A Sequential Latent Topic-based Readability Model for Domain-Specific Information Retrieval.

Conference or Workshop Item

A Sequential Latent Topic-based Readability Model for Domain-Specific Information Retrieval

Wenya Zhang[1], Dawei Song[1,2], Peng Zhang[1]
, Xiaozhao Zhao[1], and Yuexian Hou[1]

[1] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, China
[2]The Computing Department, The Open University, United Kingdom
{wenyazhang, dwsong, pzhang, zxz, yxhou}@tju.edu.cn,

Abstract. In domain-specific information retrieval (IR), an emerging problem is how to provide different users with documents that are both relevant and readable, especially for the lay users. In this paper, we propose a novel document readability model to enhance the domain-specific IR. Our model incorporates the coverage and sequential dependency of latent topics in a document. Accordingly, two topical readability indicators, namely *Topic Scope* and *Topic Trace* are developed. These indicators, combined with the classical *Surface*-level indicator, can be used to rerank the initial list of documents returned by a conventional search engine. In order to extract the structured latent topics without supervision, the hierarchical Latent Dirichlet Allocation (hLDA) is used. We have evaluated our model from the user-oriented and system-oriented perspectives, in the medical domain. The user-oriented evaluation shows a good correlation between the readability scores given by our model and human judgments. Furthermore, our model also gains significant improvement in the system-oriented evaluation in comparison with one of the state-of-the-art readability methods.

Keywords: domain-specific retrieval; readability; documents reranking

1 Introduction

Conventional search engines aim to return relevant documents based on "similarity" (between a document and a query) and "popularity" (with respect to the hyperlink structure). A recently emerging relevance criteria is document readability. With the diversification of web resources and users, it is increasingly difficult for a search engine to provide different users, especially lay users, with documents in a specific domain that are not only relevant but also readable [10]. The readability plays an important role in assessing documents' relevance [21, 22], quality [1] and utility [19]. However, traditional similarity and popularity measures do not necessarily reflect the readability of the returned documents [10].

In a typical human reading process, the readability of documents can be interpreted at different levels [16, 17]. It is argued in [13] that humans' minds appear to go far beyond the data available, which means there will be a complicated process of abstraction in humans' understanding. Thus, we propose to measure documents readability from two levels. The first is the surface level readability that relates to the surface content. It can be assessed by a series of classical readability features. Beyond the surface content, a higher level, namely the topic level readability, reflects whether it is easy for a user to comprehend the hidden topics in documents. Thus, we propose a topic-based readability method, which can be used to enhance domain-specific IR by considering both the surface and topic level readability of documents.

2 Related Work

There have been various general-purpose readability measures in the literature, such as the Flesh-Kincaid Grade Level and SMOG Index [3][11]. Based on the surface-level features of a document, e.g., word length, sentence length, etc., these classical measures usually generate a numeric score that maps onto an educational grade level. To further improve the accuracy of readability computation, various statistical, semantic and syntactic features of documents have been used [9][15]. However, they are designed for traditional general-purpose texts, thus insufficient to deal with domain-specific documents. Most of the existing measures do not consider the documents' readability at a topic level, which is indeed important for domain-specific documents which often contain a large amount of domain related topics and concepts.

A concept-based approach has been proposed by Yan et al. [16, 17], which takes into account the coverage (Scope) and relatedness (Cohesion) of domain topics (concepts) within a document, with reference to a domain taxonomy. In the taxonomy, the topics are at different abstraction levels and their relationships are organized into a hierarchical tree structure [12]. Topic taxonomy encodes high-quality domain knowledge and can be used to improve a user's understanding of the content of the text [2]. The general hypothesis is that the more abstract a topic is, the more general and easier to understand the topic tends to be. A limitation is that explicit domain taxonomy may not be always available. Recently, the hierarchical Latent Dirichlet Allocation (hLDA) has been widely used to discover the latent topics from large scale data [2][20]. Thus in this paper we propose to automatically build latent topic structures to represent the domain. Moreover, Yan's model does not take into account the sequential dependency between adjacent topics which is important in understanding documents content easily and logically. Different from Yan's work, in [6, 7][14] a readability measure based on the term embedding and sequential discourse cohesion is proposed. However, it does not refer to a domain taxonomy. Nonetheless, their thought about sequential discourse cohesion gives us an inspiration for incorporating sequential dependency information within a latent topic based approach.

