Expert Systems with Applications 40 (2013) 2081–2091

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic retrieval of current evidence to support update of bibliography in clinical guidelines

A. Iruetaguena *, J.J. Garcia Adeva, J.M. Pikatza, U. Segundo, D. Buenestado, R. Barrena

Department of Computer Languages and Systems, Faculty of Informatics, University of the Basque Country UPV/EHU, Paseo Manuel de Lardizabal 1, 20018 Donostia, Spain

ARTICLE INFO

Keywords:
Clinical guidelines
Evidence-based medicine
Bibliometrics
Information retrieval
Literature-based discovery

ABSTRACT

This paper reports on a system developed to support medical experts in the process of updating clinical guidelines by automatically suggesting new articles suitable to the domain under consideration. It follows a comprehensive process based on several consecutive steps in order to (i) identify which articles from the current guideline are eligible to be updated; (ii) retrieve and filter new related articles from MEDLINE; and (iii) select the most relevant resulting articles by applying a scoring algorithm. Extensive validation is based on a set of experiments on 40 guidelines from multiple medical domains. The analysis of results shows a promising prospect as indicated by recall values greater than 90%.

1. Introduction

As defined by the Institute of Medicine, clinical guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (Field & Lohr, 1990). Clinical guidelines are used in evidence-based medicine to address the problem of clinical practice variation (Timmermans & Mauck, 2005). Their dissemination and implementation has become a cornerstone of professional decision support in the past 30 years and has come to play a pivotal role in routine clinical practice (van der Weijden, Boivin, Burgers, Schnnemann, & Elwyn, 2012).

Clinical guidelines are often relied upon by Governments with the aim of supporting the establishment of practice standards, including decisions about allocation of health-related funding (Shapiro, Lasker, Bindman, & Lee, 1993) and promoting care standardisation across the sector (Woolf, Grol, Hutchinson, Eccles, & Grimshaw, 1999). The importance of clinical guidelines will most likely increase in the future as these aspects of consistency will be used to improve quality of care (Graham, Mancher, Miller, Greenfied, & Steinberg, 2011) while keeping costs down.

Although clinical guidelines are useful, they also are subject to some limitations and challenges. The process of creating, reviewing or updating a guideline is laborious and expensive (Grol et al., 1998). For example, ideally guidelines should be updated immediately after new relevant evidence is published (Shekelle, Woolf, Eccles, & Grimshaw, 1999). Therefore, continuous literature revision and updating is required (Rosenfeld & Shiffman, 2009), but this is seldom the case. Instead, guidelines are usually updated based on a certain schedule (e.g. every two years), hence leading to frequent situations where a guideline might be out of date by the time it has been published.

Our objective consisted of offering a method for automatically obtaining new bibliography in order to support the process of the guideline update. Our hypothesis was based on the assumption that such a tool could help to decrease the time currently required to achieve this task, and consequently also reduce the economic costs involved in updating guidelines. The automatic bibliographic update supports this process of creating new versions of clinical guidelines by offering suggestions to panel experts while providing a neutral and transparent point of view.

This paper is organised as follows. Section 2 offers some context for the challenge tackled by this work. Section 3 contains a description about what data were available and how they were used. Section 4 proposes a set of experiments that were carried out in order to validate the work, along with a discussion of the results. The paper concludes with Section 5, which also suggests some ideas for future work.

2. Background

The U.S. National Library of Medicine created the bibliographic database MEDLINE, which currently includes over 19 million articles from 5600 worldwide journals in the broad areas of life sciences and biomedicine. The number of references added per week varies between 10,000 and 20,000 new instances.[1] Fig. 1 depicts the growth of the number of articles indexed by MEDLINE, from the year

* Corresponding author. Tel.: +34 943015108; fax: +34 943015590.
E-mail address: ander.iruetaguena@ehu.es (A. Iruetaguena).

