Published in final edited form as:

Methods Inf Med. 2013 October 11; 52(5): 395–402. doi:10.3414/ME12-01-0054.

Developing Topic-Specific Search Filters for PubMed with Click-Through Data

Jiao Li^{1,2} and Zhiyong Lu²

¹Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing 100020, China

²National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, U.S

Summary

Objectives—Search filters have been developed and demonstrated for better information access to the immense and ever-growing body of publications in the biomedical domain. However, to date the number of filters remains quite limited because the current filter development methods require significant human efforts in manual document review and filter term selection. In this regard, we aim to investigate automatic methods for generating search filters.

Methods—We present an automated method to develop topic-specific filters on the basis of users' search logs in PubMed. Specifically, for a given topic, we first detect its relevant user queries and then include their corresponding clicked articles to serve as the topic-relevant document set accordingly. Next, we statistically identify informative terms that best represent the topic-relevant document set using a background set composed of topic irrelevant articles. Lastly, the selected representative terms are combined with Boolean operators and evaluated on benchmark datasets to derive the final filter with the best performance.

Results—We applied our method to develop filters for four clinical topics: nephrology, diabetes, pregnancy, and depression. For the nephrology filter, our method obtained performance comparable to the state of the art (sensitivity of 91.3%, specificity of 98.7%, precision of 94.6%, and accuracy of 97.2%). Similarly, high-performing results (over 90% in all measures) were obtained for the other three search filters.

Conclusion—Based on PubMed click-through data, we successfully developed a high-performance method for generating topic-specific search filters that is significantly more efficient than existing manual methods. All data sets (topic-relevant and irrelevant document sets) used in this study and a demonstration system are publicly available at http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/downloads/CQ_filter/

Keywords

Information Retrieval; PubMed Search Filter; PubMed Log Analysis; Clinical Topic

1. Introduction

To facilitate literature retrieval in biomedicine, search filters have been built in academic search engines such as PubMed [1], Health-evidence.ca [2] and EMBASE (Biomedical Answers) [3]. These search filters enable users to focus their searches to specific topics of interest. According to their different design objectives, existing search filters can be

generally classified into two types: methods-based filters and topic-specific filters. The methods-based filters help identify clinical research of high methodological merit such as treatment, diagnosis, prognosis, causation, prediction and etc. [4] Such filters have been successfully integrated in PubMed, known as PubMed Clinical Queries (PubMed-CQ) [5]. More recently, a different kind of filters (known as topic-specific filters) was developed with the aim of helping users identify studies in specialized clinical disciplines/topics such as nephrology [6, 7], geriatrics [8, 9], occupational injury [10], alcohol-impaired driving [11], and off-label drug use [12]. For example, by using a nephrology-specific filter, a user can search for kidney disease-associated low back pain by simply querying 'low back pain' so that the search results will be automatically narrowed down to a set of 'low back pain' relevant articles within nephrology.

Despite the increasingly wide adoption of search filters and their demonstrated benefits for literature retrieval, the number of developed filters remains modest. One major limiting factor of the traditional filter development methods is that they require human experts to manually review articles and select filter terms, making it a very time-consuming and labor-intensive process. For example, in developing a filter for nephrology, four researchers had to manually review more than 4,000 articles to determine if they contain relevant renal information. In addition, 21 clinicians and 7 medical librarians were involved in manual filter term composing [6].

In this study, we propose an automatic method to develop topic-specific filters using the PubMed user click-through data (see Figure 1). More specifically, instead of relying on humans to hand-select relevant papers for a given topic, we automatically detect topic-relevant user queries and their corresponding click-through data to assemble the topic-relevant reference set. Moreover, we use a statistical method to generate and evaluate filter terms without human intervention. In terms of filter performance, we show that our automatically generated filters perform comparably to those that were manually constructed. Finally, to demonstrate our approach is robust and scalable, a total of four different topic-specific filters (nephrology, diabetes, pregnancy, and depression) were generated in this work. We conducted this research as part of an on-going investigation for enhancing PubMed retrieval performance using query log data [13–16]. Note that despite focusing on PubMed, the proposed method for automatic filter generation is rather general and not restricted to any specific literature search platform.