In this paper, we propose a novel readability enhanced domain-specific information retrieval model. Specifically, two latent topical indicators, i.e., *Topic Scope* and *Topic Trace*, are proposed. They capture the sequential dependency

of topics at different granularities, through mapping a document onto an automatically constructed topic taxonomy. *Topic Scope*, as originally proposed by Yan et al. [16], reflects the overall coverage of domain topics in a document. *Topic Trace* tracks how the sequence of topics occurring in a document traverses on the topic taxonomy. Additionally, we use the ratio of complex words as an indicator of the document's surface level readability. The individual indicators and their combinations can be used to measure, from different perspectives, the readability of a document. Based on the documents' readability scores, we can then rerank the initial list of results generated by a convention search engine.

3 SEQUENTIAL LATENT TOPIC BASED READABILITY COMPUTATION

3.1 Topic Taxonomy Extraction and Topic Identification

A topic taxonomy can be extracted from a collection of documents. As a tree structure, it consists of topics (nodes) that are at different abstraction levels and connected by the subsumption relationships (edges). In this paper, we use a nonparametric generative procedure, namely the hierarchical Latent Dirichlet Allocation (hLDA), to generate a tree structure of topics by means of a nested Chinese Restaurant Process (nCRP) and Bayesian nonparametric inference. Each topic can be represented as a probability distribution over words in the vocabulary. In the extracted topic taxonomy, the deeper a topic is, the more specific it tends to be. Thus, the root has the broadest meaning, while the leaves are the most specific ones. Figure 1 shows a fragment of topic taxonomy extracted from the CLEF eHealth 2013 medical collection [5].

A domain-specific document can then be mapped onto the topic taxonomy through a topic identification process. In this paper, we identify topics contained in a document based on the occurrence of top 10 probability words from the underlying distributions of topics. Therefore, a document can be represented as a sequence of identified topics, i.e., $d = (t_1, t_2, ..., t_n)$,as illustrated in Figure 3.

3.2 Topical Readability Indicators

After the topic identification, we propose two topical readability indicators. *Topic Trace*, tracks the the identified topics sequentially on the taxonomy. Another indicator, *Topic Scope* reflects the coverage of the identified topics in a document.

Topic Trace (TT) This indicator is based on the hypothesis that the topical line to compose a document is like the planning of travels among a number of scenery spots. A good traveling plan can help tourists visit as many scenery spots as possible with as little expense as possible and as small bumpy leap as possible. Similarly, a well-organized (thus more readable) document should introduce the related topics sequentially with little *Topical Expense* and small *Topical Leap*, which can reflect the coherence among sequential topics defined as *Topic Trace* here. Thus, the *Topic Trace* for a document d_i can be calculated as in Equation (1),

