CupQ: A New Clinical Literature Search Engine

Jesse Wang[1][a] and Henry Kautz[2][b]

[1]*Department of Translational Biomedical Science, University of Rochester Medical Center, Rochester, NY, U.S.A.*
[2]*Department of Computer Science, University of Rochester, Rochester, NY, U.S.A.*

Keywords: Applied Computing, Life and Medical Sciences, Health Care Information Systems.

Abstract: A new clinical literature search engine, called CupQ, is presented. It aims to help clinicians stay updated with medical knowledge. Although PubMed is currently one of the most widely used digital libraries for biomedical information, it frequently does not return clinically relevant results. CupQ utilizes a ranking algorithm that filters non-medical journals, compares semantic similarity between queries, and incorporates journal impact factor and publication date. It organizes search results into useful categories for medical practitioners: reviews, guidelines, and studies. Qualitative comparisons suggest that CupQ may return more clinically relevant information than PubMed. CupQ is available at https://cupq.io/.

1 INTRODUCTION

The task of staying updated with advances in medicine remains a challenging aspect of clinical practice. An average of about two biomedical documents is added to the literature every minute (Fiorini et al., 2018). The widely used PubMed digital library often does not deliver clinically relevant results within a reasonable time frame (Agoritsas et al., 2014; Ho et al., 2016; Davies, 2011). Other resources, such as UpToDate, NEJM Journal Watch, and ACP Journal Club, rely on the expensive and time-consuming process of using human curators to manually comb the literature for clinical information. The current utilities for medical information retrieval may be inadequate for continuing medical education and consequently may be hindering efforts to improve patient care.

PubMed is a biomedical digital library built and maintained by the United States National Center for Biotechnology Information. It often requires users to select filters, identify MeSH terms, and generate boolean entries to distill relevant clinical results (Russell-Rose and Chamberlain, 2017; Lindsey and Olin, 2013). The complexity of PubMed may contribute to low search satisfaction among healthcare professionals (Agoritsas et al., 2014; Ho et al., 2016; Davies, 2011). Moreover, the newly released Best

    [a] https://orcid.org/0000-0001-8269-1930
    [b] https://orcid.org/0000-0001-5219-2970

Match relevance algorithm does not incorporate important metrics such as journal rank and semantic similarity (Fiorini et al., 2018). These ranking signals also appear to be missing in the related search tool, PubMed Clinical Queries. To better fulfill the information needs of medical practice, PubMed may require further improvements.

This paper discusses the development of a new medical literature search engine called CupQ. The system uses Word2Vec to generate word embeddings for comparing semantic similarity between queries and documents (Mikolov et al., 2013; White et al., 2015). It also considers journal impact factor (JIF) and publication date. Results are organized by reviews, guidelines, and clinical studies. Documents written in English and published in journals listed in the medicine subject area of ScimagoJR are returned. Example search results suggest that CupQ may be more effective than PubMed for returning relevant clinical information. This publication aims to encourage utilization of CupQ for staying updated with medical literature.

2 RELATED WORK

Lu provides a survey of search tools for biomedical literature, including Quertle, MEDIE, and Semantic MEDLINE (Lu, 2011).

Quertle is a semantic search engine utilizing over 250 million subject-verb-object (SVO) associations

225

to provide relevant publications (Rindflesch et al., 2011). It also features "Power Terms" that allow users to search topics. Example terms, denoted by a dollar sign ($) prefix, include "$Amino Acids," "$Biomarkers," and "$Chemicals." In addition, Quertle differentiates capitalizations, such as the "WHO" abbreviation for the World Health Organization and the "who" pronoun. Search results are presented in two tabs. One tab lists results derived from its semantic-based algorithm. Another tab lists results obtained from a standard PubMed search. Quertle was developed and is currently maintained by a for-profit private enterprise. The exact details of its search process are consequently unavailable to the public.

MEDIE also aims to incorporate grammatical meaning into its search algorithm (Ohta et al., 2006). It returns documents that match the user's desired SVO relations. For example, the query "what does p53 activate" would produce results that contain sentences matching "activate" and "p53" as the verb and object, respectively. Queries in MEDIE are first annotated with part-of-speech tags through the Enju head-driven phrase structure grammar parser. Genes and diseases are also annotated through a dictionary comparison approach. After annotation, results returned from a standard keyword search are filtered based on the predicate structure of their sentences. Users are shown the results with sentences matching the specified semantic relations.