2. Methods

Our approach consists of four separate steps (see Figure 2), similar to the manual filter development workflow [6, 17], but in an automated fashion. First, we processed click-through data collected from PubMed logs as follows: 1a) detecting topic t of interest through semantic analysis of user queries; 1b) collecting the corresponding clicked articles for topic-relevant article set D^{t+} . Second, we constructed a topic irrelevant article set D^{t-} , which was then combined with D^{t+} to become the article pool D^t for filter development. Third, we compiled a list of filter candidate terms by parsing single words, multiword phrases, and Medical Subject Headings (MeSH) [18] terms* from D^t . Subsequently, we evaluated each filter's ability to represent articles in D^{t+} against those in D^{t-} by computing their Odds Ratios. Lastly, with Boolean operators, we combined those most discriminative single-term filter candidates to generate multi-term filters for obtaining the best filter performance. Each step in our workflow is described in details as below.

*"MeSH terms" were used in the cases when identifying filter candidate terms from the document pool because most of the PubMed articles are indexed with MeSH terms where UMLS concepts are not applicable.

2.1 Collecting and processing log data

In this study, since we focused on clinical topics we used click-through data directly from PubMed Clinical Queries (PubMed-CQ)—one of the widely used PubMed search features by clinicians [19]. Figure 3 shows an example of user interactions in PubMed-CQ with a search query ‘sunitinib AND kidney’. Such interactions indicated that (1) the user has special interests in clinical studies because she uses the PubMed-CQ page instead of the general PubMed search page; (2) the user seeks specific information in the nephrology discipline because her search terms contain a topic word ‘kidney’; and (3) the clicked article is likely to be relevant to the user’s query (i.e. containing nephrology related information). The PubMed-CQ click-through data records each user action in a user session. A more detailed description on PubMed log data can be found in our earlier work [20]. Here, we focused on one type of user behavior: a user enters a query followed by abstract click(s). That is, we excluded all queries with no abstract clicks. In total, we collected 3 months (March to May, 2010) worth of PubMed-CQ log data.

To identify clinical topics from user searches, we first used MetaMap [21, 22] to identify UMLS[†] (Unified Medical Language Systems) [23] concepts in user queries. Next, if a concept is found in a query, its more general concept(s) according to the UMLS hierarchy will also be associated to the query. For example, a user query ‘ESRD quality indicators’ was first annotated with UMLS concept ‘Kidney Failure, Chronic’ (CUI=C0022661) by MetaMap because ‘ESRD’ is an abbreviation for ‘End Stage Renal Disease’, an alternative name of this concept. Furthermore in UMLS, ‘Kidney Failure, Chronic’ is linked to several more general concepts such as ‘Renal Insufficiency’ and ‘Kidney Disease’. Thus, the original user query ‘ESRD quality indicator’ was also annotated to these more general UMLS concepts.

2.2 Constructing the article pool

In topic-specific filter development, an article pool needs to contain both topic-relevant articles D^{+} and irrelevant articles D^{-} . Traditionally, filter developers select a set of journals and then sample a set of articles within those journals to construct the article pool [6]. Next, a group of domain experts is asked to determine whether each article in the pool is relevant to the topic or not. By contrast, we constructed the topic-relevant set D^{+} through collecting the articles that users clicked. Given a topic t , we first identified its corresponding root concepts in UMLS (designated as the topic-relevant root concepts hereafter) by hand. Second, by using the ‘has a broader/narrower relationships’ between UMLS concepts, we systematically computed all the sub-concepts of those topic-relevant root concepts found in the previous step. Third, we automatically identified a set of relevant queries $\{q_1, q_2, \dots, q_n\}$ associated with either the topic-relevant root concepts or their sub-concepts (see section 2.1). Lastly, for each relevant query, we included all of its corresponding clicked articles $\{D_1, D_2, \dots, D_n\}$ to the topic-relevant set D^{+} . On the other hand, we constructed the topic-irrelevant set D^{-} based on keyword searches, excluding any articles containing either topic-relevant root concept terms or their sub-concept terms. To ensure that the articles in D^{-} are relevant to clinical studies, we required articles in D^{-} to be published in one of the 121 “Core Clinical Journals” also known as “Abridged Index Medicus” (AIM) [24]. Following the lead of [6], the pooled article set $D^t \{D^{+}, D^{-}\}$ was further split into a development set D_{dev}^t and a validation set D_{val}^t by using a ratio of three to two.

[†]MetaMap returns UMLS concepts when applied to user queries.

2.3 Identifying filter candidate terms

A topic-specific filter aims to best define article set D^{t+} and distinguish it against set D^{t-} . In this study, filter candidate terms—also known as discriminative features—were drawn from words and phrases from the article title and abstract, as well as from MeSH terms using an NCBI internal NLP toolkit. As a result, a list of filter candidate terms $F = \{f_1, f_2, \dots, f_k\}$ is compiled, where $f_i \in F$ is also distinguished by its source (i.e., title, abstract or MeSH).

