A Task-based Comparison of Linguistic and Semantic Document Retrieval Methods in the Medical Domain

Mohammad Shafahi[1], Qing Hu[2,3], Hamideh Afsarmanesh[1], Zhisheng Huang[2], Annette ten Teije[2], and Frank van Harmelen[2]

[1] Informatics Institute, Faculty of Science, University of Amsterdam, Science Park 904, Amsterdam, The Netherlands,
{m.shafahi,h.afsarmanesh}@uva.nl

[2] Department of Computer Science, VU University Amsterdam, De Boelelaan 1081, Amsterdam, The Netherlands
{qhu400,huang,annette,Frank.van.Harmelen}@cs.vu.nl

[3] College of Computer Science and Technology, Wuhan Univesity of Science and Technology, Wuhan, China

Abstract. Text-based and semantics-based methods are both studied intensively as methods for document retrieval. In order to gain insight in the respective merits of these two approaches, we have performed a controlled experiment where we executed a real-life task using both text-based and semantics-based techniques.
To maximise the lessons that we could draw about the two approaches, we have performed an experiment where we used the same task (searching papers from the scientific literature needed for updating a medical guideline), the same test-case (updating the 2004 Dutch national breast-cancer guideline), the same gold standard (the updated 2012 Dutch national breast-cancer guideline) and the same corpus (PubMed). We then performed this task using two different methods: retrieving papers based on keywords (text-based approach) and retrieving papers based on semantic annotations (semantics-based approach). Based on this experiment, we discuss the insights that we gained from this dual set of experiments.

Keywords: document retrieval, keyword search, semantic annotation, concept-based search, relation-based search

1 Introduction

Many domains involve retrieving items from large text corpora. Examples are searching for web-pages, searching scientific literature, or question answering over a text corpus. Classical information retrieval techniques use text-based methods for selecting and ranking the most relevant documents for a query. Typical examples are N-gram similarity, vector-space models over words, probabilistic language models, etc. [10].

There is an increasing interest in the use of semantic methods for retrieving items from large corpora (eg [5]). In such techniques, words from both the query and the items in the corpus are mapped to concepts and relations in a knowledge source (typically an ontology), and retrieval is then based on semantic proximity in the background ontology.

Attempts to understand the circumstances that determine the effectiveness of each approach have a long history (e.g. [9]). This paper contributes to this understanding of the respective merits of text-based and semantics-based information retrieval.

To this end, we have performed an experiment where we used the same task (searching papers from the scientific literature needed for updating a medical guideline), the same test-case (updating the 2004 Dutch national breast-cancer guideline), the same gold standard (the updated 2012 Dutch national breast-cancer guideline) and the same corpus (PubMed). We then performed this task using two different methods: retrieving papers based on keywords (text-based approach) and retrieving papers based on semantic annotations (semantics-based approach)that in this case applies concepts and relations extracted from corpus. Based on this experiment, we discuss the insights that we gained from this dual set of experiments.

The rest of the paper is organized as follows. Section 2 describes the evaluation-task. Section 3 describes our experimental setup (gold standard, corpora, metrics). Sections 4 and 5 describe the text-based and semantics-based methods respectively. Section 6 interprets and concludes the results from our experiments.

2 Description of the task

A medical guideline, alternatively called clinical guideline, is a document which is designed with the aim of guiding medical decisions and criteria for diagnosis, management, and treatment in specific areas of health-care. Medical guidelines have been proved to be valuable for clinicians, nurses, and other health care professionals [16]. Evidence-based medical guidelines are developed based on the best available evidence in biomedical science and clinical practice. Guideline recommendations in evidence-based medical guidelines are annotated with their underlying evidence and their evidence classes.

Evidence-based medical guidelines are expected to be updated regularly and frequently, so that medical guidelines can accommodate the latest research findings. However, such a requirement on timely and regularly update of a medical guideline has been proved to be difficult for two reasons. First, the number of medical publications and the size of medical information is very large (for example, PubMed[4] alone contains more than 24 million citations for biomedical literature from MEDLINE[5]). Second, large volumes of new medical findings occur every day (PubMed is growing at a rate of 750.000 papers per year[6], ie roughly one new paper every minute).

[4] http://www.ncbi.nlm.nih.gov/pubmed
[5] http://www.nlm.nih.gov/bsd /pmresources.html
[6] `http://www.nlm.nih.gov/pubs/factsheets/medline.html`

Consequently, it usually takes about five years to release a new update of a medical guideline. However, such an update frequency significantly lags behind the occurrence of new medical findings. Thus, automatically finding new and relevant evidences for timely and regularly updates of medical guidelines has become one of the important challenges in medical information retrieval.

