Methods for Linking EHR Notes to Education Materials

Jiaping Zheng1, Hong Yu, PhD2

1University of Massachusetts, Amherst MA; 2University of Massachusetts Medical School, Worcester MA

Abstract

It has been shown that providing patients with access to their own electronic health records (EHR) can enhance their medical understanding and provide clinically relevant benefits. However, languages that are difficult for non-medical professionals to comprehend are prevalent in the EHR notes, including medical terms, abbreviations, and domain-specific language patterns. Furthermore, limited average health literacy forms a barrier for patients to understand their health condition, impeding their ability to actively participate in managing their health. Therefore, we are developing a system to retrieve EHR note-tailored online consumer-oriented health education materials to improve patients' health knowledge of their own clinical conditions. Our experiments show that queries combining key concepts and other medical concepts present in the EHR notes significantly outperform (more than doubled) a baseline system of using the phrases from topic models.

Introduction

It has been shown that providing patients with access to their own electronic health records (EHR) can enhance medical understanding and provide clinically relevant benefits, including increased medication adherence1. Patients have also expressed interests in accessing their own records2. However, EHR notes present unique challenges to the average patients. Since these notes are not usually targeted at the patients3, languages that may be difficult for non-medical professionals to comprehend are prevalent, including medical terms, abbreviations, and medical domain-specific language patterns.

Limited health literacy forms yet another barrier for the average patients to understand their health conditions, impeding their ability to actively participate in managing their health. It is estimated in the National Assessment of Adult Literacy that the average American has a reading level between the 7th and 8th grade4. It is also reported in the same assessment that about 75 million Americans (36%) have basic or below basic health literacy. Patients with limited health literacy are at a disadvantage to understand written health-related information and make appropriate decisions. For example, health literacy is shown to be associated with health status and health-related knowledge in people living with HIV-AIDS. Limited health literacy prevents the patient from identifying out-of-range laboratory test results5. One study identified poor health literacy as a potential factor in lower cancer screening rates within recommended guidelines6.

To address the aforementioned challenges, we are developing a NoteAid system7 to translate medical jargon to lay language and to link medical notes to targeted easy-to-understand education materials from trusted resources. We speculate that NoteAid has the potential to improve patients' understanding of their own clinical conditions and health knowledge, which in turn can lead to improved self-managed care and clinical outcome. In this study, we report the development and evaluation of systems for linking EHR notes to online consumer-oriented health education materials. We explore Information Retrieval methods to generate queries from the topic phrases and medical concepts in the EHR notes.

Related Work

There is a wealth of work in health literacy and comprehension. The medical jargon, which is prevalent in the EHR notes, is one evident difficulty in patients’ understanding8. Mappings between the medical and consumer terminologies are created to bridge the gap between them9,10. Unsupervised methods are employed to identify difficult terms and definitions are retrieved using commercial search engines11. Tools are developed to substitute difficult terms with easier synonyms, and split long sentences to shorter ones12. Providing definitions of medical jargons is also shown to improve EHR notes’ readability. For instance, our NoteAid system7 identifies medical concepts and fetches definitions from UMLS, Medline Plus, and Wikipedia and evaluation has shown significant improvement in self-reported comprehension.

Improving readability does not fully exploit the benefits of access to EHR notes. High quality information obtained through education materials can potentially lead to better outcomes13. The Patient Clinical Information System14...
provides patients with online information resources and educational aides, and evaluations by patients have been
positive. However, no automated systems have been reported. Infobuttons was developed mainly to assist
physicians, and was not designed for patients. The Infobutton Manager Project15,16 links EHR notes to other
information resources (e.g., drug databases, Google, PubMed, AskHERMES17). However, Infobuttons was
developed mainly to assist physicians, and was not designed for patients. PERSIVAL is another physician-centric
system that accepts user provided queries to retrieve personalized results from a patient care library18. EHR notes are
used to build topics from consumer health texts. Probabilistic topic modeling is also utilized to recommend
education materials to patients with diabetes19. Education materials are ranked according to frequencies of terms and
topics in a given EHR note. The authors show that the top two recommended documents are significantly more
relevant than a randomly selected document from the same domain. Structured data from EHR is also used to
retrieve health-related information20.

