, and ce is a candidate expansion concept. Pr is the maximum likelihood estimation function (i.e. the probability that a medical record contains a concept tr), while $\Pr(tr, ce)$ is the joint probability of tr and ce, estimated as the fraction of records where they co-occur.

The calculated EMIM of each trigger and candidate expansion pair is integrated into the retrieval score of a medical record for a query as follows:

$$\begin{aligned} score\,(r, Q) \;=\; & \sum_{t \in Q} score(r, t) + \qquad (4) \\ & \lambda_S \sum_{tr \in matches(Q)} \sum_{ce \in synonyms(tr)} \Big[\\ & EMIM\,(tr, ce) \cdot score(r, ce) \Big] \end{aligned}$$

where λ_S is a parameter to weight the score of the expanded concepts. $matches(Q)$ uses the medical resources (e.g. MeSH, SNOMED) to identify concepts occurring in the query Q, where each identified concept is called a trigger concept (and denoted tr); $synonyms(tr)$ uses the medical resources to define a set of synonyms for trigger concept tr - each synonym is added to the query as a candidate expansion, denoted ce. We set λ_S to 0.3, based on our experiment to test the effective value of λ_S on the training set of topics provided by TREC.

Moreover, we follow the work of Li et al. [10] to expand concepts with their hyponyms in the ontologies. Indeed, the approach is to estimate the similarity of the concepts based on the distance between the original concept and the expanding concept. The similarity between two concepts is calculated as follows:

$$Sim(tr, ce) = e^{-\alpha \cdot l(tr,ce)} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} \tag{5}$$

where tr is a trigger concept and ce is a candidate expansion concept, $l(tr, ce)$ returns the distance (i.e. the number of nodes) between tr and ce in the ontology hierarchy, and h is the depth of the ontology hierarchy; α and β are parameters. In our participation, we set $\alpha = 0.2$ and $\beta = 0.6$ as recommended in [10].

The hyponym similarity is integrated with the medical record retrieval score as follows:

$$\begin{aligned} score\,(r, Q) \;=\; & \sum_{t \in Q} score(r, t) + \qquad (6) \\ & \lambda_H \sum_{tr \in matches(Q)} \sum_{ce \in hyponyms(tr)} \Big[\\ & Sim(tr, ce) \cdot score(r, ce) \Big] \end{aligned}$$

where λ_H is a parameter to weight the score of the hyponym expanding concepts. As before, $matches(Q)$ identifies the trigger concepts (tr) occurring in the query; $hyponyms(tr)$ uses the ontology hierarchy to the hyponyms of tr, which are added to the query as candidate expansions (denoted ce). In this work, based on our experiment to test the effective value of λ_H on the training set of topics, we set λ_H to 0.5.

In our participation, we apply both Equations (4) and (6) to expand a query with both synonym and hyponym concepts associated to the concepts in the queries.

3.3 Inferring on Report Types

We found that the medical records of a patient with a particular disease often belong to a particular hospital department (i.e. report type). For example, patients with heart disease frequently have cardiology reports from the cardiology department. Therefore, we first apply our concept representation technique to identify medical concepts in medical records and queries (described in Section 3.2) and

[5]http://en.wikipedia.org/wiki/Chronic_airway_obstruction
[6]http://www.ncbi.nlm.nih.gov/mesh

then use the identified query concepts to promote certain types of medical records within the voting technique. For instance, queries mentioning bone breakages are unlikely to be concerned with cardiology reports. Therefore, the types of records containing concepts within the same MeSH hierarchy as the query concepts should be promoted within the voting technique. Indeed, we create a separate collection where each 'document' corresponds to a subtype. Each subtype document contains medical concepts, in a truncated form, that appear in the medical records of that subtype in the University of Pittsburgh NLP Repository. Figure 2 depicts a sample subtype document, for a radiology report of subtype chest, containing concepts such as lung cancer, in the form of truncated MeSH identifiers (e.g. A1.5, A3.2, and A4.1). The weight of the subtype document is then calculated for integration into expCombSUM (Equation (1)), using the score of the concepts in the original query, as:

$$w(r) = \sum_{c \in matches(Q)} score(subtype(r), c) \tag{7}$$

where $matches(Q)$ identifies all MeSH concepts occurring in query Q, and $subtype(r)$ is the subtype of record r. The $score()$ function measures the similarity of the subtype document representing the subtype of record r (i.e. $subtype(r)$). The score of a particular subtype is based on the score of the associated subtype document for a given c. In this work, we apply the DPH weighting model to score subtype documents.

```
<DOC>
<DOCNO>RAD:CHEST</DOCNO>
A1.5 A3.2 A4.1
A1.3 A3.2 A9.4
</DOC>
```

Figure 2: Example of a subtype document.