For a set of articles, a single term candidate f_i can serve as either a positive filter (selecting articles containing f_i) or a negative one (excluding articles containing f_i). We used Odds Ratio () to rank all positive and negative filters. As shown in the following equation, $P(f_i|t^+)$ and $P(f_i|t^-)$ are conditional probabilities of filter f_i in topic-relevant and irrelevant sets, respectively. The positive and negative values of $\theta(f_i)$ correspond to the positive and negative filters. The greater an absolute value $|\theta(f_i)|$ is, the more discriminative the filter f_i is.

$$\theta(f_i) = \log \frac{P(f_i|t^+)(1 - P(f_i|t^-))}{P(f_i|t^-)(1 - P(f_i|t^+))}$$

We calculated $\theta(f_i)$ for each filter term candidate using the development set, resulting in a ranked list of positive and negative filters in the descending order of $|\theta(f_i)|$.

2.4 Combining filter candidate terms

To generate the final filter F_{book} we first used the Boolean operator “OR” to combine multiple top-ranked filter candidate terms in each category (positive or negative) separately. Next, we used the Boolean operator “NOT” to connect the two categories. For example, if two positive filter candidate terms $\{f_1, f_2\}$ and two negative terms $\{f_3, f_4\}$ were selected, the final filter was constructed as $(f_1 \text{ OR } f_2) \text{ NOT } (f_3 \text{ OR } f_4)$. In order to obtain an optimal filter, we evaluated different combinations of positive and negative single filter terms using benchmark data sets. Note that in order to optimize the filter performance, different numbers of positive and negative terms in the final filter F_{book} may be selected for different topics.

2.5 Topic-specific filter evaluation metrics

Sensitivity, *Specificity*, *Precision*, and *Accuracy* were used to evaluate filter performance (see Table 1). *Sensitivity* is the proportion of topic relevant articles identified; *Specificity* is the proportion of topic irrelevant articles not identified; *Precision* is the proportion of identified articles with topic relevant information; and *Accuracy* is the proportion of all articles correctly identified by a filter.

We first tested and compared performance of filters F_{book} made up of different filter candidate terms on the development set. Next, on the validation set we tested those filters that performed well on the development set.

3. Results

To help understand how our method works in practice and compare it with the state of the art, we developed a filter for nephrology as a specific case study. Moreover, to demonstrate its scalability and robustness, we also generated three other filters of different topic types.

3.1 Renal-relevant PubMed queries

We manually identified five UMLS concepts ('Kidney Disease', 'Kidney', 'Renal Replacement Therapy', 'Renal Circulation' and 'Nephrology') across different semantic types as renal-relevant root concepts. As such, these five root concepts and their sub-concepts in UMLS were used to identify corresponding user queries that are relevant to renal studies. Table 2 shows the five renal-relevant root concepts, their UMLS identifiers (CUIs), semantic types, and sample sub-concepts. Using all the related UMLS concepts, we identified 672 unique queries in our query log relevant to the renal topic.

3.2 Article pool for renal filter development

Through tracking the clicked articles in the 672 renal-related search queries, we collected 862 articles for the topic relevant set D^{+} . In the renal irrelevant article set D^{-} construction, we excluded 13,855 core clinical journal articles that contain either renal-relevant root concepts or their sub-concepts. From the remaining set, we randomly selected 3468 articles as the renal-irrelevant set so that the ratio of D^{+} vs. D^{-} is consistent with the previous study [6] for comparison. Further, the pooled set was split into a development set D_{dev}^t of 2,598 articles and validation set D_{val}^t of 1,732 articles with a ratio of three to two.

3.3 Renal filter candidate terms

From the 4,330 articles in D^t , we generated a total of 194,135 filter candidate terms, including 24,692 words/phrases from titles, 133,802 words/phrases from abstracts, and 7,697 MeSH terms. We calculated Odds Ratio () for each filter candidate using the development set. Table 3 shows the top 20 positive filter candidate terms and their performance. As can be seen, these top ranked filters using the development set achieved equally good performance on the validation set.