Contributions of this work are: 1) we designed approaches to generate effective queries from the long and noisy
EHR data, resulting in more than doubled performance improvement over the baseline, 2) we demonstrated that
identifying concepts in EHR data is important in generating queries, and 3) we built a corpus of EHR note specific
education materials.

Materials and Methods

Education Materials

MedlinePlus (http://www.nlm.nih.gov/medlineplus/) provides current and reliable information about over 900 health
topics to users in consumer-oriented lay language. Additionally, the medical encyclopedia section includes over
4000 articles about diseases, tests, symptoms, injuries, and surgeries. We include in this study the textual narratives
in the “health topics”, “drugs, supplements, and herbal information”, and “medical encyclopedia” sections of the
MedlinePlus as the collection of educational materials. There are a total of approximately 9400 articles in this
collection, which we designate as MedlinePlus.

We index the MedlinePlus documents with Galago21, an advanced open source search engine. Galago implements
the inference network retrieval model22. This model calculates the probability of the user’s information needs being
satisfied given a document in a directed acyclic graph. This framework is applied in many information retrieval
tasks, and shown to be successful23.

Information Retrieval (IR) Systems

We develop two strategies to link EHR notes to external education materials. The first is based on traditional IR in
which we use the entire EHR notes to retrieve relevant education materials. Since EHR text is not patient-oriented,
we also experiment with substituting medical jargons with consumer terms as queries. In our experiments, we return
500 relevant documents for each EHR note.

In the second strategy, we investigate several query generation approaches, whereby short queries are built from an
EHR note to retrieve relevant education materials. We first explore topic detection. Full EHR notes typically discuss
diverse aspects of the patient’s conditions, including diagnoses, medication, procedures, etc. We trained Latent
Dirichlet Allocation (LDA) topic models24 from over 6000 de-identified EHR notes. These EHR notes are
tokenized, and stop words are removed before the models are learned. The resulting models are then used to infer
the topic distribution of each test note.

Traditional LDA models extract distributions over individual word tokens for each topic. However, medical
concepts often contain more than one token. We employ turbo topics25 to find phrases from these topics. This
method builds significant n-grams based on a language model of arbitrary length expressions. To translate the topics
into queries, from each of the inferred topics whose combined probability is over 80%, the top 5 phrases are selected
as query terms. We use this system as our baseline.

To concentrate on medical concepts, we next train another LDA model from the UMLS26 Concept Unique
Identifiers (CUIs) of the medical concepts contained in the EHR notes, disregarding the textual content. The
Corporation provides the preferred terms of the top 5 relevant CUIs from each prominent topic from the test EHR notes are
issued as queries to retrieve relevant documents. To make comparisons to our different approaches, we do not index
our MedlinePlus corpus with concept CUIs.
We also more directly focus on the medical concepts by selecting the top ones based on their inverse document frequency (IDF) from the EHR note corpus we used to learn LDA models. Concepts that appear less often in a large corpus are more unique to the EHR, thus presumably more important for the patient.

Furthermore, we developed a supervised machine learning approach to identify key concepts that need explanation by external education materials. These key concepts can be considered in a broad sense topics, as they also capture various aspects of the EHR note content. A physician independently assigned relevant MedlinePlus documents to 20 EHR notes (details are described in the following Evaluation Data and Metrics section). Training data was generated by the first author who manually annotated phrases (key concepts) in the EHR notes that are also titles of relevant MedlinePlus documents. A Conditional Random Fields (CRF) model was then trained to predict the key concepts. We explored lexical, morphological, word shape, UMLS semantic type, and section information as learning features.

Due to the sparseness of the key concepts, many relevant education documents may be missed. We thus extracted all concepts from an EHR note using MetaMap and constructed a query using all concepts. In order to limit concepts to domain-specific medical terms, we filtered concepts with the following semantic types: acquired abnormality, antibiotic, cell or molecular dysfunction, clinical attribute, diagnostic procedure, disease or syndrome, experimental model of disease, finding, laboratory procedure, laboratory or test result, organ or tissue function, pathologic function, physiologic function, pharmacologic substance, sign or symptom and therapeutic or preventive procedure.

Finally, we evaluated a two-stage design that first retrieves a small number of documents (we empirically set up the threshold to be 20) using the key concepts before complementing them with more documents from the full set of concepts for broader coverage.