[1] <https://www.nlm.nih.gov/pubs/factsheets/medline.html>.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.10.015

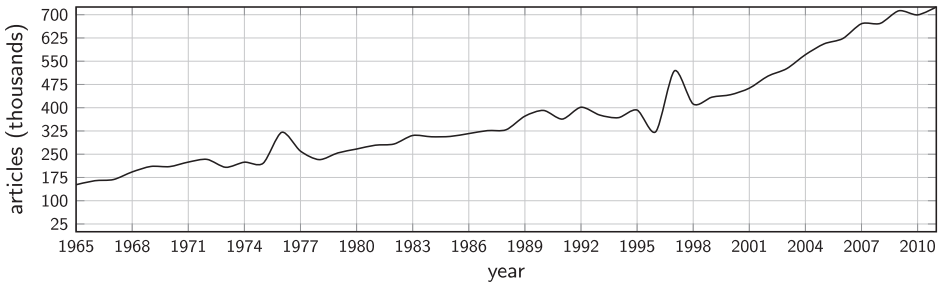

Fig. 1. Number of articles indexed by MEDLINE by year from 1965 to 2011.

1965, when there were 151,635 articles, to the year 2011, when 724,831 instances were indexed.[2]

MEDLINE includes a controlled vocabulary of concepts, called MeSH, for the purpose of supporting both indexing and searching articles. Each article in MEDLINE is indexed by about 10–15 concepts.

Readers can access MEDLINE through the Web-based service PubMed, which is based on the federated search engine Entrez. The latter also offers access to this bibliographic database (e.g. full-text articles, citations, etc.) to third party software components through a Web service, which is described in more detail in Section 3.

Based on this overwhelming and ever increasing amount of bibliographic resources, it becomes clear that medical practitioners must deal with massive amounts of existing knowledge if they want to keep up with the most recent developments in their area. Unfortunately, traditional approaches for literature search employing manual methods of information retrieval are not enough to satisfy this challenge due to some limitations including: (i) the inability to recognize connections between articles, (ii) the obtained results are too restrictive (i.e. low recall), and (iii) the results are too broad (i.e. low precision) (Ramampiaro, 2010).

Literature based discovery (LBD) is a current area of research that tries to address these limitations by finding hidden, new connections between references embedded in published literature. In our case, we aimed at bibliography contained in well-established clinical practice guidelines.

It was Swanson, 1986, Swanson (1988), Swanson (1990), Swanson (1991), Swanson (2008) who first proposed the discovery of unseen relationships among references in a bibliographic collection. Swanson defined LBD as the process of finding complementary structures in disjoint science literature. In his multiple studies, he demonstrated how this discovery of new knowledge from the literature could contribute to providing solutions to particular medical problems. This approach differs from the one followed by empirical sciences where the discovery of new knowledge is obtained through laboratory experiments and observation of nature (Swanson, 1988).

In this context, a standard search request to MEDLINE retrieves the explicit knowledge that pertains to the query. However, while the implicit knowledge inferred from the retrieved information might be valuable, it is not directly available. Swanson's ABC model of discovery (Swanson, 1991) tackled this challenge by establishing the hypothesis that if the entities (such as variables or concepts) A and B are directly related, and so are the entities B and C, then A and C might be indirectly related.

In recent years LBD has been successfully applied to the process of finding new therapeutic applications of certain drugs, such as thalidomide to treat myasthenia gravis, chronic hepatitis C, Helicobacter pylori-induced gastritis, and acute pancreatitis (Weeber et al., 2003), or chlorpromazine to reduce cardiac hypertrophy (Wren, Bekeredjian, Stewart, Shohet, & Garner, 2004).

[2] Data retrieved from https://www.nlm.nih.gov/bsd/index_stats_comp.html.

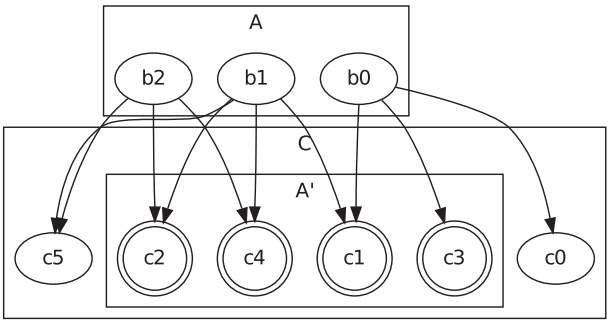

Fig. 2. An example of applying Swanson's ABC model of literature discovery to clinical guidelines.