3.4 Combined filter for renal study

We combined the top positive filter candidate terms with the Boolean operator "OR" and assessed the combined filters on the development set. As shown in Figure 4, when adding single filters together, the overall sensitivity increased quickly while specificity and precision remained relatively unchanged. When the top 40 positive filter candidate terms were combined, our final filter (after translated into PubMed format [26]) achieved a sensitivity of 87.8%, specificity of 98.7%, precision of 94.2%, and accuracy of 96.5% on the development set.

As a comparison, we applied the highest sensitivity (Renal_sen) and highest specificity filter (Renal_spc) from the previous study [6] to our data sets and showed their results in Table 4. Note that unlike Renal_spc, no improved results were found in our filter when combining negative terms with the Boolean operator "NOT".

3.5 Filters for diabetes, pregnancy, and depression

Table 5 shows three additional filters developed by our method for diabetes (UMLS semantic type: disease), pregnancy (UMLS semantic type: organism function), and depression (UMLS semantic type: mental or behavioral dysfunction) respectively. The numbers of selected positive filter candidate terms for diabetes, pregnancy and depression are 15, 13, and 13 respectively. Similar to our renal filter, no negative terms were needed for these three topics using our method. As can be seen in Table 5, the selected terms in the filters for Diabetes, Pregnancy, and Depression are indeed relevant to their corresponding topics. The three filters all achieved high performance.

4. Discussion

4.1 Performance comparison

As mentioned earlier, our approach differs from the existing methods in several ways. First, we identify topic-relevant articles through query logs as opposed to hand selection. Additionally, a topic-irrelevant article set is also automatically generated. Second, filter candidate terms are automatically generated and evaluated based on the indexing terms such as MeSH. As such, we show that our approach can be scaled to generate filters for different topics in an efficient and cost-effective manner.

With regard to the filter performance, we show with the results in Table 4 that our automatically generated filters achieved comparable performances in sensitivity and specificity to those manually constructed ones. As a matter of fact, by comparing the actual terms[§] in our final renal filter vs. two existing filters Renal_sen and Renal_spc in Table 4, we see a significant overlap—50% of the filter terms in Renal_sen and 80% of positive filter terms in Renal_spc are also found in our renal filter.

4.2 Using MeSH vs. non-MeSH terms in search filters

Our filter candidate terms were drawn from words and phrases in the article title and abstract, as well as MeSH terms. There are two main reasons why our filter candidate selection is not limited to MeSH terms, which are created for describing the subject of journal articles [26] and were used as major features in previous biomedical literature mining studies [27–29]. First, MeSH terms curated by librarians do not always represent article topics in the author's and reader's view [30–32]. Second, not all articles are indexed with MeSH terms. Even for those articles that are eventually indexed with MeSH terms, it is estimated that there are on average 90 days delay in doing so [33]. For these reasons, features from the article itself (*i.e.*, words and phrases from the title and abstract) were included together with MeSH terms as filter candidate terms in this study.

4.3 Limitations of this research

Inspired by our earlier work of using query log data for improving retrieval effectiveness in PubMed [14, 20], in this work we used click-through data to model the human review process for finding topic-relevant articles by leveraging user interaction—clicking articles—as positive feedback. That is, the clicked articles are assumed to be relevant to the given search topics. Such a strategy results in the topic-relevant reference set construction at substantially lower cost and eases the burden of hiring a group of domain experts. Despite the reported success of using clicks as relevance judgment [34, 35], potential noise and bias may arise [36–38]. This led to two limitations of our current study. First, the clicked articles may be noisy because users may click and view articles with attractive titles but not relevant to their search topics. To maximally account for such a user behavior, we specifically made use of log data from PubMed-CQ because its users tend to focus on retrieving high-quality studies for clinical care [19]. Second, relying on click-through data limits our method to those topics that are associated with sufficient numbers of search queries or user clicks. In this regard, we leave the evaluation of how user relevance judgments based on click-through data may affect the results of our topic-specific filter development as future work. Finally, separate large-scale user-studies are warranted to assess the practical benefits of such automatically generated search filters on PubMed retrieval.

[§]Our filter terms do not contain the truncation symbol that is used at the end of a word to search for all terms that begin with that basic word. Hence, we consider as a match when one of the variations (e.g., nephritis) can be found given a truncated term (e.g., nephri*).