Semantic MEDLINE, similar to Quertle and MEDIE, utilizes linguistic information (Rindflesch et al., 2011). In particular, it extracts normalized representations of semantic relations. For example, the phrase "Genes AFFECTS Circadian Rhythms" was parsed from the title "Clock genes are the genes that control circadian rhythms in physiology and behavior." The extraction process was developed using the SemRep natural language processing platform, which depends on the National Library of Medicine's Unified Medical Language System. Semantic categories and relationships are derived from this collection. The process was conducted on about 25 million MEDLINE abstracts and produced more than 26 million semantic relations.

3 METHODS AND MATERIALS

3.1 Server Architecture

An instance of CupQ uses two networked servers. A dedicated storage server is used to maintain persistent information, including a MySQL database, a MongoDB database, and other files. The storage server

also performs operations relating to data download and extraction. Another server containing high memory capacity is used for tokenization, embedding, indexing, searching, and website hosting. The storage server contains an Intel Core i7-4790K 4.0 GHz processor, Ballistix Sport 32 GB DDR3 RAM, and a Samsung 850 Evo 1 TB SSD. The memory server is a Dell R710 with dual Intel Xeon X5687 3.6 GHz processors, 288 GB PC3-10600R RAM, and a Samsung 850 Evo 256 GB SSD.

3.2 Data Download and Extraction

MEDLINE/PubMed data is downloaded via FTP as a directory of compressed XML files. MD5 checksums are compared to ensure file integrity. Specific XML elements related to title, abstract, journal, authors, and publication date are parsed and inserted into a MongoDB collection. The most recent journal information from ScimagoJR and Journal Citation Reports is also downloaded. Documents published in journals listed in the medicine subject area of ScimagoJR are labeled. Each document in this subset is assigned the JIF of its publishing journal. Subsequent operations are performed only on this document subset.

3.3 Tokenization and Embedding

Tokens are extracted from titles and abstracts by splitting text on space and hyphen characters. The LuiNorm API is used for token normalization. Stopwords, except for those fully capitalized, are removed. Then, the Genism library is used to run Word2Vec, generating a vector representation for each token (Řehůřek and Sojka, 2010). Vectors of 100 elements are produced using skip-gram and a window size of 100 without sentence boundaries for 10 epochs. Embeddings for document titles are computed as the sum of each token embedding multiplied by the log ratio of the corpus size to the number of documents containing the token.

3.4 Inverted Indexing

A Java hash map with keys as integers and values as integer array lists is instantiated. Keys represent numeric token identifiers (TIDs). Values represent document PubMed identifiers (PMIDs). For each document, a hash set of title and abstract TIDs is created. The document PMID is added to the array list for each TID. Key-value pairs are stored in a MySQL table comprised of two integer columns, the first for TIDs and the second for PMIDs, with the primary key set

over both columns. New MEDLINE/PubMed documents are automatically downloaded, processed, and indexed on a weekly basis.

3.5 Document Retrieval

A Java search server tokenizes the search string and computes a weighted sum vector representation. The token contained in the fewest documents is passed to the inverted index, which returns a list of PMIDs. Only results written in English and containing all search tokens are retained. Errata, retracted documents, and documents published before the year 1990 are removed. Document information, including publication date, publication type, title embedding, and TIDs are stored in an object array. Results are organized by publication type into array lists for reviews, guidelines, and studies.

After assigning documents into publication categories, a relevance score is computed for each document. A different relevance calculation is used for each publication category. Document lists are then sorted by relevance in descending order. The top 500 results are retained and cached into a MySQL table. A sublist containing results to be displayed for the user's requested page number is obtained. Display information including title, abstract, author abbreviations, journal ISO abbreviation, and publication year is retrieved from disk. Search results are returned to the web server as a JSON payload for HTML rendering.

The document relevance score is the sum of several min-max normalized subscores multiplied by empirically configured boosting factors. A semantic score is computed as the cosine similarity between the query vector and the title vector. A title count score is set to one if a title contains all search tokens and zero otherwise. A date score is computed as an estimated number of days. A journal score is set to the JIF. If a document is published over twenty years ago, the relevance score is fractioned by a tenth. If any of the subscores are zero, then the relevance score is zero. Different sets of boosting factors are used for each category (Table 1).