Fig. 2 shows an example of how we decided to apply this approach in order to discover new bibliography from existing clinical guidelines. In this context, A represents a clinical guideline and B represents the corresponding guideline bibliography composed by a collection of articles $B = \{b_0, b_1, \ldots, b_n\}$. For each of these articles b_i, $0 \leqslant i \leqslant n$ it is possible to find a group of related articles $C_i = \{c_i^0, c_i^1, \ldots, c_i^m\}$. After obtaining all the articles related to B, $\langle C_0, C_1, \ldots, C_n \rangle$, their union results in a long collection of unique articles $C = \{c_0, c_1, \ldots, c_k\}$ related to B. Our work was based on the hypothesis that the guideline A could have its bibliography B updated by finding its related articles C and selecting the most relevant article instances, as represented by A' in Fig. 2.

3. Materials and methods

This section describes the key resources and techniques employed to test the hypothesis proposed in Section 2. Fig. 3 provides a graphical overview of the process, which is divided into several steps that will be detailed in the next sections.

Section 3.1 describes the first step, which consisted of randomly selecting a set of clinical guidelines as the basis of the experiments. Next, for each of the guidelines, the associated bibliography had to be retrieved from MEDLINE by means of the Entrez utilities, as will be explained in Section 3.2. Section 3.3 provides details on how these retrieved articles were filtered in order to maintain a minimum level of quality. The remaining articles were scored by applying the approach proposed in Section 3.4, and only those above a certain threshold were kept. The general algorithm is detailed in Section 3.5.

3.1. Clinical guidelines

National Guideline Clearinghouse is a database of evidence-based clinical practice guidelines maintained by the Agency for Healthcare Research and Quality of the U.S. Department of Health and Human Services. By July 2011, it stored 2548 different