We have taken this task of finding medical publications which are relevant for updating a given guideline as our benchmark. A medical guideline is usually a document of more than hundred pages of text and tables. The essence of the guideline is captured in numerous recommendations (called "conclusions"), each of them in the form of a short paragraph, for example:

> "A descriptive study found that women who undergo breast reconstruction immediately following the mastectomy are more satisfied with the aesthetic result and experience greater psychosocial well being than women who undergo secondary reconstruction."
> (1st conclusion in Section 1.2.6 (on page 25/117), from the Dutch National Breast Cancer Guideline, 2004 - considered as conclusion nr. 12 in our experiments)

Such guideline conclusions are typically annotated with somewhere between 1 to 10 citations to the medical literature that provide the evidence for the conclusion. For example, the above recommendation was supported by three citations to the literature, from the years 1984, 1995 and 2000.

Our benchmark task is now to find for each conclusion in a guideline all the recent medical publications which are relevant for making an updated version of that conclusion.

3 Description of guideline, corpus, gold standard, and metrics

Guideline: For investigating the behaviour of both the text-based and the semantics-based methods of document retrieval for the purposes of finding new evidences to update the conclusions of a medical guideline, we have selected the Dutch National Guideline for Breast Cancer from 2004 (version 1.0, [12][7]). The guideline is a document of 117 pages, listing around 50 recommendations ("conclusions") in total, each the length of 1 to 2 sentences. This guideline is in daily use nationwide.

Corpus: As the corpus for our text-based experiment, we have used the PubMed query service. This service allows querying of titles and abstracts of 24 million publications from the biomedical scientific literature.
As the corpus for our semantics-based experiment we have used the query service of BioMed Xplorer [15]. BioMed Xplorer is built on top of SemMedDB[8] [8], containing semantic annotations in the form of triples which have been extracted

[7] For our experiments we used a certified English translation of the document
[8] `http://skr3.nlm.nih.gov/SemMedDB/`

from PubMed and annotated with the PubMed-ID of the paper(s) from which the relation was extracted. In our experiments we have used a version of SemMedDB that hosts more than 70 million statements extracted from PubMed papers. Furthermore, the concepts and relations that form the statements in SemMedDB have been linked to corresponding concepts and relations from Linked Life Data [11] and Bio2RDF [4].

Gold Standard: When our search methods (either text-based or semantics-based) search for publications in PubMed, which are relevant for updating a particular guideline recommendation, how should we measure their success? In other words, how can it be decided whether the returned publications are indeed those relevant for updating the guideline? For this purpose, we use the updated 2012 revision of the 2004 Dutch National Breast Cancer guideline [13] and create our gold standard from it. In the revised 2012 guideline, we have identified 16 corresponding and/or matching example conclusions from 2004, that while each of the pairs addresses the same or similar subject, they have distinct revised statements in the 2012 version of the guideline. Thus indicating a clear revision of their 2004 conclusion. All other conclusions in the 2012 version were either directly copied from the 2004 version and not updated, or they were entirely new and could not be interpreted as a revision of a conclusion from 2004. For these 16 conclusions, we have then identified the *new* publication evidences that were listed from them (sometimes a revision also listed some of the evidences from the previous version (these we will call *hits* for our search results). These *hits* (all evidences that were actually used in the 2012 revision) are the gold standard for our search methods: ideally the search methods would suggest all the hits and only the hits.

Metric: In practice, of course, our search methods will not return all the hits, and they will return not only the hits but also other papers. Let *count* be the number of papers returned by the search method and *hit* be the number of papers that have been referenced in the conclusion of the guideline, then we would like *count* to be as small as possible, while containing the maximal number of *hits*. In a realistic scenario, a guideline revision committee may consider many dozens of papers for a single conclusion, but certainly not more than a few hundred, putting a stringent upper bound on a realistic value for *counts*. Now let *Relevant hits* (RH) be the number of papers returned by the search method that have been referenced in at least one of the conclusions of the guideline. We would also then like RH to be maximal for each concept or relationship used for querying. In other words, we would like $RH/count$ to approach 1.

4 Description and results of the text-based method

In the text-based method, PubMed queries are generated in the form of medical terms which appear in a guideline conclusion. Construction of this query proceeds in the following steps [7]:

1. We use Xerox's NLP tool [1, 2] to identify the medical terms which appear in the guideline conclusion (formulation from 2004).