4 RESULTS

4.1 User Interface

CupQ provides a simple user interface that includes a search bar for entering queries and a tab bar for selecting publication categories (Figures 1–2).
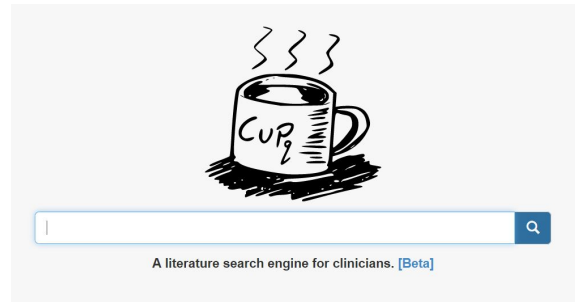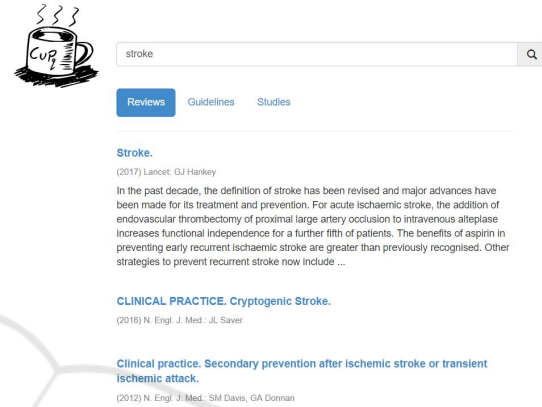

Figure 1: CupQ home page.

Figure 2: CupQ results page showing the top three results for the query "stroke."

4.2 Search Results

The top ten results for several queries and filters were compared between CupQ and PubMed. Note that the PubMed sidebar does not allow for selection of document types to be excluded. For example, it does not allow inclusion of documents that are reviews but not systematic reviews. Defining document types using advanced search strings in PubMed returns different results than selecting document types in the sidebar, perhaps because the Best Match ranking algorithm weighs document types in the search string whereas the sidebar selection behaves as a simple binary filter. Search result comparisons were made with the PubMed sidebar because its interface is more similar to the CupQ tab bar. Searches were performed on January 28, 2019.

4.2.1 Myocardial Infarction (Reviews)

For the query "myocardial infarction," CupQ returned review results from high impact factor journals, including *New England Journal of Medicine* (JIF = 79.26), *Lancet* (JIF = 53.254), and *BMJ* (JIF = 23.562) (Table 2). The titles of the first two results, "Acute myocardial infarction," were highly relevant to the query with a cosine similarity of 0.986. Both

Table 1: Publication Category Boosting Factors.

Category	Title Cosine	Title Count	Date	Journal
Reviews	4	3	1	2
Guidelines	6	8	1	4
Studies	3	5	1	1

documents were published in 2017 issues of *New England Journal of Medicine* and *Lancet*. Other results referenced common concepts related to myocardial infarction, including coronary reperfusion strategies, percutaneous coronary intervention, electrocardiogram, and ST-segment elevation. Incorporation of JIF and Word2Vec query-title cosine similarity may explain the effective prioritization of high impact factor journals and titles containing semantically related concepts to myocardial infarction.

PubMed contrastingly returned no results from *New England Journal of Medicine*, *Lancet*, or *BMJ* (Table 3). The title of the first result was less relevant to the query with a cosine similarity of 0.832. Moreover, the first result was published in an unranked journal by JIF. Although there was a document with the highly relevant title "Acute myocardial infarction," it was published in a 2013 issue of *Disease-A-Month*, a relatively low impact factor journal (JIF = 0.891). PubMed did not return the newer documents with the same title from *New England Journal of Medicine* and *Lancet*. In addition, only one PubMed title referenced an aforementioned common topic related to myocardial infarction, percutaneous coronary intervention. PubMed and CupQ shared no common results. These observations suggest that PubMed may not effectively incorporate JIF in the context of query-title semantic similarity.

4.2.2 Depression (Guidelines)

There were more similarities between CupQ and PubMed for the query "depression" when searching for guidelines (Tables 4–5). The same document published in a 2016 issue of *JAMA* (JIF = 47.661) appeared as the top result for both search engines. However, PubMed lacked the result "Screening for Depression in Children and Adolescents: U.S. Preventive Services Task Force Recommendation Statement," published in a 2016 issue of *Annals of Internal Medicine* (JIF = 19.384). This was unusual behavior because PubMed was able to return a document with the same title and year, albeit from a lower impact factor journal, *Pediatrics* (JIF = 5.515). Unlike PubMed, CupQ may return relevant results by estimating importance via JIF.