$$Trace(d_i) = Expense(d_i)^{-1} * e^{-\lambda * Leap(d_i)} \quad (1)$$

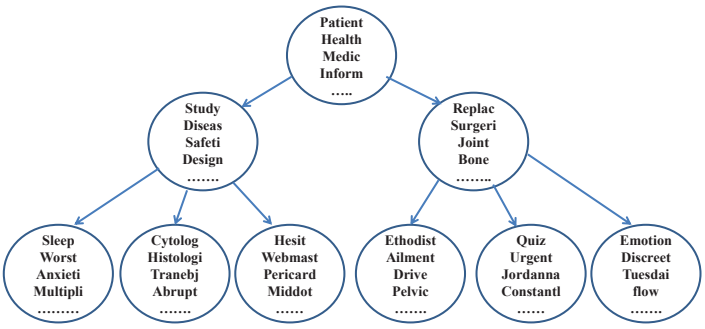

Fig. 1. A fragment of automatically constructed topic taxonomy from CLEF eHealth 2013 medical collection (construction details will be shown in Section 4). In this fragment, the root topic consists of top 4 high probability words of "Patient, Health, Medic Inform" (stemmed by the Porter stemmer) which are general concepts (topics) in medical domain. Its children nodes "Study, Diseas, Safeti" and "Replac, Surgeri, Joint" have relatively specific meaning, while its grandchildren nodes are more specific.

where $Expense(d_i)$ and $Leap(d_i)$ refer to the *Topical Expense* and *Topic Leap*, respectively, and λ is a parameter to control the influence of the *Leap* on the trace score ($\lambda = 0.001$, the optimal values by experiments). A high trace score means the high readability of the document.

Topical Leap means the bumpiness when the identified topics sequentially traverse on the topical taxonomy, as defined in Equation (2). H_{t_j} denotes the depth of topic t_j in the taxonomy.

$$Leap(d_i) = \sum_{t_j, t_{j+1}} \left| H_{t_j} - H_{t_{j+1}} \right| \tag{2}$$

Topical Expense reflects the difficulty to parse the identified topics sequentially. Hypothesizing that the topical expense of a document is inversely related to the overall coherence among the topics within the document, we measure it as follows:

$$Expense(d_i) = \left(\frac{\sum_{t_j} ConCoh(t_j)}{|MC| - 1}\right)^{-1} \tag{3}$$

where MC is the size of the set of identified topics and $ConCoh(t_j)$ computes the contextual coherence, simplified as cc_{t_j}, of t_j in term of its average topical similarity with its surrounding topics (i.e., context).

Specifically, to compute $ConCoh(t_j)$, we use a sliding window [4] with fixed size M (an odd number, $M = 5$ is the optimal value by experiments in this paper) which takes the center topic as the current topic, while the other surrounding topics within the window as contextual topics, as illustrated in Figure 2. The contextual coherence of the current topic can be computed as in Equation (4):

$$ConSim(C_j) = \frac{\sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}} e^{-|m|} * Sim(C_j, C_{j+m})}{|M|} \tag{4}$$

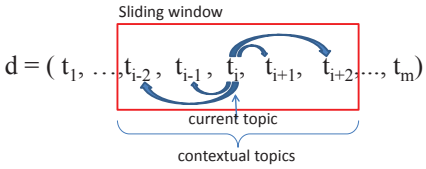

Fig. 2. Sliding window for contextual coherence with $M = 5$.

$Sim(t_j, t_{j+1})$ calculated as in Equation (5), is the topical similarity between the current topic t_j and a context term t_{j+1} within the the sliding window. $e^{|m|}$ means sequential dependency between ti and ti+m gets stronger when they are closer in the sliding window. m the relative distance between the t-wo topics in the window. Thus, we can get a global topic trace vector, i.e., $tv(d) = (cc_{t_1}, cc_{t_2}, ..., cc_{t_n})$, for each document.

One way for calculating similarity between two topics has been shown in Equation (5) [8]. L means the shortest path, and H is the depth of the most specific subsumer. The constants α and β are set to 0.2 and 0.6, respectively (the optimal values by experiments).

$$Sim(t_i, t_{i+m}) = e^{-\alpha L} \frac{e^{\beta H} - e^{-\beta H}}{e^{\beta H} + e^{-\beta H}} \tag{5}$$

By now we have defined all the components in Equation (1) for calculation of document trace, i.e., $Trace(d_i)$. The score of $Trace(d_i)$ falls into the range of (0,1). Figure 3 gives an example, where the $Trace(d_i) = 0.42 * e^{-0.005} = 0.417$. For d_j, the $Trace(d_j) = 0.17 * e^{-0.005} = 0.169$. It turns out that d_i is more readable. Furthermore, from the structure perspective, d_i would seem to be more concise and logical than d_j.