2. In the same way we collect medical terms from the heading of the guideline-section in which the conclusion appears.

3. We use a co-occurrence based ranking measure to rank the extracted terms (the ranking measure is computed by counting co-occurences of the term in the PubMed corpus)

4. We construct a query as the conjunction of the top k ranked terms, where k is determined by heuristically balancing the size of *counts* and *hits*.

In [6], we develop a heuristic function which considers the balance of the hits of original evidences and the counts to evaluate the search results, in order to find the best answer for k. These results are then compared against the evidence items for the corresponding conclusion in the 2012 version of the guideline, giving us the score of the query in terms of the number of *hits*. We repeat this procedure for each of the 16 recommendations from the 2004 guideline which have a revised version in the 2012 guideline. The results of this experiment are reported in

Conclusion	Goal	*Hits*	Count	%
C1	5	2	60	40%
C2	2	1	166	50%
C3	4	1	36	25%
C4	14	0	49	0%
C5	2	1	28	50%
C6	2	0	33	0%
C7	2	1	333	50%
C8	8	3	140	38%
C9	2	1	89	50%
C10	5	3	1628	60%
C11	5	3	281	60%
C12	3	0	82	0%
C13	5	5	9911	100%
C14	3	1	72	33%
C15	2	0	372	0%
C16	2	1	324	50%
Total	66	23	13604	
Nr. of *hits* > 0		12		
Average	4	1	850	35%

Table 1. Results of the text-based method.

table 1. Conclusions are numbered C1 to C16. The table shows that the text-based method found some hits for 11 out of the 16 guideline conclusions, with an average of 35% of all evidence items retrieved. The total number of returned PubMed entries (the count) ranges from below 100 (reasonable) to a few hundred (problematic), with two outliers over 1000.

5 Description and results of the semantic-based methods

In the semantic-based method, the BioMed Xloprer [15] is searched based on the semantic concepts and relations that are extracted from the text of the guideline. The 16 conclusions from the guideline are used as the base also for this method, but here we aim to reach suitable matching criteria for discovering relevant evidences from BioMed Xplorer, using the extracted semantics instead of the terms and keywords. We introduce two semantic-based search methods, namely a concept-based method and a relation-based method. Through some experiments, we also measure the results of our methods against the gold standard.

Aiming to optimize the identification of relevant evidence items for updating the guideline, in our first semantic-based approach we construct queries for BioMed Xloprer based on concepts generated out of the same keywords as those used in the text-based method. We use the Meta-maps tool [3] for mapping the keywords into their relevant concepts, and formulate the queries in SPARQL. This concept-based search approach for evidences and the example experiment for it are further descried in (section 5.1). In our second semantic-based approach, the queries constructed for BioMed Xplorer are based on the relations that we extract out of the text in the guideline related to the 16 conclusions, e.g. the abstracts. We use the SemRep tool [14] to automate the extraction of relations, and construct SPARQL queries to BioMed Xplorer from their conjunction. However, an abstract typically precedes several conclusions at the same time, so the produced results also need to be evaluated against the hits in the corresponding group of conclusions, and not per conclusion. Further description of this method and its two experiments are addressed in (section 5.2).

5.1 Detecting new evidences using concepts extracted from keywords out of the conclusions and headers

In this semantic-based method, BioMed Xplorer [15] queries are generated encapsulating concepts that may represent either the *subject* or the *object* of a triple relationship. The following steps are followed in this approach:

1. Starting with the 25 keywords that are related to 16 conclusions, as identified in section 4, we first apply UMLS medical concepts used by the MetaMap tool [3], to map each keyword into its related set of concept(s).
2. For each generated concept, we formulate a SPARQL query as the conjunction of the identified relationships, addressing the RDF triples in which the concept is either the subject or the object.
3. The BioMed Xplorer is then queried, to search for annotated publications as evidences related to each conclusion.