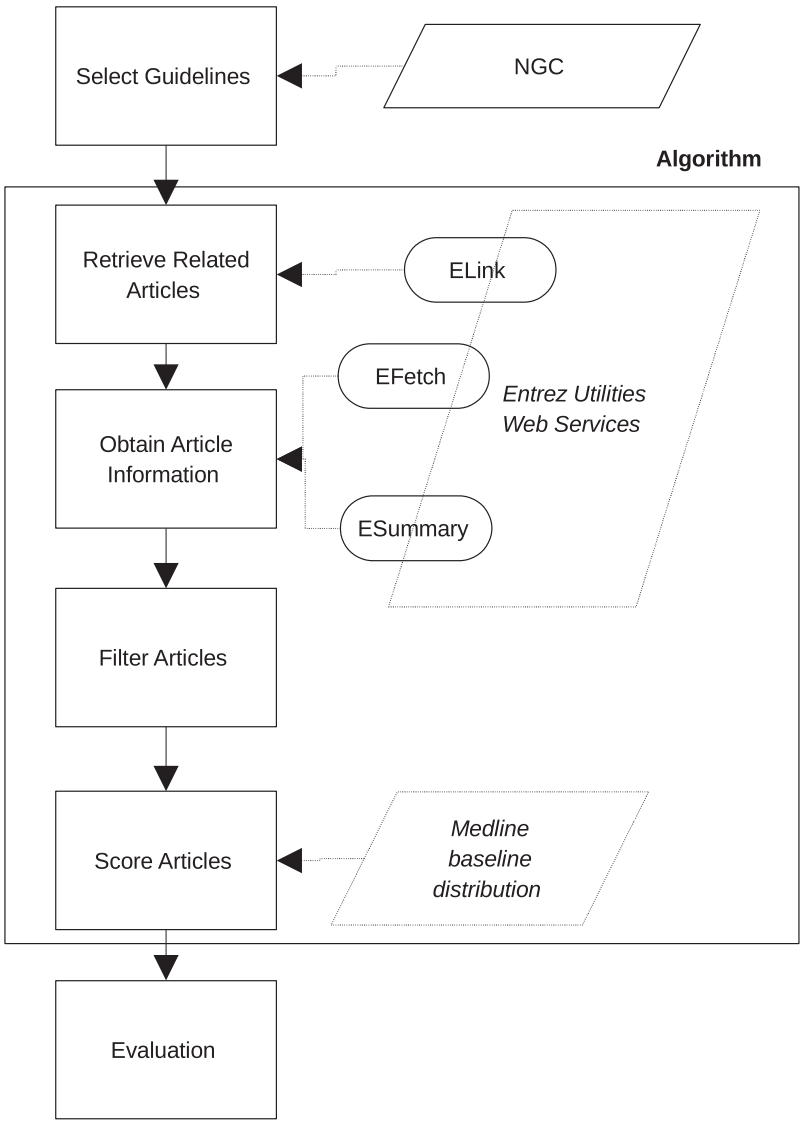

Fig. 3. Experiment overview where major steps are described.

guidelines, of which 804 had at least two different versions that could be accessed from the University of the Basque Country. 5% of the guidelines that have two versions (i.e. 40 instances) were randomly selected for this experiment.

These selected guidelines are shown in Tables A.1 and A.2, where Column V_1 indicates the reference to the first version of the guideline, while column V_2 corresponds to the reference to the updated version. These references are listed in Tables A.3–A.7. N_1 is the number of bibliographic entries from V_1 that are included in MEDLINE. N_2 is the number of bibliographic entries from V_2 that are included in MEDLINE and that have been published after the publication date of V_1. Finding the articles that constitute the group N_2 is the target of this experiment. The maximum and minimum numbers of bibliographic entries in V_1 version were 904 and 4 respectively, while the mean value was 115.9 ± 156.5. In V_2 those values were 305, 4 and 63.67 ± 63.2 respectively.

3.2. Bibliographic information access

Once the experiment guidelines were selected, we utilized the Entrez service to retrieve detailed information for each article. The Entrez Utilities Web service of the The National Center for Biotechnology Information allows access to these data via a programmatic interface based on the Simple Object Access Protocol (SOAP). The following services were used in this project:

- **ESummary** retrieves document summaries from a list of primary article IDs.
- **EFetch** retrieves records in the requested format from a list of one or more IDs or from the user's environment; it is used to obtain information such as the abstract, concepts or Journal Subsets.
- **Elink** checks for the existence of related articles from a list of one or more primary IDs.

3.3. Bibliography filtering

After articles were found, we required methods to delimit the space of results and measure the quality of the medical literature.

The journals indexed in MEDLINE are tagged using the Journal Subset codes, which enables restricting the results to specialized journals or articles on specialized topics. Having a literature collection, it is possible to define a Journal Subset space as the sum of the Journal Subsets of each article. This characteristic has been used

in this project as a heuristic to restrict the number of results, by filtering articles whose Journal Subsets were not included in the Journal Subset space. The following is an example of the filtering algorithm: Having an initial literature collection $D = \{d_1, d_2\}$ where d_1 Journal Subset is $\{s_1, s2\}$ and d_2 Journal Subset is $\{s_1\}$, the Journal Subset Space would be $\{s_1, s_2\}$ (the union of both Journal Subsets). When a new article d_3 with Journal Subset $\{s_1, s_3\}$ is evaluated, it is rejected or filtered because it has a Journal Subset s_3, which is not included in the original Journal Subset Space ($s_3 \notin \{s_1, s_2\}$).

However, further filtering was required in order to increase the quality of results. Randomised trials are most valuable means for evaluating therapeutic interventions (Rosenfeld & Shiffman, 2009) and they provide high quality evidence (Guyatt et al., 2008). We selected the set of criteria by Rosenfeld and Shiffman (2009) in order to identify quality literature in MEDLINE. The followings are the eligibility conditions that they propose:

1. "Randomized controlled trial" or "controlled clinical trial" as publication type, or
2. "Randomized", "placebo", or "randomly" in the abstract, or
3. "Clinical trials as topic" as a MeSH term or
4. "Trial" in the title, and
5. Restricts the final set (1 or 2 or 3 or 4) by excluding "animals" as MeSH term.

3.4. Article scoring

In addition to filtering, we required a method to order (or rank) articles depending on their relevance, this being measured by the score proposed in this section.