5. Conclusion

In this study, we developed a computational method to automatically design topic-specific filters using PubMed click-through data. Our automatic method achieved comparable performance against existing methods that heavily rely on human reviews and curation. Moreover, we showed that our method is applicable to a wide range of clinical topics in an efficient and cost-effective manner. We believe our method is beneficial to other real-world applications with regards to literature search service.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, the National Key Technology R&D Program of China (Grant No. 2013BAI06B01), and the Peking Union Medical College Research Fund for Young Scientists. The authors would like to thank Chih-Hsuan Wei for implementing a demonstration system for our work, Drs. W. John Wilbur, Won Kim, Sun Kim and Rezarta Islamaj Dogan for their helpful comments and discussion on this work, and Bethany Harris for proofreading our manuscript.

References

- Shariff SZ, Sontrop JM, Haynes RB, Iansavichus AV, McKibbin KA, Wilczynski NL, et al. Impact of PubMed search filters on the retrieval of evidence by physicians. *CMAJ*. 2012; 184(3):E184–E190. [PubMed: 22249990]
- Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC Med Res Methodol*. 2012; 12:51. [PubMed: 22512835]
- Golder S, Loke YK. The performance of adverse effects search filters in MEDLINE and EMBASE. *Health Info Libr J*. 2012; 29(2):141–151. [PubMed: 22630362]
- Jenkins M. Evaluation of methodological search filters--a review. *Health Info Libr J*. 2004; 21(3):148–163. [PubMed: 15318913]
- PubMed's Clinical Queries [internet]. Bethesda (MD): National Library of Medicine (US); [cited 2012 Dec]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/clinical>
- Garg AX, Iansavichus AV, Wilczynski NL, Kastner M, Baier LA, Shariff SZ, et al. Filtering Medline for a clinical discipline: diagnostic test assessment framework. *BMJ*. 2009; 339:b3435. [PubMed: 19767336]
- Iansavichus AV, Haynes RB, Shariff SZ, Weir M, Wilczynski NL, McKibbin KA, et al. Optimal search filters for renal information in EMBASE. *Am J of Kidney Dis*. 2010; 56(1):14–22. [PubMed: 20231047]
- van de Glind EM, van Munster BC, Spijker R, Scholten RJ, Hooft L. Search filters to identify geriatric medicine in Medline. *J Am Med Inform Assoc*. 2012; 19(3):468–472. [PubMed: 21946235]
- Kastner M, Wilczynski NL, Walker-Dilks C, McKibbin KA, Haynes RB. Age-specific search strategies for Medline. *J Med Internet Res*. 2006; 8(4):e25. [PubMed: 17213044]
- Beahler CC, Sundheim JJ, Trapp NI. Information retrieval in systematic reviews: challenges in the public health arena. *Am J Prev Med*. 2000; 18(4 Suppl):6–10. [PubMed: 10793275]
- Goss C, Lowenstein S, Roberts I, Diguseppi C. Identifying controlled studies of alcohol-impaired driving prevention: designing an effective search strategy. *J Inf Sci*. 2007; 33(2):151–162.
- Mesgarpour B, Muller M, Herkner H. Search strategies to identify reports on "off-label" drug use in EMBASE. *BMC Med Res Methodol*. 2012; 12:190. [PubMed: 23272771]
- Lu Z, Xie N, Wilbur WJ. Identifying related journals through log analysis. *Bioinformatics*. 2009; 25(22):3038–3039. [PubMed: 19734155]
- Lu, Z.; Wilbur, WJ.; McEntyre, JR.; Iskhakov, A.; Szilagyi, L. Finding query suggestions for PubMed. *Proceedings of the American Medical Informatics Association 2009 Annual Symposium*; 2009 Nov 14–18; San Francisco, USA: AMIA; 2009.