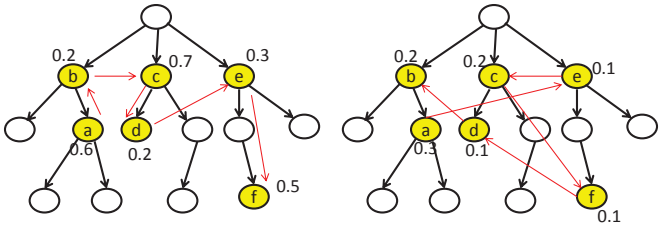

Fig. 3. Topic sequence for $d_i = (t_a, t_b, t_c, t_d, t_e, t_f)$ (left) and $d_j = (t_a, t_e, t_c, t_f, t_d, t_b)$ (right), and their corresponding global topic trace are $tv(d_i) = (0.6, 0.2, 0.7, 0.2, 0.3, 0.5)$ and $tv(d_j) = (0.3, 0.1, 0.2, 0.1, 0.1, 0.2)$, respectively. Thus $Expense(d_i) = [(0.6 + 0.2 + 0.7 + 0.2 + 0.3 + 0.5)/6]^{-1} = 0.42^{-1}$ and $Expense(d_j) = 0.17^{-1}$; $Leap(d_i) = 0.001 * 5 = 0.005$ and $Leap(d_j) = 0.005$.

Topic Scope (TS) Based on a general hypothesis that the overall lower taxonomy depths of identified topics in the taxonomy would indicate a better document readability, we also employ the average tree depth of the identified topics to calculate the topic scope. Compared with Yan's work [16], we measure the document scope on topic level rather than conceptual level. As shown in Equation (6), n_t is the number of identified topics, while $depth(t_i)$ represents the depth of the identified topic t_i on the topic taxonomy. Falling in (0,1), the higher the scope score is, the more readable the document tends to be.

$$Scope(d_i) = e^{-(\frac{\sum_{i=1}^{n} depth(t_i)}{n_t})} \tag{6}$$

3.3 Document Reranking based on Readability

We combine the two levels of readability to calculate the overall readability score of d_i as follows:

$$ReadScore(d_i) = \frac{x * Scope(d_i) + y * Trace(d_i)}{1 + z * Surface(d_i)} \tag{7}$$

where $Surface(d_i)$ measures the surface level readability of the document. Specifically, x, y and z are explored to control the weight of three readability indicators, respectively. Both limited to (0,1), $x + y = 1$, and z is 0 or 1. Thus, *ReadScore* can be normalized into (0,1). The larger the *ReadScore* is, the more readable the document will be.

As shown in Equation (8), we employ the ratio of complex words that are not in the the Dale-Chall word list [3] to calculate the surface level readability, where $ComplexWords$ is the number of complex words and $TotalWords$ is the number of total words in the document.

$$Surface(d_i) = \frac{ComplexWords}{TotalWords} \tag{8}$$

After we get the readability score, in the same way as in [16], we use Equation (9) to compute the total score for reranking, where $RelScore(d_i)$ is the relevance score returned by a conventional search engine. m controls the weight of relevance score in documents reranking, while n controls the weight of readability score.

$$Score(d_i) = RelScore(d_i, Q)^m \cdot e^{-(ReadScore(d_i))^n} \tag{9}$$

4 EXPERIMENTS AND RESULTS

In order to evaluate our proposed model, both user-oriented and system-oriented evaluations have been carried out. The former aims to find out how well our model's prediction is correlated with human judgment on document readability, while the latter aims to evaluate how effectively our model can improve document ranking in medical information retrieval.