**Evaluation Data and Metrics**

Twenty de-identified EHR progress notes are selected to test our systems’ performance. A physician read each note, and manually identified relevant education materials from the MedlinePlus documents. For example, a note about various conditions and symptoms of liver disease is linked to an education document on alcoholic disease to discourage the patient from drinking alcohol. On average, each note is linked to 26.4 MedlinePlus documents. This collection of 20 annotated EHR-note-education-materials is then used as the gold standard for evaluating our NoteAid IR systems.

To evaluate the IR systems, we use the Mean Average Precision (MAP) metric, a common standard in the IR community to evaluate ranked retrieval results. Set-based measures such as precision and recall metrics cannot distinguish the order the results are presented in a ranked retrieval context.

Average Precision (AveP) for each test document is the average of precision at each point where a relevant document is found:

$$AveP(R, D_k) = \frac{\sum_{d_i \in D_k \cap R} P(R, D_i)}{|R|}$$

where $R$ is the gold standard, $D_k$ is the top k retrieved results, $d_i$ is the i-th ranked result in $D_k$, and $D_i$ is the results from 1 to i-th ranked document. $P(R, D_i)$ is the precision score for the $D_i$ documents:

$$P(R, D_i) = \frac{|R \cap D_i|}{|D_i|}$$

Then, for a given set of queries $Q$, MAP can be calculated as

$$MAP(Q) = \frac{\sum_{q \in Q} AveP(R_q, D_{k,q})}{|Q|}$$

where $q$ is a query in $Q$, $R_q$ is the gold standard results of $q$, and $D_{k,q}$ is the top k retrieved results of $q$.

**Experiment Results and Discussion**

As described in the Materials and Methods section, our traditional IR-based approach issues each of the 20 test EHR notes as the query to the MedlinePlus document index and retrieves top 500 relevant documents. Unlike the topics or concepts based systems, this system does not use the sequential dependence model because of the length of the EHR notes. The performance is shown in Table 1.
We found that the top 10 retrieved results of the baseline system for each of the EHR note are nearly identical, with minimal order variations. We found that none of the top-10 retrievals is a true relevant document according to our gold standard. The results are not surprising. EHR notes are written by physicians, containing domain-specific medical jargons. In contrast, consumer-oriented education materials are written in lay language, a different text genre. In addition, the full text of an EHR note may contain noise to the extent that distinguishing content is drowned out. For example, an EHR sentence “I am glad to see Ms. Smith today” provides little information other than the gender of the patient, which may still be identified from other parts of the note. Search engines are not optimized to process queries as long as over 500 tokens, and cannot automatically filter out the noise without significant adaptations. The unique language and style in these medical notes makes the filtering all the more difficult.

To investigate the gap between medical language and lay language, we substituted the medical concepts recognized by MetaMap with their consumer-oriented counterparts created by the Consumer Health Vocabulary (CHV) Initiative. The result (Full text HER with CHV terms) as shown in Table 1 more than doubled. The gap highlights the issue that patients may have difficulty finding relevant health information without assistance.

Table 1. System performance.

| System                        | MAP@500 | Increased Folds |
|-------------------------------|---------|----------------|
| Full text EHR                 | 0.0091  |                |
| Full text EHR with CHV terms  | 0.0240  |                |
| LDA (Baseline)                | 0.0489  | -              |
| LDA on concepts               | 0.0410  | 0.84           |
| IDF-filtered concepts         | 0.0681  | 1.39           |
| Key concepts                  | 0.0921  | 1.88           |
| All concepts                  | 0.0968  | 1.98           |
| Key concepts and then all concepts | 0.1114 | 2.28           |

As described in the Method section, we explored several approaches to generate shorter queries from an EHR note to address the problems of using full notes. In these approaches, except when specified otherwise, sequential dependence model was used to capture the dependencies in a multi-word query term. In this model, given a query, documents are ranked based on features of documents containing a single query term, two query terms sequentially appearing in the query, and two query terms in any order. This model has been shown to be effective in many applications.

Our first method is to learn Latent Dirichlet Allocation (LDA) topic models and generate queries from top terms of the most prominent topics. We employed turbo topics to find phrases from these topics. Three models are learned with 20, 50, and 100 topics. The one with 100 topics performed the best. System result using this model is shown in Table 1 (row LDA). The improvement over the full text method is statistically significant using a paired t-test with p<0.05.