An example output of the Metamap for the keyword mastectomy follows:

C0024881:Mastectomy
C0191849:Mastectomy (Excision of breast tissue)
C0024886:Mastectomy (Simple mastectomy)

In the case when a keyword is mapped to multiple concepts, as demonstrated above for the keyword "mastectomy", then the "union" of all these concepts is used for formulating the query. We compare our results against corresponding evidence items for each conclusion in the 2012 guideline, calculating the score for the query on each concept, in terms of the number of *hits*. This procedure is repeated for every concept generated from each of the 25 keywords, as extracted from 16 conclusions related to the 2004 guideline. Table 2 reports on the results of this experiment. Please note that the highlighted cells of the table demonstrate from which conclusion a keyword has been extracted. For example the keyword "excision" has been extracted from conclusion 5 (i.e. C5). It also indicates that the search for "excision" through BioMed Xplorer has discovered 1 out of the 2 evidences in the gold standard for this conclusion. Furthermore, the search for the "excision" keyword has also discovered 1 relevant evidence for each conclusion C1, C3, C8, and C16. Over all this keyword has been found relevant in discovery of 5 papers referenced within the 16 conclusions (i.e 5 RH-relevant hits). The latter discovery of evidences indicate that this semantic-based approach can further enhance and benefit from concepts in other conclusions in the guideline, for identifying the needed evidences. We would later on use this fact in section 5.2.

Although the results in table 2 suggest that some selected on the keywords for the text-based method are very suitable for a semantic based method (i.e dcis for C1 that has 100% recall), it also shows that out of these 25 keyword, 12 were not suitable for retrieving any of the goal publications. This is even move visible when noticing that out of the 12 unsuitable keywords, three of them not only have a hit of zero but also a count of zero. Another interesting finding is that for the keywords "resection" and "excision" although in the text-based method they are considered as 2 different entities, in this sematic-based method they are mapped into a common concept and as such they are considered as one entity (hence providing the same results).

5.2 Detecting new evidences using extracted relations from the conclusions and abstracts in the guideline

In this semantic-based method, BioMed Xplorer queries are generated in the form of triples, representing a medical statement, extracted first from guideline conclusions and second from the abstracts corresponding to the conclusions. Construction of this type of query involves the following steps:

1. We use the SemRep tool [14] to extract medical statements from the text which appear in the guideline conclusion (formulation from 2004).
2. We transform the extracted medical statements into a RDF triples.
3. BioMed Xplorer is queried to find the annotated publications for each of the queries.

Example extracted relations from conclusion number 12 (i.e C12) using SemRep

Mammaplasty (C0085076) TREATS Woman (C0043210)
Reconstructive Surgical Procedures (C0524865) TREATS Woman (C0043210)

Keyword	RH	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	Count
dcis	10	5	2	2										1				3615
local excision	6	2	1	1				1							1			3063
radiotherapy	0																	105
recurrence	0																	3938
margin of excision	0																	4795
breast	16	4	1	2	2			1		1	2	2		1	1	1		62069
bct	1									0								524
irradiation	0																	103
survival rate	0																	360
boost	0																	0
resection	5	1			1	1			1								1	149902
excision	5	1			1	1			1								1	149902
age	2							1										43288
primary	0																	587
mastectomy	23	1		1	4				1	1	2	3	1	3	3		1	5357
survival	0																	8
systemic therapy	2							1			1							6097
reconstruction	11												1	4	1	2	2	37261
breast cancer	36			1	9	2	1		5	2	3	2	2	2	1			107960
skin-sparing mastectomy	0																	0
autologous	0																	0
silicone	2															1		4562
breast reconstruction	9												2	2	2		1	2573
complications	3																2	154062
local	0																	391
Goal		5	2	4	14	2	2	2	8	2	5	5	3	5	3	2	2	
Hits		5	2	2	0	1	0	0	0	0	2	3	2	4	2	1	2	
%		100	100	50	0	50	0	0	0	0	40	60	65	80	65	50	100	

Table 2. Results of extracted concepts from keywords method.

These results are then compared against the evidence items for the conclusions in the guideline, giving us the score of the query in terms of the number of *hits*. We repeat this procedure for each of the 16 conclusion from the 2004 guideline.

The results of this experiment are reported in table 3. Please note that only the conclusions for which SemRep managed to extract relationships are shown in the table. Table 3 suggests that although this method is only capable of extracting relationships in 6 out of 16 conclusion cases, when possible this method achieves comparable and in some cases better precision then the text-based methods. It is also interesting to point out that based on the RH results, one can con-

Conclusion	Goal	Extracted relationships	Hit	RH	%	Count
C2	2	Excision TREATS Noninfiltrating Intraductal Carcinoma	0	1	50%	60
C3	4	Adjuvant therapy USES Tamoxifen	0	1	25%	339
		Breast LOCATION_OF Tamoxifen	0	0	0%	37
		Therapeutic procedure TREATS Neoplasm	0	1	25%	3799
C4	14	Modified radical mastectomy PRECEDES Radiation therapy	0	0	0%	8
C9	2	Primary Carcinoma PART_OF Breast	0	0	0%	1425
C12	3	Mammaplasty TREATS Woman	1	2	67%	179
		Reconstructive Surgical Procedures TREATS Woman	0	0	0%	251
C14	3	Reconstructive Surgical Procedures METHOD_OF Mastectomy	0	0	0%	32

Table 3. Results of extracted relations from conclusion text.

clude that the "TREATS" and "USES" relations are the most suited statements for retrieving the proper literature, as other relationship types have a RH of zero in all cases. We have then used the lessons learned in the above experiment, to improve our relation-based approach as follows.