Aiming to provide valuable and relevant documents to lay users, CLEF e-Health 2013 dataset [5] contains 50 test queries and one million English documents covering a broad set of medical topics. The initial search results were

returned by the TF-IDF model in Lemur. All documents have been stemmed by Porter stemmer and filtered by SMART 571 stop word list. As a comparison, MeSH (Medical Subject Headings), an existing medical taxonomy, had been used to calculate Yan's model (*Scope*, the most effective indicator). Since it is expensive to construct taxonomy by hLDA on all documents, we employed the top 20 returned documents for all queries as the same method used in [18]. Specifically, we limited the vocabulary to be the 29795 words that appeared in more than 5 documents and a number of meaningless symbols were removed, such as "[", "-", "&", "$" etc. As a result, 634 topics have been nested in a topic taxonomy with a depth of 8, of which a fragment has been shown in Figure 1.

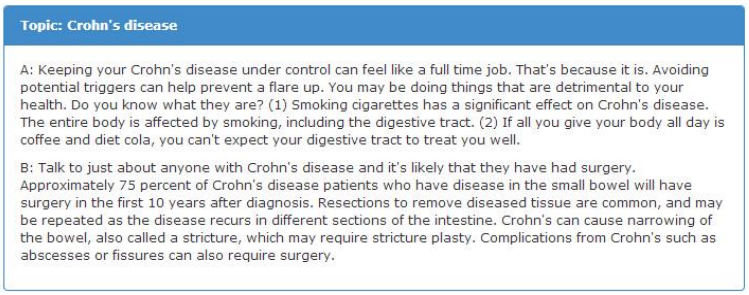

Fig. 4. Sample pair of medical passages with different topic scopes for Query1-Crohn's disease in the first task of user-oriented evaluation.

User-oriented Evaluation. In this evaluation, users were instructed to answer a series of questions related to the readability of the passages selected from CLEF eHealth 2013. We only selected 6 simple queries (Query1-Crohn's disease; Query2-Scar; Query3-Lightheaded; Query4-Liver transplantation; Query5-C.diff; Query6-Cardiac arrest) to avoid exhausting users. For each query, two user tasks, corresponding to topic scope and topic trace respectively, were performed independently with different sets of users to avoid the learning effect. In the first task, one pair of medical passages with different topic scopes (preselected from the top returned documents for the query in initial search results and labeled as passage "A" and "B", each of which are limited to 80-90 words,

Table 1. Calculation Matrix for Similar Rates of Users' Judgements.

	PassageA	PassageB
Topic Scope (TS, by users)	n_{as}	n_{bs}
Topic Trace (TT, by users)	n_{at}	n_{bt}
Average Readability Score (by users)	$Read(A)$	$Read(B)$

as shown in Figure 4) are presented to a set of users. Through actual reading, the users were asked to answer the following questions: (1) Filtering question; (2) Scope related question; (3) Readability score for A; (4) Readability score for B. Detailed information has been shown in Figure 5.

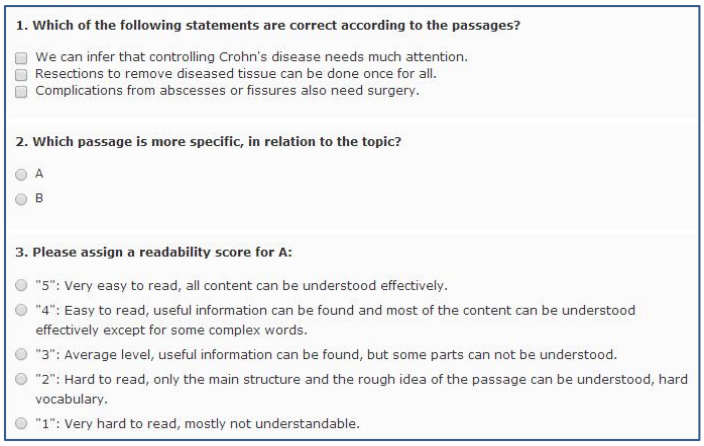

Fig. 5. Detailed questions for the first task in user-oriented evaluation.