15. Islamaj Dogan R, Lu Z. Click-words: learning to predict document keywords from a user perspective. *Bioinformatics*. 2010; 26(21):2767–2775. [PubMed: 20810602]
16. Neveol A, Islamaj Dogan R, Lu Z. Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction. *J Biomed Inform*. 2011; 44(2):310–318. [PubMed: 21094696]
17. Wilczynski NL, Morgan D, Haynes RB. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak*. 2005; 5:20. [PubMed: 15969765]
18. Medical Subject Headings (MeSHR) [internet]. Bethesda (MD): National Library of Medicine (US); [cited 2012 Dec]. Available from: <http://www.nlm.nih.gov/mesh/>
19. Corrao S, Colomba D, Arnone S, Argano C, Di Chiara T, Scaglione R, et al. Improving efficacy of PubMed Clinical Queries for retrieving scientifically strong studies on treatment. *J Am Med Inform Assoc*. 2006; 13(5):485–487. [PubMed: 16799123]
20. Islamaj Dogan R, Murray GC, Neveol A, Lu Z. Understanding PubMed user search behavior through log analysis. *Database (Oxford)*. 2009:bap018. [PubMed: 20157491]
21. Aronson, AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings of the American Medical Informatics Association 2001 Annual Symposium*; 2001 Nov 3–7; Washington DC, USA: AMIA; 2001.
22. MetaMap [internet]. Bethesda (MD): National Library of Medicine (US) [cited 2012 Dec]; Available from: <http://metamap.nlm.nih.gov/>
23. Unified Medical Language System R (UMLSR) [internet]. Bethesda (MD): National Library of Medicine (US) [cited 2012 Dec]; Available from: <http://www.nlm.nih.gov/research/umls/>
24. Core clinical journals [internet]. Bethesda (MD): National Library of Medicine (US); [cited 2012 Dec]. Available from: <http://www.nlm.nih.gov/bsd/aim.html>
25. Croft, WB.; Metzler, D.; Strohmman, T. *Search engines: information retrieval in practice*. 2. Pearson Education Inc; 2010.
26. Indexed Field in PubMed [internet]. Bethesda (MD): National Library of Medicine (US); [cited 2012 Dec]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Search_Field_Descrip
27. McCray AT, Gefeller O, Aronsky D, Leong TY, Sarkar IN, Bergemann D, et al. The birth and evolution of a discipline devoted to information in biomedicine and health care. As reflected in its longest running journal. *Methods Inf Med*. 2011; 50(6):491–507. [PubMed: 22146913]
28. Kastrin A, Peterlin B, Hristovski D. Chi-square-based scoring function for categorization of MEDLINE citations. *Methods Inf Med*. 2010; 49(4):371–378. [PubMed: 20091016]
29. Yen YT, Chen B, Chiu HW, Lee YC, Li YC, Hsu CY. Developing an NLP and IR-based algorithm for analyzing gene-disease relationships. *Methods Inf Med*. 2006; 45(3):321–329. [PubMed: 16685344]
30. Neveol, A.; Islamaj Dogan, R.; Lu, Z. Author keywords in biomedical journal articles. *Proceedings of the American Medical Informatics Association 2010 Annual Symposium*; 2010 Nov 13–17; Washington DC, USA: AMIA; 2010.
31. Lu Z, Kim W, Wilbur WJ. Evaluation of query expansion using MeSH in PubMed. *Inf Retr Boston*. 2009; 12(1):69–80. [PubMed: 19774223]
32. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol*. 2002; 31(1):150–153. [PubMed: 11914311]
33. Huang M, Neveol A, Lu Z. Recommending MeSH terms for annotating biomedical articles. *J Am Med Inform Assoc*. 2011; 18(5):660–667. [PubMed: 21613640]
34. Macdonald, C.; Ounis, I. Usefulness of quality click-through data for training. *Proceedings of the Workshop on Web Search Click Data*; 2009 Feb 9; Barcelona, Spain: ACM; 2009.
35. Joachims, T. Optimizing search engines using clickthrough data. *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining*; 2002 Jul 23–26; Edmonton, Alberta, Canada: ACM SIGKDD; 2002.

36. Joachims, T.; Granka, L.; Pan, B.; Hembrooke, H.; Gay, G. Accurately interpreting clickthrough data as implicit feedback. *Proceedings of the 28th International Conference on Research and Development in Information Retrieval*; 2005 Aug 15–19; Salvador, Brazil: ACM SIGIR; 2005.
37. Carterette, B.; Jones, R. Evaluating search engines by modeling the relationship between relevance and clicks. *Proceedings of the 21st Annual Conference on Neural Information Processing Systems*; 2007 Dec 3–6; Vancouver, B.C., Canada: NIPS; 2007.
38. Xu, J.; Chen, C.; Xu, G.; Li, H.; Abib, ERT. Improving quality of training data for learning to rank using click-through data. *Proceedings of the 3rd International Conference on Web Search and Data Mining*; 2010 Feb 3–6; New York, USA: ACM WSDM; 2010.

Figure 1.
Overview of our method for automatically developing topic-specific filters

Figure 2.
Workflow for filter development based on users' searches and click actions

Figure 3.

Users' search behavior in PubMed-CQ.

(A) Snapshot of the interface of PubMed-CQ for searching by clinical study category. A user inputs search terms such as 'sunitinib AND kidney' in the search box; selects a category (e.g., Therapy) and a scope (e.g., Narrow). (B) An overview of the user's interactions with it. Upon the above search terms and selections sent to PubMed-CQ, a list of articles is retrieved. The user can browse the returned result and click to review abstracts of interest.