All result titles in CupQ contained the query "depression." A problem with PubMed was that the ti-

tle of the fourth result did not contain the query. Although this result was published within the last two years in a high impact factor journal, *CA: A Cancer Journal for Clinicians* (JIF = 244.585), it did not specifically focus on depression. This document encompassed strategies for addressing multiple conditions in patients with breast cancer, including chemotherapy-induced nausea, vomiting, and peripheral neuropathy. Although this document may be more appropriate as a top ten result for the query "depression breast cancer," it addresses too many topics other than depression to be a top ten result for the query "depression."

4.2.3 Stroke (Studies)

When searching for studies about stroke, all result titles from CupQ and PubMed contained the query (Tables 6–7). CupQ only returned results from *New England Journal of Medicine* whereas PubMed returned no results from this journal. The first result returned by PubMed was published in *Clinical Neurology and Neurosurgery* (JIF = 1.736). The highest impact factor journal returned by PubMed was *Lancet Neurology* (JIF = 27.144). The first result from CupQ was published in 2018 whereas the first result from PubMed was published 2017. Moreover, CupQ results were published from 2017 to 2018 whereas PubMed results were published from 2007 to 2018. CupQ may prioritize recent, high impact factor results whose titles contain the query.

5 DISCUSSION AND CONCLUSION

Search engine performance can be assessed through a variety of approaches. Precision and recall can be measured, assuming a binary relevance model and an existing standard for relevance (Hawking et al., 2001). User task studies may demonstrate performance with respect to specific search objectives but may require statistical adjustment for prior user experience with comparative search tools (Taksa et al., 2008). Click-through rates may provide another indication of performance given a high volume of web traffic (Fiorini et al., 2018). This paper qualitatively compared re-

Table 2: Reviews returned by CupQ for the query "myocardial infarction."

No	Title	Journal	Year
1	Acute Myocardial Infarction.	New England Journal of Medicine	2017
2	Acute myocardial infarction.	Lancet	2017
3	Myocardial infarction due to percutaneous coronary intervention.	New England Journal of Medicine	2011
4	Primary PCI for myocardial infarction with ST-segment elevation.	New England Journal of Medicine	2007
5	Acute myocardial infarction.	Lancet	2008
6	Use of the electrocardiogram in acute myocardial infarction.	New England Journal of Medicine	2003
7	Future treatment strategies in ST-segment elevation myocardial infarction.	Lancet	2013
8	Reperfusion strategies in acute myocardial infarction and multivessel disease.	Nature Reviews Cardiology	2017
9	Coronary microvascular obstruction in acute myocardial infarction.	European Heart Journal	2016
10	Management of patients after primary percutaneous coronary intervention for myocardial infarction.	BMJ	2017

Table 3: Reviews returned by PubMed for the query "myocardial infarction."

No	Title	Journal	Year
1	Myocardial infarction with non obstructive coronary arteries (MINOCA): a whole new ball game.	Expert Review of Cardiovascular Therapy	2017
2	Type 2 myocardial infarction due to supply-demand mismatch.	Trends in Cardiovascular Medicine	2017
3	Assessment and classification of patients with myocardial injury and infarction in clinical practice.	Heart	2017
4	Multivessel versus culprit lesion only percutaneous coronary intervention in cardiogenic shock complicating acute myocardial infarction: A systematic review and meta-analysis.	European Heart Journal: Acute Cardiovascular Care	2018
5	Exosomes and cardiac repair after myocardial infarction.	Circulation Research	2014
6	Acute myocardial infarction.	Disease-A-Month	2013
7	Perioperative myocardial infarction/injury after noncardiac surgery.	Swiss Medical Weekly	2015
8	MicroRNAs in myocardial infarction.	Nature Reviews Cardiology	2015
9	Galectin-3 and post-myocardial infarction cardiac remodeling.	European Journal of Pharmacology	2015
10	Type 2 myocardial infarction: the chimaera of cardiology?	Heart	2015

sults between CupQ and PubMed for specific queries and publication categories. Because CupQ was recently launched in January 2019, future work will include analyses of click-through rates when there is significant traffic.