$$SimilarRate(TS) = \begin{cases} n_{as}/(n_{as}+n_{bs}), & \text{if } Read(A) > Read(B) \\ n_{bs}/(n_{as}+n_{bs}), & \text{if } Read(A) < Read(B) \end{cases} \quad (10)$$

Table 2. Similar Rate for Users' Judgements

Qid	Query1	Query2	Query3	Query4	Query5	Query6	AvgSimlarRate
TS (by users)	**0.70**	0.10	**0.85**	0.25	0.40	**0.80**	**0.52**
TT (by users)	**0.75**	**0.80**	**0.70**	**0.95**	**0.95**	**0.90**	**0.84**

In the second task, another pair of passages, also manually selected from the top returned documents, with different topic occurrence sequences (i.e., different topic traces), are used for another set of users to answer the same question in the first task, except that the question (2) is replaced by "Which passage describes the topic more logically and smoothly?"("more logically and smoothly" refers to better trace). In question (3) and (4) of both tasks, the readability score "5" means the simplest to read, while "1" means the hardest to read.

The evaluation was conducted through Amazon Mechanical Turk which targets at "crowdsourcing" of Human Intelligence Tasks (HIT) in large scale. Only

Table 3. Pearson Correlation Coefficient for User Evaluation

	TT+SI $(x = 0, y = 1, z = 1)$	TS+TT+SI $(x = y = 0.5, z = 1)$	Yan(Scope)
Pearson	**0.63**	**0.22**	**0.18**

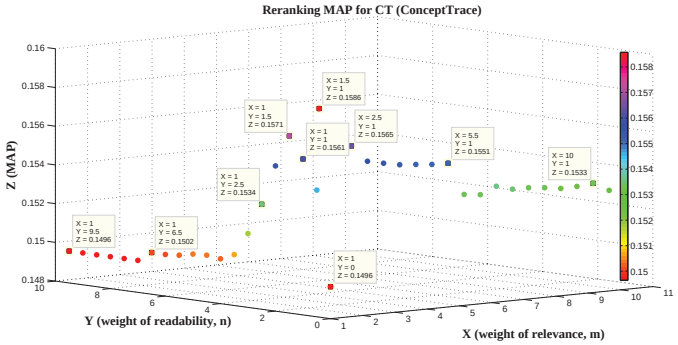

Fig. 6. Reranking MAP for CT (Concept Trace).

the high-qualification turkers are used (i.e., HIT Approval Rate (%) $\geq$ 95). We filtered the data of turkers who did not answer the filtering question (i.e., question (1)) correctly, whose dwell time was less than 40s or whose individual HIT is uncompleted. As a result, we collected the high-quality data of 20 Mechanical Turk users for each pair. For every pair of passages, we computed the average

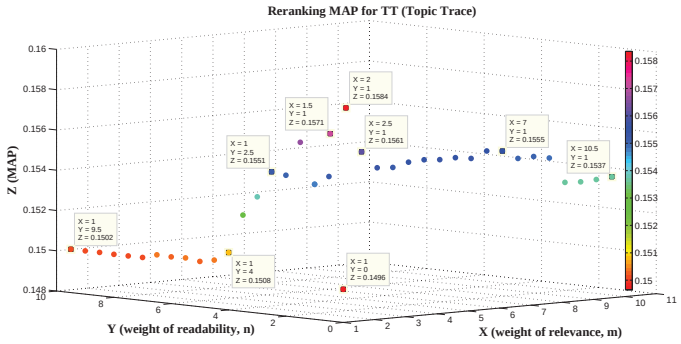

Fig. 7. Reranking MAP for TT (Topic Trace).

readability score for each passage (with average standard deviation 0.89), and we calculated the consistency of users' judgements on topic scope (topic trace) with average readability score in terms of *Similar Rate*. Through referring to Table 1, we derived *SimilarRate* for topic scope in Equation (10), where n_{as},