3.4 Age/Gender Scoring

Often, medical records contain the age and gender of the associated patients and the inclusion criteria frequently focuses on finding a cohort of patients with certain age range and gender. To meet this requirement, we improve the ability of the retrieval system to identify the target age range and gender of patients specified in the query. We use crowdsourcing to both build a dictionary of terms describing patients of various ages (e.g. elderly, teenager), and to assign age ranges to these terms. In particular, we use workers from the Amazon Mechanical Turk marketplace[7]. Examples of terms and their corresponding age ranges are shown in Table 2. In addition, we define terms to determine the gender of the patient of each visit, which are shown in Table 3.

During indexing, we identify the age ranges in a medical record based on explicitly stated patient information in the record and define the gender of the patient based on the majority occurrences of male and female related terms in the medical records for the visit. If the frequencies of the male and female terms are tied, then the gender of the record is assumed to be undefined.

Then, for a given query, we identify in the query any age range and a gender requirement for the cohort. During retrieval, the medical records of the patients having the gender and age ranges relevant to the query are ranked higher than the medical records of other patients.

[7] https://www.mturk.com/mturk/welcome

Age range	Terms in queries
birth-12	small, young, little
Teenage	snotty, student, energetic
20s	vivacious, college-aged individuals
30s	working class, prime
40s	parent, middle aged, mature
50s	mature, older, wise
60s	older, wise, retired
70s	senior, gray, retired
80s	senior, retired, ancient
More than 90	fragile, senior, elderly

Table 2: Examples of age ranges and their associated terms.

Gender	Terms in queries or records
male	male, he, his, him, gentleman, man
female	female, she, her, hers, lady, woman

Table 3: Patient genders and their associated terms.

4. RUNS AND RESULTS

We perform all runs using the Terrier retrieval platform[8]. During indexing, the medical records from the University of Pittsburgh NLP Repository have standard stopwords removed, and the Porter English stemmer is applied. We submitted 4 automatic runs in our participation in the Medical Records track 2011 search task:

- uogTrMDeNFo: This run tests our NegFlag negation handling, the DPH weighting model (Equation (2)), the expCombSUM voting approach (Equation (1)), and our age/gender scoring (See Sections 2.1-2.3 and 3.4). All other submitted runs build upon this run.
- uogTrDeNIo: This run integrates ICD text and Wikipedia as external resources for visit expansion (See Section 3.1).
- uogTrDeNfCE: This run focuses on dealing with the vocabulary mismatch issue. It employs the synonym and hyponym concept expansion approach using concepts identified in the MeSH, SNOMED and ICD using Equations (4) and (6) (See Section 3.2).
- uogTrDeNSo: This run employs concept expansion using MeSH, and focuses the voting technique on the most important medical record types for each query using Equation (7) (See Section 3.3).

Additionally, we perform five additional runs where the age/gender scoring is not applied to permit the methodical evaluation of the performance of our approaches, as follows:

- uogTrB: The baseline deploys only the DPH weighting model (Equation (2)) to rank medical records and the expCombSUM voting technique (Equation (1)) to identify relevant visits based on the ranked medical records.

[8] http://terrier.org

- uogTrMDeNFoNF: This run deploys our NegFlag approach, the DPH model (Equation (2)) to rank medical records, and the expCombSUM (Equation (1)) to identify relevant visits. The other remaining additional runs are built upon this run.
- uogTrDeNIoNF: This run uses ICD text and Wikipedia as external resources for record expansion (See Section 3.1).
- uogTrDeNfCENF: This run employs the synonym and hyponym expansion approach using concepts identified in the MeSH, SNOMED and ICD using Equations (4) and (6) (See Section 3.2).
- uogTrDeNSoNF: This run deploys the concept expansion using MeSH, and focuses the voting technique on the most important medical record types for each query using Equation (7) (See Section 3.3).

Run	Submitted	bpref	r-prec	P@10
TREC Best (A)	N/A	0.5520	0.4400	0.6560
TREC Median	N/A	0.4120	0.3090	0.4760
uogTrB	✗	0.4773	0.3864	0.5353
uogTrMDeNFoNF	✗	0.4816	0.3901	0.5618
uogTrDeNIoNF	✗	0.4830	0.3882	0.5588
uogTrDeNfCENF	✗	0.4293	0.2779	0.4382
uogTrMDeNSoNF	✗	0.3292	0.1176	0.1765
uogTrMDeNFo	✔	0.4912	**0.4018**	**0.5706**
uogTrDeNIo	✔	**0.4930**	0.4010	0.5676
uogTrDeNfCE	✔	0.4857	0.3881	0.5353
uogTrDeNSo	✔	0.4655	0.3614	0.5382

Table 4: Results of the submitted runs to the Medical Records track.