So far, queries are in the form of triples and represent a medical statement/relationship just based on the guideline conclusions. To enhance this approach instead of only using the guideline conclusions we also use the abstracts provided for each set of the conclusions as the input for extracting the triples. In relation to the 16 conclusions these are 7 abstracts in the guideline that we can use to extract more suitable relations, covering all 16 conclusions. Also based on the results in table 3 we have decided to focus only on "TREATS" and "USES" statements for our retrieval task. As such the construction of the query is done as follows:

1. We use the SemRep tool [14] to extract medical statements from the abstracts which appear for the guideline conclusions and the conclusions them selves (formulation from 2004).
2. Out of the resulted medical statements we only select the statements that have "TREATS" or "USES" as their predicate.
3. We transform the filtered set of medical statements into RDF triples.
4. BioMed Xplorer is queried to find the annotated publications for each of the queries.

When evaluating the results of this experiment in table 4 the overall hit rate is lower than the hit rate of the text-based approach (i.e 26 % compared to 35 %). Although comparing abstracts A1 (i.e for Conclusions 1-3) and A3 (i.e for Conclusions 4-9) in table 4, indicates that the longer text considered for each conclusion, the better the results of our approach. It is important to point out that given larger input, this approach can further improve the quality of its results, and improve over the text-based approach, even if removing the problematic relations that have a count of more than 1000).

Abstract	Conclusions	Goal	Hit	%	Extracted relationships	Hit	RH	Count
A1	C1-C3	11	6	55%	Excision TREATS Noninfiltrating Intraductal Carcinoma	1	1	60
					Adjuvant therapy USES Tamoxifen	0	1	339
					Therapeutic procedure TREATS Neoplasm	0	1	3799
					Radiation therapy TREATS Noninfiltrating Intraductal Carcinoma	3	4	145
					Therapeutic procedure TREATS Invasive Carcinoma	0	0	59
					Therapeutic procedure TREATS Noninfiltrating Intraductal Carcinoma	2	3	156
A2	C4-C9	30	4	13%	Pharmacotherapy TREATS Woman	0	1	802
					Pharmacotherapy TREATS Malignant neoplasm of breast	1	2	2420
					Reexcision TREATS Neoplasm	0	0	17
					Radiation therapy TREATS Woman	3	10	664
					Clinical Research USES Clinical Trials, Phase II	0	0	50
A3	C10-C11	10	6	60%	Modified radical mastectomy TREATS Indicated	0	0	0
					Operative Surgical Procedures TREATS Male population group	0	0	525
					Operative Surgical Procedures TREATS Malignant neoplasm of breast	0	4	2198
					Modified radical mastectomy TREATS Neoplasm	0	0	10
					Chemotherapy, Adjuvant TREATS Patients	1	1	2178
					Modified radical mastectomy TREATS Patients	1	1	297
					Radiation therapy TREATS Patients	5	11	19081
A4	C12	3	1	33%	Mammaplasty TREATS Woman	0	2	179
					Reconstructive Surgical Procedures TREATS Woman	0	0	251
					Mammaplasty TREATS Patients	1	3	710
A6	C15	2	0	0%	Mammaplasty USES Prosthesis	0	1	64
					Implantation procedure USES Silicones	0	0	43
					Prosthesis USES Silicones	0	1	355
A7	C16	2	0	0%	Radiation therapy TREATS Patients	0	11	19081
					Reconstructive Surgical Procedures TREATS Woman	0	0	251
					Radiation therapy TREATS Complication	0	1	127
					Mammaplasty USES Prosthesis	0	1	64

Table 4. Results of extracted relations from abstract text.

6 Interpretation of the results and Conclusion

In this paper, we have reported on two kinds of semantic experiments for documents retrieval in the medical domain, namely: a concept-based method and a relation-based method. The results of those experiments are compared against a keyword-based experiment with the same task, the same test-case, the same gold standard, and the same corpus.