Table 2 shows the top 10 n-grams from 7 topics. It is clear that while topics like the first one capture medical concepts, others like the second one do not. The LDA results also highlight the noisy nature of the EHR notes. Queries formed by including the generic or noisy terms such as “continue on” will not benefit retrieval results. Examining the retrieval results, we found that when the most prominent topics include medical concepts, the top 10 results usually contain at least one relevant document (sometimes as the first result). When only generic topics are identified, relevant documents are absent in the top 10 results. For instance, three true relevant documents (first, second, and fifth in the top 10) are retrieved for an EHR note with topic 1 in Table 2. On the other hand, none of the top 10 results is relevant for a note for which only topic 2 is identified as prominent.

Table 2. Top 10 n-grams from 7 topics using the LDA model.

| Topic ID | Phrases with the highest probability                                                                 |
|----------|-------------------------------------------------------------------------------------------------------|
| 1        | dialysis, hemodialysis, catheter, renal failure, renal, coumadin, line, picc line, dialysis catheter, failure |
| 2        | job id, today, point, continue on, reasonable, try to, continue, yesterday, left, right                |
| 3        | continue, patient, job id, pain, patient has, normal, patient s, white count, secondary to, culture    |
| 4        | liver, ascites, normal, tenderness, fluid, stable, elevated, today, edema, chest                      |
To eliminate generic terms and focus on the medical terms, we trained a LDA model solely from the UMLS Concept Unique Identifiers (CUIs) of the medical concepts contained in the EHR notes. The corresponding phrases of the top 5 relevant CUIs from each prominent topic are issued as queries. Since the CUIs implicitly represent entire terms, sequential dependence model is not used. The number of topics is selected based on the MAP score. We found that learning 100 topics yielded higher score than 20 or 50 topics. The best MAP score of a LDA model is shown in Table 1 (LDA on concepts). It is also statistically significant over the baseline system, using the same test as in the LDA model on text. On the other hand, the LDA model trained from the UMLS concepts did not outperform one trained from the full note. This drop of performance may be in part due to the noise introduced by the MetaMap system. Analysis of the retrieval results show that this model identified 2 more topics per EHR note on average, or 8 more terms in the resulting query than did the textual LDA model. This increase in the number of query terms could dilute the probability assigned to the terms, rendering the query string less effective. In addition, some query concepts, such as Carney’s Syndrome, are not present in our MedlinePlus document collection.

Our manual evaluation shows that LDA is unable to capture the important medical concepts that most need explanations in simple language, as shown in some of the topics in Table 2. We therefore directly selected medical concepts recognized by MetaMap, based on their inverse document frequency (IDF) from a large de-identified EHR corpus. The number of concepts was empirically set to 10. The system performance is shown in Table 1 (IDF-filtered concepts). This represents a nearly 40% increase over the baseline.

Medical concepts are very effective at filtering out the noise in EHR notes. However, high frequency concepts such as “blood pressure” may be erroneously filtered out regardless of its importance in the note. We trained a CRF model from the 20 EHR notes to predict key medical concepts, using phrases that match education material titles. The performance of this CRF model using leave-one-out cross validation is shown in Table 3. System performance using these key phrases as queries is presented in Table 1 (Key concepts). Using the key medical concepts shows a significant improvement over the LDA model learned from UMLS CUIs using t-test. The improvement over the LDA model learned from text is smaller. Although this improvement is not statistically significant, this model achieves nearly double the performance of the LDA model.

| Table 3. Key concept identification performance. |
|-----------------------------------------------|
| Precision                                      |
| Recall                                        |
| F1                                            |
| Average number of key phrases per document    |
| 45.77%                                        |
| 26.51%                                        |
| 31.76%                                        |
| 7.95                                          |

One drawback of this approach is that the identified key phrases fail to cover the scope of the medical concepts contained in an EHR note. First, since the key phrases are rather sparse in the EHR notes, the CRF model cannot learn from enough examples, thus only approximately a quarter of them are extracted. Secondly, the sparseness leads to a low coverage of the entire set of concepts in an EHR note. In fact, on average only 23.6% of the concepts are annotated as key phrases in each note.