Table 4 shows the results of our submitted and additional runs compared to the TREC Median. First, we find that all of our submitted runs and uogTrB markedly outperform the TREC Median on all measures. This shows that these approaches are effective for ranking patient visits. Next, comparing the uogTrB baseline with uogTrMDeNFo, we observe that our negation handling approach improves retrieval performance. In addition, our age/gender scoring is also effective for this task, since all the submitted runs outperform their corresponding unsubmitted run without the age/gender scoring (e.g. uogTrMDeNFo vs uogTrMDeNFoNF). Moreover, our approach to integrate ICD descriptions and Wikipedia (uogTrDeNIo) also results in performance improvement over the uogTrMDeNFo run upon which it builds. Therefore, our negation handling, age/gender scoring approaches, and visit enrichment with ICD information are indeed effective for the medical records search task. In contrast, the concept expansion and the inference on report type approaches decrease the retrieval performance. The improper parameter settings (e.g. λ_S, λ_H) may be the reason of this performance. We believe that the retrieval performance would be further improved if more representative training topics were available.

5. CONCLUSIONS

Participating in the Medical Records track of TREC 2011 using our Terrier IR platform, we focused on enhancing the Voting Model to retrieve the relevant visits with medical-related knowledge. In addition, we proposed to handle negation in medical records in Section 2.3, while we defined alternative approaches to semantically expand terms in the queries using external resources in Section 3. As shown in Table 4, our results attest the effectiveness of our enhanced Voting Model and the deployed domain-specific approaches.

6. REFERENCES

[1] Amati, G. *Probability models for information retrieval based on Divergence From Randomness.* PhD thesis, University of Glasgow. 2003.

[2] Amati, G., Ambrosi, E., Bianchi, M., Gaibisso, C., Gambosi, G. FUB, IASI-CNR and University of Tor Vergata at TREC 2007 Blog Track, In *Proc. of TREC*, 2007.

[3] Aronson, A. R. and Rindflesch, T. C. Query expansion using the UMLS Metathesaurus. In *Proc. of AMIA*, 1997.

[4] Averbuch, M., Karson, T., Ben-Ami, B., Maimond, O., and Rokachd, L. Context-sensitive medical information retrieval. In *Proc. of ACL*, 2003.

[5] Chapman, W., Bridewell, W., Hanbury, P., Cooper, G., Buchanan, B. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J. Biomedical Informatics*, 5:301–310, 2001.

[6] Hersh, W., Bhupatiraju, R., Ross, L., Cohen, A., Kraemer, D., and John, P. TREC 2004 Genomics Track Overview. In *Proc. of TREC*, 2004.

[7] Hersh, W., Cohen, A., Yang, J., Bhupatiraju, R. and Hearst, M. TREC 2005 Genomics Track Overview. In *Proc. of TREC*, 2005.

[8] Jain, H., Thao, C., Huimin Z. Enhancing electronic medical record retrieval through semantic query expansion. *Information Systems and E-Business Management*, 2010.

[9] Koopman, B., Bruza, P., Sitbon, L., and Lawley, M. Analysis of the effect of negation on information retrieval of medical data. In *Proc. of ADCS*, 2010.

[10] Li, Y. and Bandar, Z.A. and Mclean, D. An approach for measuring semantic similarity between words using multiple information sources. *Trans. Knowledge and Data Engineering*, 15(4):871–882, 2003.

[11] Limsopatham, N., Santos, R., Macdonald, C., Ounis, I. Disambiguating biomedical acronyms using EMIM. In *Proc. of SIGIR*, 2011.

[12] Macdonald, C. *The Voting Model for People Search.* PhD Thesis, University of Glasgow, 2009.

[13] Macdonald, C., and Ounis, I. Voting for candidates: Adapting data fusion techniques for an expert search task. In *Proc. of CIKM*, 2006.

[14] Macdonald, C., and Ounis, I. Key blog distillation: Ranking aggregates. In *Proc. of CIKM*, 2008.

[15] Ounis, I. and Amati, G. and Plachouras, V. and He, B. and Macdonald, C. and Lioma, C. Terrier: A High Performance and Scalable Information Retrieval Platform. In *Proc. of OSIR at SIGIR*, 2006.

[16] Stokes, N., Li, Y., Cavedon, L., and Zobel, J. Exploring criteria for successful query expansion in the genomic domain. *J. Information Retrieval*, 12(1):17-50, 2009.

[17] van Rijsbergen, C. J. A theoretical basis for the use of co-occurrence data in information retrieval. *J. Documentation*, 33(2):106–199, 1977.

[18] Voorhees, E., and Tong, R. Overview of the TREC 2011 Medical Records Track. In *Proc. of TREC*, 2011.

[19] Zhou, W., Yu, C. and Meng, W. A system for finding biological entities that satisfy certain conditions from texts. In *Proc. of CIKM*, 2008.

[20] Zhou, W., Yu, C., Smalheiser, N., Torvik, V., and Hong, J. Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature. In *Proc. of SIGIR*, 2007.