From the first experiment (concept based) in section 5.1 which we consider the concepts generated from a keyword that may appear either as a subject or an object in a triple, we can see that the method can find the goal evidences for 11 conclusions out of the 16, compared with the 12 out of 16 evidences obtained when using the text-based method. Also the precision of concept-based approach is still quite low in comparison to the text-based method, namely its counts are still quite large. This could be due to the fact that the concept-based method uses UMLS as its concept ontology, and PubMed uses the MeSH ontology for indexing keywords. As MeSH is covered in UMLS, it might be the case that PubMed performs a concept-based search, based on MeSH, using the keywords provided by our text-based method. Therefore, the text-based approach returns results similar to those of our concept-based method. Investigating the validity of this educated guess is planned as one of our future work.

From the second experiment (relation-based) in section 5.2 however, we achieve good results. In this method, where we extract both the concepts and relations directly from the text of the guideline, we observe that the counts are much smaller than those in the concept-based method. Consequently, with relation-based method, we can achieve much higher precisions, compared to the concept-based method. But the precision is still lower then those of the text-based method. However, we achieve much better results with the goal evidence discovery. With the relation-based method, we discover 14 out of 16, when compared to text-based method that finds 12 out of 16. Furthermore, the total number of hits in the concept-based method (26 hits) is larger than the total number of hits in the text-based method (23 hits). In future work, we would like to further investigate which criteria of the environment (e.g. guideline) would prove to be more suitable for adopting one of these methods.

For the evaluation of the results of text-based methods, we have previously invited three medical professionals from the MAASTRO clinic in the Netherlands to score the guideline update tool with respect to various properties such as functionality, efficiency, usability, reliability and quality of use [7]. We plan to perform a similar evaluation for the results of semantics-based methods as a future work.

References

1. Salah Ait-Mokhtar, Berry De Bruijn, Caroline Hagege, and Pajolma Rupi. Initial prototype for relation identification between concepts, D3.2. Technical report, EURECA Project, 2013.
2. Salah Aït-Mokhtar, Jean-Pierre Chanod, and Claude Roux. Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(2):121–144, 2002.
3. AR Aronson. Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In *Proceedings of AMIA Symposium*, pages 17–21, 2001.

4. F Belleau, MA Nolin, N Tourigny, P Rigault, and J Morissette. Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *Journal of Biomed Inform*, 41(5):706–716, 2008.
5. Julio Gonzalo, Hang Li, Alessandro Moschitti, and Jun Xu. Sigir 2014 workshop on semantic matching in information retrieval. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 1296–1296, New York, NY, USA, 2014. ACM.
6. Qing Hu, Zhisheng Huang, Annette den Teije, and Frank van Harmelen. Detecting new evidence for evidence-based guidelines using a semantic distance method. In *Proceedings of the 15th Conference on Artificial Intelligence in Medicine(AIME 2015)*, 2015.
7. Qing Hu, Zhisheng Huang, Annette ten Teije, Frank van Harmelen, M Scott Marshall, and Andre Dekker. A topic-centric approach to detecting new evidences for evidence-based medical guidelines. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (HealthInf2016)*, 2016.
8. H Kilicogl, D Shin, M Fiszman, G Rosemblat, and TC Rindflesch. Semmeddb: a pubmed-scale repository of biomedical semantic predications. *Bioinformatics*, 28(23):3158–3160, 2012.
9. Karen E. Lochbaum and Lynn A. Streeter. Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing & Management*, 25(6):665 – 676, 1989.
10. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
11. V. Momtchev. Expanding the pathway and interaction knowledge in linked life data. In *International Semantic Web Challenge*, 2009.
12. NABON. Guideline for the treatment of breast carcinoma 2004. Technical report, Nationaal Borstkanker Overleg Nederland (NABON), 2004.
13. NABON. Breast cancer, dutch guideline, version 2.0. Technical report, Integraal kankercentrum Netherland, Nationaal Borstkanker Overleg Nederland, 2012.
14. T.C. Rindflesch and M Fiszman. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, 36(6):462–477, 2003.
15. Mohammad Shafahi, Hayo Bart, and Hamideh Afsarmanesh. Biomed xplorer - exploring (bio)medical knowledge using linked data. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOINFORMATICS 2016)*, 2016.
16. Steven Woolf, Richard Grol, Allen Hutchinson, Martin Eccles, and Jeremy Grimshaw. Clinical guidelines:potential benefits, limitations, and harms of clinical guidelines. *BMJ*, 318(7182):527–530, 1999.