Therefore, we processed each note to identify all medical concepts, and generated queries using these concepts. As described in the Method section, the concepts are extracted using MetaMap and only those of the following semantic types are retained: acquired abnormality, antibiotic, cell or molecular dysfunction, clinical attribute, diagnostic procedure, disease or syndrome, experimental model of disease, finding, laboratory procedure, laboratory or test result, organ or tissue function, pathologic function, physiologic function, pharmacologic substance, sign or symptom and therapeutic or preventive procedure. The full set of medical concepts enables the system to achieve a slightly better performance than the limited-coverage key concept approach, retrieving more relevant documents, albeit at low ranks in the retrieval results. The performance is shown in Table 1 (All concepts). This result is similar to the Key concept approach in significance tests.

Combining the ability of the full-set approach to cover more concepts and the advantage of the key phrase approach to identify relevant documents more precisely, we adopt a two-stage approach in which key phrases are first used to
retrieve 20 most relevant documents, and the full set of concepts are then generated as a second query to retrieve more documents. The retrieval results from this query are appended to the results from the first query. System performance is shown in Table 1 (Key concepts and then all concepts). This approach outperforms all the other approaches. The improvement is statistically significant in every case except for the “all concepts” approach.

**Conclusion, Limitations and Future Plan**

It has been shown that providing patients with access to their own EHR can enhance their understanding and provide clinically relevant benefits. However, the difficult language in EHR notes and limited average health literacy present a challenge. To address these problems, we are developing a system to retrieve EHR note-tailored online consumer-oriented health education materials.

In our experiments, we have shown that using the full text of an EHR note is ineffective at retrieving relevant education materials. Identifying key concepts of an EHR note and then using the key concepts as query terms result in significantly improved performance. Furthermore, a two-stage approach in which key concepts are complemented by other medical concepts in the note outperforms other approaches, such as topic models or simple aggregation of medical concepts as queries.

One limitation of our design is that only one physician provided relevancy judgments. Additional annotators would provide a more rigorous set of gold standard, allowing us to measure inter-annotator agreement. Secondly, our education material collection only includes the MedlinePlus documents, which may not cover all the topics that a user needs. Lastly, a larger test EHR note collection could improve the key concept identification performance.

There are several directions we can explore in our future research. First, we plan to optimize the weights in our generated query terms as they are now equally weighted. In addition, we also plan to evaluate whether patients’ comprehension improve when their EHR notes are tailored to education materials. Lastly, we intend to explore methods to integrate both the narrative and structured data in EHR and link them to education materials.

**Acknowledgement**

Research reported in this publication was in part supported by the Health Services Research & Development of the US Department of Veterans Affairs Investigator Initiated Research 1I01HX001457-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US Department of Veterans Affairs.

**References**

1. Delbanco, T. et al. Inviting patients to read their doctors’ notes: a quasi-experimental study and a look ahead. *Ann. Intern. Med.* **157**, 461–470 (2012).
2. Wiljer, D. et al. Getting results for hematology patients through access to the electronic health record. *Can. Oncol. Nurs. J. Rev. Can. Nurs. Oncol.* **16**, 154–164 (2006).
3. Mossanen, M. et al. Surgical pathology and the patient: a systematic review evaluating the primary audience of pathology reports. *Hum. Pathol.* (2014). doi:10.1016/j.humpath.2014.07.008
4. National Assessment of Adult Literacy (NAAL). at <http://nces.ed.gov/naal/>
5. Zikmund-Fisher, B. J., Exe, N. L. & Witteman, H. O. Numeracy and literacy independently predict patients’ ability to identify out-of-range test results. *J. Med. Internet Res.* **16**, e187 (2014).
6. Oldach, B. R. & Katz, M. L. Health literacy and cancer screening: A systematic review. *Patient Educ. Couns.* **94**, 149–157 (2014).
7. Polepalli Ramesh, B., Houston, T., Brandt, C., Fang, H. & Yu, H. Improving patients’ electronic health record comprehension with NoteAid. *Stud. Health Technol. Inform.* **192**, 714–718 (2013).
8. Keselman, A. et al. Towards consumer-friendly PHRs: patients’ experience with reviewing their health records. *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* 399–403 (2007).
9. Zeng, Q. T. & Tse, T. Exploring and developing consumer health vocabularies. *J. Am. Med. Inform. Assoc. JAMIA* **13**, 24–29 (2006).
10. Smith, B. & Fellbaum, C. Medical WordNet: A New Methodology for the Construction and Validation of Information Resources for Consumer Health. in *Proceedings of the 20th International Conference on Computational Linguistics* (Association for Computational Linguistics, 2004). doi:10.3115/1220355.1220409
11. Elhadad, N. Comprehending technical texts: predicting and defining unfamiliar terms. *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* 239–243 (2006).
12. Kandula, S., Curtis, D. & Zeng-Treitler, Q. A Semantic and Syntactic Text Simplification Tool for Health Content. *AMIA Annu. Symp. Proc.* **2010**, 366–370 (2010).