Table 4. MAP Comparisons for CLEF eHealth 2013 (symbol † means p < 0.05 with paired t-test)

	Baseline	**Yan**	**CT**	**TS**	**TT**	**SI**
MAP	0.1496	0.1515	0.1586	0.1504	0.1584	0.1548
(x,y,z)	-	(1,0,0)	(1,0,0)	(1,0,0)	(0,1,0)	(0,0,1)
(n,m)	-	(2.5,1)	(1.5,1)	(1,3)	(2,1)	(5,1)
(-/+)	-	**+1.27%**	**+5.92%** †	+0.53%	**+5.88%** †	+3.48%
	-	-	**TS+TT**	**TT+SI**	**TS+SI**	**TS+TT+SI**
MAP	-	-	0.1571	0.1583	0.1505	0.1570
(x,y,z)	-	-	(0.5,0.5,0)	(0,1,1)	(1,0,1)	(0.5,0.5,1)
(n,m)	-	-	(1,1)	(1,1)	(3.5,1)	(1,0.5)
(-/+)	-	-	+5.01%	**+5.82%** †	+0.60%	+4.95%

n_{bs}, n_{at}, n_{bt} means the number of users who picked the corresponding choice. $Read(A)$ and $Read(B)$ are the average readability scores assigned by all users. In addition, we calculate it for topic trace in the same way.

Table 2 summarizes the results of *Similar Rate*, in which "TT, by users" shows a good average similar rate with 0.84 among users. It means that users tend to assign higher readability score to the passage with better topic trace.

In addition, we calculated the Pearson Correlation Coefficient between average assigned readability scores and that computed by our model and Yan's, which have been shown in Table 3 with best tuned combing parameters (i.e., x, y and z in Equation (7)). "TT+SI" (combination of Topic Trace and Surface Indicator) has the highest coefficient among all combinations, and "TS+TT+SI" (combination of all indicators) also correlates more closely with average assigned score than Yan's model,which also implies the potential of our proposed model.

System-oriented Evaluation. We also conducted system experiment to examine the proposed indicators and combinations of them to rerank the top 20 documents for all 50 test queries in CLEF eHealth 2013. To explore the relative effect of readability and relevance, we tuned the weights of m and n. Parts of the tuning results have been shown in Figure 6 and 7, through which we can infer that by integrating a certain weight of readability, i.e., n (for instance in Figure 6, when n is around 1), we can get consistent improvement by increasing weight of relevance, i.e, m.

Specifically, we compared the reranking MAP of each indicator and some combinations of them, and picked up their best performance to do the significance test. Detailed results have been shown in Table 4, in which "CT" (*Concept Trace* that implements the idea of *Trace* by referring to MeSH), gains the highest improvement of 5.92% that is better than Yan's model. Meanwhile, "TT" (*Topic Trace*) and "TT+SI" (combination of *Topic Trace* and *Surface Indicator*) also improve the reranking performance significantly. Compared with "CT", "TT" (*Topic Trace*) is competitive by constructing taxonomy automatically, which indicates the good potential of the idea of *Trace*.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a sequential latent topic-based document readability model for domain-specific information retrieval. In our model, two topical readability indicators, namely *Topic Scope* and *Topic Trace* have been developed, which can capture the overall coverage and sequential trace of the latent topics in the document, respectively. Compared with Yan's work [16], on one hand, our model does not require referencing to an existing domain taxonomy. Instead, we automatically construct a latent topic taxonomy from the data. Therefore, our approach is more general and applicable to any domains that may not have an existing taxonomy. On the other hand, we take advantage of the sequential information between adjacent latent topics. Through user-oriented evaluation, our proposed readability indicators and the re-ranking model demonstrate a good correlation with human judgments. Furthermore, our model outperforms a state of the art concept-based model.

In the future, we plan to improve topic taxonomy construction by incorporating n-grams. Meanwhile, refined algorithms and more suitable combinations of readability indicators will be tested.