Figure 4.
Performance of combined filters for the renal topic on the development set

Table 1

Definition of topic-specific filter evaluation metrics.

Filter (F_{bool})	Topic relevant article set	Topic irrelevant article set
Articles identified	a	b
Articles not identified	c	d

- Sensitivity= $a/(a+c)$: proportion of topic relevant articles identified
- Specificity= $d/(b+d)$: proportion of topic irrelevant articles not identified
- Precision= $a/(a+b)$: proportion of identified articles with topic relevant information
- Accuracy= $(a+d)/(a+b+c+d)$: proportion of all articles dealt with correctly by filter

Table 2

Relevant UMLS concepts for the renal topic detection

Root concept	CUI	Semantic type	Sample sub-concepts
Kidney Disease	C0022658	Disease or Syndrome	Nephrosclerosis
Kidney	C0022646	Body Part, Organ, or Organ Component	Kidney Pelvis
Renal Replacement Therapy	C0022671	Therapeutic or Preventive Procedure	Renal Dialysis
Renal Circulation	C0035070	Organ or Tissue Function	Renal Plasma Flow
Nephrology	C0027712	Biomedical Occupation or Discipline	---

Table 3

Top 20 positive filter candidate terms in the renal filter development

Filter (<i>f_i</i>)		<i>f_i</i>	Filter Performance				
Term	Source		Data set	Sensitivity	Specificity	Precision	Accuracy
kidney	title	5,851	Development	0.145	1.000	1.000	0.830
			Validation	0.139	0.999	0.980	0.828
kidney failure, chronic	MeSH	5,835	Development	0.143	1.000	1.000	0.829
			Validation	0.186	1.000	1.000	0.838
dialysis	abstract	5,819	Development	0.141	1.000	1.000	0.829
			Validation	0.139	1.000	1.000	0.829
serum creatinine	abstract	5,701	Development	0.128	1.000	1.000	0.826
			Validation	0.148	0.999	0.981	0.830
kidney transplantation	MeSH	5,572	Development	0.114	1.000	1.000	0.824
			Validation	0.096	1.000	1.000	0.820
renal	title	5,548	Development	0.331	0.998	0.977	0.865
			Validation	0.348	0.999	0.984	0.869
kidney disease	abstract	5,532	Development	0.108	1.000	0.982	0.822
			Validation	0.180	1.000	1.000	0.837
proteinuria	abstract	5,532	Development	0.108	1.000	0.982	0.822
			Validation	0.122	0.999	0.977	0.824
glomerular	abstract	5,512	Development	0.106	1.000	0.982	0.822
			Validation	0.099	0.999	0.971	0.820
creatinine	MeSH	5,449	Development	0.101	1.000	0.981	0.821
			Validation	0.093	0.999	0.970	0.819
chronic kidney	abstract	5,383	Development	0.095	1.000	0.980	0.819
			Validation	0.104	1.000	1.000	0.822
renal	abstract	5,362	Development	0.607	0.993	0.954	0.916
			Validation	0.672	0.991	0.947	0.927
renal dialysis	MeSH	5,361	Development	0.095	1.000	1.000	0.820
			Validation	0.130	1.000	1.000	0.827

Filter (f_i)		Filter Performance				
		Data set	Sensitivity	Specificity	Precision	Accuracy
renal disease	abstract	Development	0.087	1.000	0.978	0.818
		Validation	0.107	0.999	0.974	0.822
creatinine /blood	MeSH	Development	0.081	1.000	0.977	0.817
		Validation	0.067	0.999	0.958	0.814
creatinine	abstract	Development	0.203	0.999	0.972	0.840
		Validation	0.058	0.999	0.952	0.812
glomerular filtration	abstract	Development	0.077	1.000	0.976	0.816
		Validation	0.058	0.999	0.952	0.808
filtration rate	abstract	Development	0.075	1.000	0.975	0.816
		Validation	0.055	0.999	0.950	0.811
renal failure	abstract	Development	0.137	0.999	0.973	0.828
		Validation	0.159	0.998	0.948	0.831
kidney neoplasms	MeSH	Development	0.075	1.000	1.000	0.816
		Validation	0.075	1.000	1.000	0.816

Table 4

Comparison with existing renal filters[‡]