The CupQ ranking algorithm prioritizes title relevance, JIF, and publication date. It assumes that users place the most emphasis on title content when determining the relevance of a result. It also assumes that users weigh the reliability and importance of information, represented by JIF, either greater than or equal to the recency of information. Although JIF is not neces-

sarily representative of individual articles in a journal, it may serve as a useful approximation for physicians who may have limited time to search for information (Garfield, 2006; Seglen, 1997; Saha et al., 2003). In addition, CupQ only returns information published in journals that are listed in the medicine subject area of ScimagoJR. This unique implementation of title relevance, JIF, publication date, and journal category may enable CupQ to return relevant clinical information.

Table 4: Guidelines returned by CupQ for the query "depression."

No	Title	Journal	Year
1	Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement.	JAMA	2016
2	Confronting depression and suicide in physicians: a consensus statement.	JAMA	2003
3	Screening for Depression in Children and Adolescents: U.S. Preventive Services Task Force Recommendation Statement.	Annals of Internal Medicine	2016
4	Screening for depression in adults: U.S. preventive services task force recommendation statement.	Annals of Internal Medicine	2009
5	Screening for Depression in Children and Adolescents: US Preventive Services Task Force Recommendation Statement.	Pediatrics	2016
6	European Psychiatric Association Guidance on psychotherapy in chronic depression across Europe.	European Psychiatry	2016
7	Management of Depression in Patients With Cancer: A Clinical Practice Guideline.	Journal of Oncology Practice	2016
8	Screening for Depression in Adults: Recommendation Statement.	American Family Physician	2016
9	Evidence-based interventions to improve the palliative care of pain, dyspnea, and depression at the end of life: a clinical practice guideline from the American College of Physicians.	Annals of Internal Medicine	2008
10	Clinical pathway for the screening, assessment and management of anxiety and depression in adult cancer patients: Australian guidelines.	Psycho-oncology	2015

Table 5: Guidelines returned by PubMed for the query "depression."

No	Title	Journal	Year
1	Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement.	JAMA	2016
2	Consensus Recommendations for the Clinical Application of Repetitive Transcranial Magnetic Stimulation (rTMS) in the Treatment of Depression.	Journal of Clinical Psychiatry	2018
3	European Psychiatric Association Guidance on psychotherapy in chronic depression across Europe.	European Psychiatry	2016
4	Clinical practice guidelines on the evidence-based use of integrative therapies during and after breast cancer treatment.	CA: A Cancer Journal for Clinicians	2017
5	Depression: The Treatment and Management of Depression in Adults (Updated Edition).	National Collaborating Centre for Mental Health (UK)	2010
6	ACG Clinical Guideline: Preventive Care in Inflammatory Bowel Disease.	American Journal of Gastoenterology	2017
7	Screening for Depression in Children and Adolescents: US Preventive Services Task Force Recommendation Statement.	Pediatrics	2016
8	Management of Depression in Patients With Cancer: A Clinical Practice Guideline.	Journal of Oncology Practice	2016
9	Confronting depression and suicide in physicians: a consensus statement.	JAMA	2003
10	Screening for Depression in Adults: Recommendation Statement.	American Family Physician	2016

Table 6: Studies returned by CupQ for the query "stroke."

No	Title	Journal	Year
1	MRI-Guided Thrombolysis for Stroke with Unknown Time of Onset.	New England Journal of Medicine	2018
2	Clopidogrel and Aspirin in Acute Ischemic Stroke and High-Risk TIA.	New England Journal of Medicine	2018
3	Rivaroxaban for Stroke Prevention after Embolic Stroke of Undetermined Source.	New England Journal of Medicine	2018
4	Tenecteplase versus Alteplase before Thrombectomy for Ischemic Stroke.	New England Journal of Medicine	2018
5	Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging.	New England Journal of Medicine	2018
6	Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct.	New England Journal of Medicine	2018
7	Patent Foramen Ovale Closure or Antiplatelet Therapy for Cryptogenic Stroke.	New England Journal of Medicine	2017
8	Long-Term Outcomes of Patent Foramen Ovale Closure or Medical Therapy after Stroke.	New England Journal of Medicine	2017
9	Patent Foramen Ovale Closure or Anticoagulation vs. Antiplatelets after Stroke.	New England Journal of Medicine	2017
10	Cluster-Randomized, Crossover Trial of Head Positioning in Acute Stroke.	New England Journal of Medicine	2017

Table 7: Studies returned by PubMed for the query "stroke."