13. Bodenheimer, T., Lorig, K., Holman, H. & Grumbach, K. Patient self-management of chronic disease in primary care. *JAMA* **288**, 2469–2475 (2002).

14. Cimino, J. J., Patel, V. L. & Kushniruk, A. W. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. *Int. J. Med. Inf.* **68**, 113–127 (2002).

15. Cimino, J. J. Use, usability, usefulness, and impact of an infobutton manager. *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* 151–155 (2006).

16. Cimino, J. J., Elhanan, G. & Zeng, Q. Supporting infobuttons with terminological knowledge. *Proc. Conf. Am. Med. Inform. Assoc. AMIA Annu. Fall Symp. AMIA Fall Symp.* 528–532 (1997).

17. Cao, Y. *et al.* AskHERMES: An online question answering system for complex clinical questions. *J. Biomed. Inform.* **44**, 277–288 (2011).

18. Mckeown, K. R. *et al.* PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information. in *In Proceedings of The First ACM+IEEE JCDL*. 331–340 (2001).

19. Kandula, S., Curtis, D., Hill, B. & Zeng-Treitler, Q. Use of topic modeling for recommending relevant education material to diabetic patients. *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* **2011**, 674–682 (2011).

20. Doupi, P. & van der Lei, J. Towards personalized Internet health information: the STEPPS architecture. *Med. Inform. Internet Med.* **27**, 139–151 (2002).

21. Croft, W. B., Metzler, D. & Strohman, T. *Search Engines: Information Retrieval in Practice.* (Addison-Wesley, 2010).

22. Turtle, H. & Croft, W. B. Evaluation of an Inference Network-based Retrieval Model. *ACM Trans Inf Syst* **9**, 187–222 (1991).

23. Metzler, D., Strohman, T., Zhou, Y. & Croft, W. B. *Indri at TREC 2005: Terabyte Track.* (2004).

24. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet Allocation. *J Mach Learn Res* **3**, 993–1022 (2003).

25. Blei, D. M. & Lafferty, J. D. Visualizing Topics with Multi-Word Expressions. ArXiv09071013 Stat (2009). at <http://arxiv.org/abs/0907.1013>

26. Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* **32**, D267–270 (2004).

27. Lafferty, J. D., McCallum, A. & Pereira, F. C. N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. in 282–289 (Morgan Kaufmann Publishers Inc., 2001). at <http://dl.acm.org/citation.cfm?id=645530.655813>

28. Aronson, A. R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc. AMIA Annu. Symp. AMIA Symp.* **17**–21 (2001).

29. Metzler, D. & Croft, W. B. A Markov Random Field Model for Term Dependencies. in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 472–479 (ACM, 2005). doi:10.1145/1076034.1076115

30. Balasubramanian, N., Allan, J. & Croft, W. B. A Comparison of Sentence Retrieval Techniques. in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 813–814 (ACM, 2007). doi:10.1145/1277741.1277922

31. Cartright, M.-A., Feild, H. A. & Allan, J. Evidence Finding Using a Collection of Books. in *Proceedings of the 4th ACM Workshop on Online Books, Complementary Social Media and Crowdsourcing*. 11–18 (ACM, 2011). doi:10.1145/2064058.2064063

32. Bendersky, M., Croft, W. B. & Smith, D. A. Two-stage Query Segmentation for Information Retrieval. in *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 810–811 (ACM, 2009). doi:10.1145/1571941.1572140