Combined Filter	Data set	Sensitivity	Specificity	Precision	Accuracy
Our filter	Development	87.8%	98.7%	94.2%	96.5%
	Validation	91.3%	98.7%	94.6%	97.2%
Renal_spc [6]	Development	69.2%	99.7%	98.1%	93.6%
	Validation	71.9%	99.6%	98.0%	94.1%
Renal_sen [6]	Development	93.8%	98.5%	94.0%	97.6%
	Validation	95.4%	98.8%	94.3%	97.9%

- **Ours:** “kidney”[tiab] OR “renal”[tiab] OR “nephropathy”[tiab] OR “dialysis”[tiab] OR “proteinuria”[tiab] OR “glomerular”[tiab] OR “creatinine”[tiab] OR “filtration rate”[tiab] OR “nephrotic”[tiab] OR “nephropathy”[tiab] OR “hemodialysis”[tiab] OR “nephritis”[tiab] OR “kidney failure, chronic”[mh] OR “kidney diseases”[mh] OR “creatinine”[mh] OR “renal dialysis”[mh] OR “kidney neoplasms”[mh] OR “carcinoma, renal cell”[mh] OR “proteinuria”[mh] OR “kidney failure, chronic”[mh] OR “kidney transplantation”[mh]
- **Renal_spc:** (“renal replacement therapy”[majr] OR “kidney diseases”[majr] OR kidney*[ti] OR nephri*[ti] OR renal[ti] OR “kidney”[majr:noexp] OR “renal dialysis”[mh] OR “kidney function tests”[majr] OR “proteinuria”[majr:noexp] OR glomerul*[ti]) NOT (“kidney neoplasms”[majr] OR pyelonephritis[majr:noexp] OR “urinary tract infections”[majr] OR “nephrolithiasis”[majr])
- **Renal_sen:** “kidney diseases”[mh] OR “renal replacement therapy”[mh] OR renal[tw] OR kidney*[tw] OR (nephre*[tw] OR nephri*[tw] OR nephroc*[tw] OR nephrog*[tw] OR nephrol*[tw] OR nephron*[tw] OR nephrop*[tw] OR nephros*[tw] OR nephrot*[tw]) OR proteinuria[tw]

[‡]The PubMed search field descriptions and tags in Table 4 and 5 can be found at http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Search_Field_Descrpt. Note that we appended the tag [tiab] for filter candidates from the title or abstract because PubMed does not have separated tags for each of them.

Table 5

Topic-specific filters for PubMed search

Topic	Topic type	Filter Performance				
		Data set	Sensitivity	Specificity	Precision	Accuracy
Diabetes	disease	Development	94.2%	97.7%	91.1%	97.0%
		Validation	94.1%	98.3%	93.3%	97.5%
Pregnancy	organism function	Development	99.0%	99.5%	98.2%	99.4%
		Validation	98.9%	99.2%	97.0%	99.2%
Depression	mental/behavioral dysfunction	Development	97.1%	99.2%	96.8%	98.8%
		Validation	96.1%	99.5%	98.0%	98.8%

- **Diabetes filter:** “diabetes”[tiab] OR “mellitus”[tiab] OR “diabetic”[tiab] OR “glycemic”[tiab] OR “plasma glucose”[tiab] OR “blood glucose”[tiab] OR “glucose levels”[tiab] OR “diabetes mellitus, type 2”[mh] OR “hemoglobin a, glycosylated”[mh] OR “diabetes mellitus, type 1”[mh] OR “insulin/therapeutic use”[mh] OR “diabetes complications”[mh] OR “diabetes, gestational”[mh] OR “hypoglycemic agents”[mh] OR “diabetic neuropathies”[mh]
- **Pregnancy filter:** “pregnancy”[tw] OR “pregnant”[tiab] OR “birth”[tiab] OR “gestation”[tiab] OR “pregnancies”[tiab] OR “mothers”[tiab] OR “maternal”[tiab] OR “labor”[tiab] OR “perinatal”[tiab] OR “postpartum”[tiab] OR “trimester”[tiab] OR “prenatal”[tiab] OR “mother”[tiab]
- **Depression filter:** “depression”[tw] OR “antidepressive agents”[mh] OR “depressive disorder”[mh] OR “psychiatric status rating scales”[mh] OR “serotonin uptake inhibitors/therapeutic use”[mh] OR “cognitive therapy”[mh] OR “personality inventory”[mh] OR “depressive”[tiab] OR “antidepressant”[tiab] OR “anxiety”[tiab] OR “antidepressants”[tiab] OR “mood”[tiab] OR “disorder mdd”[tiab]