No	Title	Journal	Year
1	Hereditary cerebral small vessel disease and stroke.	Clinical Neurology and Neurosurgery	2017
2	Imaging Markers of Post-Stroke Depression and Apathy: a Systematic Review and Meta-Analysis.	Neuropsychology Review	2017
3	Role of Total, Red, Processed, and White Meat Consumption in Stroke Incidence and Mortality: A Systematic Review and Meta-Analysis of Prospective Cohort Studies.	Journal of the American Heart Association	2017
4	Endarterectomy achieves lower stroke and death rates compared with stenting in patients with asymptomatic carotid stenosis.	Journal of Vascular Surgery	2017
5	The Course of Activities in Daily Living: Who Is at Risk for Decline after First Ever Stroke?	Cerebrovascular Diseases	2017
6	Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis.	Lancet Neurology	2009
7	Acupuncture lowering blood pressure for secondary prevention of stroke: a study protocol for a multicenter randomized controlled trial.	Trials	2017
8	Decreased Serum Brain-Derived Neurotrophic Factor May Indicate the Development of Poststroke Depression in Patients with Acute Ischemic Stroke: A Meta-Analysis.	Journal of Stroke and Cerebrovascular Diseases	2018
9	Aerobic Exercises for Cognition Rehabilitation following Stroke: A Systematic Review.	Journal of Stroke and Cerebrovascular Diseases	2016
10	Types of stroke recurrence in patients with ischemic stroke: a substudy from the PRoFESS trial.	International Journal of Stroke	2014

ACKNOWLEDGEMENTS

REFERENCES

Agoritsas, T., Iserman, E., Hobson, N., Cohen, N., Cohen, A., Roshanov, P. S., Perez, M., Cotoi, C., Parrish, R., Pullenayegum, E., et al. (2014). Increasing the quantity and quality of searching for current best evidence to answer clinical questions: protocol and intervention design of the macplus fs factorial randomized controlled trials. *Implementation Science*, 9(1):125.

Davies, K. S. (2011). Physicians and their use of information: a survey comparison between the united states, canada, and the united kingdom. *Journal of the Medical Library Association: JMLA*, 99(1):88.

Fiorini, N., Canese, K., Starchenko, G., Kireev, E., Kim, W., Miller, V., Osipov, M., Kholodov, M., Ismagilov, R., Mohan, S., et al. (2018). Best match: new relevance search for pubmed. *PLoS biology*, 16(8):e2005343.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Jama*, 295(1):90–93.

Hawking, D., Craswell, N., Bailey, P., and Griffihs, K. (2001). Measuring search engine quality. *Information Retrieval*, 4(1):33–59.

Ho, G. J., Liew, S. M., Ng, C. J., Shunmugam, R. H., and Glasziou, P. (2016). Development of a search strategy for an evidence based retrieval service. *PloS one*, 11(12):e0167170.

Lindsey, W. T. and Olin, B. R. (2013). Pubmed searches: Overview and strategies for clinicians. *Nutrition in Clinical Practice*, 28(2):165–176.

Lu, Z. (2011). Pubmed and beyond: a survey of web tools for searching biomedical literature. *Database*, 2011.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Ohta, T., Miyao, Y., Ninomiya, T., Tsuruoka, Y., Yakushiji, A., Masuda, K., Takeuchi, J., Yoshida, K., Hara, T., Kim, J.-D., et al. (2006). An intelligent search engine and gui-based efficient medline search tool based on deep syntactic parsing. *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 17–20.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

Rindflesch, T. C., Kilicoglu, H., Fiszman, M., Rosemblat, G., and Shin, D. (2011). Semantic medline: An advanced information management application for biomedicine. *Information Services & Use*, 31(1-2):15–21.

Russell-Rose, T. and Chamberlain, J. (2017). Expert search strategies: the information retrieval practices of healthcare information professionals. *JMIR medical informatics*, 5(4):e33.

Saha, S., Saint, S., and Christakis, D. A. (2003). Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*, 91(1):42.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *Bmj*, 314(7079):497.

Taksa, I., Spink, A., and Goldberg, R. (2008). A task-oriented approach to search engine usability studies. *JSW*, 3(1):63–73.

White, L., Togneri, R., Liu, W., and Bennamoun, M. (2015). How well sentence embeddings capture meaning. In *Proceedings of the 20th Australasian Document Computing Symposium*, page 9. ACM.