Acknowledgments. This work is supported in part by Chinese National Program on Key Basic Research Project (973 Program, grant No.2013CB329304, 2014CB744604), the Chinese 863 Program (grant No. 2015AA015403), the Natural Science Foundation of China (grant No. 61272265, 61402324), and the Research Fund for the Doctoral Program of Higher Education of China (grant No. 20130032120044).

References

1. M. Bendersky, W. B. Croft, and Y. Diao. Quality-biased ranking of web documents. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 95–104. ACM, 2011.
2. D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM (JACM)*, 57(2):7, 2010.
3. J. S. Chall and E. Dale. *Readability revisited: The new Dale-Chall readability formula*. Brookline Books, 1995.
4. Y. Chen, X. Yin, Z. Li, X. Hu, and J. X. Huang. A lda-based approach to promoting ranking diversity for genomics information retrieval. *BMC genomics*, 13(Suppl 3):S2, 2012.
5. L. Goeuriot, L. Kelly, G. J. Jones, G. Zuccon, H. Suominen, A. Hanbury, H. Müller, and J. Leveling. Creation of a new evaluation benchmark for information retrieval targeting patient information needs. 2013.
6. S. Jameel, W. Lam, and X. Qian. Ranking text documents based on conceptual difficulty using term embedding and sequential discourse cohesion. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 145–152. IEEE Computer Society, 2012.
7. S. Jameel and X. Qian. An unsupervised technical readability ranking model by building a conceptual terrain in lsi. In *Semantics, Knowledge and Grids (SKG), 2012 Eighth International Conference on*, pages 39–46. IEEE, 2012.

8. J. J. Jiang and D. W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*, 1997.
9. R. J. Kate, X. Luo, S. Patwardhan, M. Franz, R. Florian, R. J. Mooney, S. Roukos, and C. Welty. Learning to predict readability using diverse linguistic features. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 546–554. Association for Computational Linguistics, 2010.
10. J. Y. Kim, K. Collins-Thompson, P. N. Bennett, and S. T. Dumais. Characterizing web content, user interests, and search behavior by reading level and topic. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 213–222. ACM, 2012.
11. R. Senter and E. Smith. Automated readability index. Technical report, DTIC Document, 1967.
12. P. Sripairojthikoon and T. Senivongse. Concept-based readability of web services descriptions. In *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, pages 853–858. IEEE, 2013.
13. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.
14. S. X. WaiLam. N-gram fragment sequence based unsupervised domain-specific document readability. 2012.
15. T. Yamasaki and K.-I. Tokiwa. A method of readability assessment for web documents using text features and html structures. *Electronics and Communications in Japan*, 97(10):1–10, 2014.
16. X. Yan, R. Y. Lau, D. Song, X. Li, and J. Ma. Toward a semantic granularity model for domain-specific information retrieval. *ACM Transactions on Information Systems (TOIS)*, 29(3):15, 2011.
17. X. Yan, D. Song, and X. Li. Concept-based document readability in domain specific information retrieval. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 540–549. ACM, 2006.
18. Z. Ye, J. X. Huang, and H. Lin. Finding a good queryrelated topic for boosting pseudorelevance feedback. *Journal of the American Society for Information Science and Technology*, 62(4):748–760, 2011.
19. E. Yilmaz, M. Verma, N. Craswell, F. Radlinski, and P. Bailey. Relevance and effort: an analysis of document utility. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 91–100. ACM, 2014.
20. Y. Zhang, A. Ahmed, V. Josifovski, and A. Smola. Taxonomy discovery for personalized recommendation. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 243–252. ACM, 2014.
21. Y. Zhang, J. Zhang, M. Lease, and J. Gwizdka. Multidimensional relevance modeling via psychometrics and crowdsourcing. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 435–444. ACM, 2014.
22. G. Zuccon and B. Koopman. Integrating understandability in the evaluation of consumer health search engines. *Proc. of MedIR*, 29, 2014.