As a general weighting method, we looked at the vector space model (VSM), an algebraic model that represents natural language documents and queries in a high-dimensional space, where each dimension of the space corresponds to a word in the document collection (Salton & Buckley, 1988). In this context, the function tf/idf (term frequency *vs* inverse document frequency) is used to estimate a weight for each term in a document, where the term frequency in the given document offers a measure of the relevance of the term within a document, while the document frequency is a measure of the global relevance of the term within a collection of documents (Baeza-Yates & Ribeiro-Neto, 1999). In particular, we considered a collection of documents D containing $|D|$ documents, so that $D = \{d_1, \ldots, d_{|D|}\}$, where a single document is identified by d_j and contains a number of terms $|d_j|$ such as $d_j = \{t_1, \ldots, t_{|d_j|}\}$. Finally, the value of tf/idf for a term contained in a document within a collection of documents was given by

$$\begin{aligned}\mathrm{tf/idf}(t_i, d_j, D) &= \mathrm{tf}(t_i, d) \cdot \log_2 \mathrm{idf}(t_i, D) \\ &= \frac{|t_i|}{\max\{\mathrm{tf}(t_1, d_j), \ldots, \mathrm{tf}(t_{|d_j|}, d_j)\}} \cdot \log_2 \frac{|D|}{|D \supset t_i|}, \end{aligned} \quad (1)$$

where $|t_i|$ is the number of times that the term t_i occurs in the document d_j (which is normalised using the maximum term frequency found in d_j) and $|D \supset t_i|$ indicates in how many documents t_i appears.

There are other alternative term weighting approaches to tf/idf, such as (Robertson & Jones, 1976) or OKAPI (Robertson & Walker, 1999). Nonetheless, because these methods are usually variations on how the frequency functions tf and idf are combined, using other term relevance functions other than tf/idf should only offer minor differences on the final results.

We elaborated on Eq. (1) in order to adapt the traditional tf/idf weight to our scenario, with two collections of documents instead of just one: (i) the MEDLINE collection P that contains $|P|$ articles, so that $P = \{p_1, \ldots, p_{|P|}\}$; and (ii) the Bibliography collection B with size $|B|$, so that $B = \{b_1, \ldots, b_{|B|}\}$, thus representing all the bibliography of a specific clinical guideline that already exists. Both, documents in B such as b_j, $0 \leqslant j \leqslant |B|$ and documents in P such as p_k, $0 \leqslant k \leqslant |P|$, are expressed by a collection of terms with size $|b_j|$ and $|p_k|$ respectively, such as $b_j = \{t_1, \ldots, t_{|b_j|}\}$ or $p_j = \{t_1, \ldots, t_{|p_j|}\}$. Expressions b_{ji} and p_{ki} represent the number of appearances of term t_i in b_j or p_k respectively. Eq. (2) defines the function scoreDB that calculates the number of appearances of term t_i in the collection P. Therefore, Eq. (3) calculates the total number of concepts in P by adding the scoreDB value for all terms in P. This is used by Eq. (4) in order to count the number of appearances of a particular term t_i in the collection B. Both Eqs. (2) and (4) are identical but applied to different document collections; P and B respectively. Eq. (5) defines the function score that estimates the relevance of a term t_i in a collection B based on the frequency of t_i in the bibliography (function freqGuide as defined by Eq. (4)) and how unusual t_i is in P.

$$\mathrm{scoreDB}(t_a) = \sum_{i=1}^{|P|} p_{ia} \quad (2)$$

$$\mathrm{totalConcepts} = \sum_{i=1}^{n} \mathrm{scoreDB}(t_i) = \sum_{i=1}^{n} \sum_{j=1}^{|P|} p_{ji} \quad (3)$$

$$\mathrm{freqGuide}(t_a, B) = \sum_{i=1}^{|B|} b_{ia} \quad (4)$$

$$\mathrm{score}(t_i, B) = \mathrm{freqGuide}(t_i, B) \frac{\mathrm{totalConcepts}}{\mathrm{scoreDB}(t_i)} \quad (5)$$

Therefore, estimating the value of tf/idf for a particular document d is based on the sum of scores of its terms $t_i \in d$, as expressed by Eq. (6).

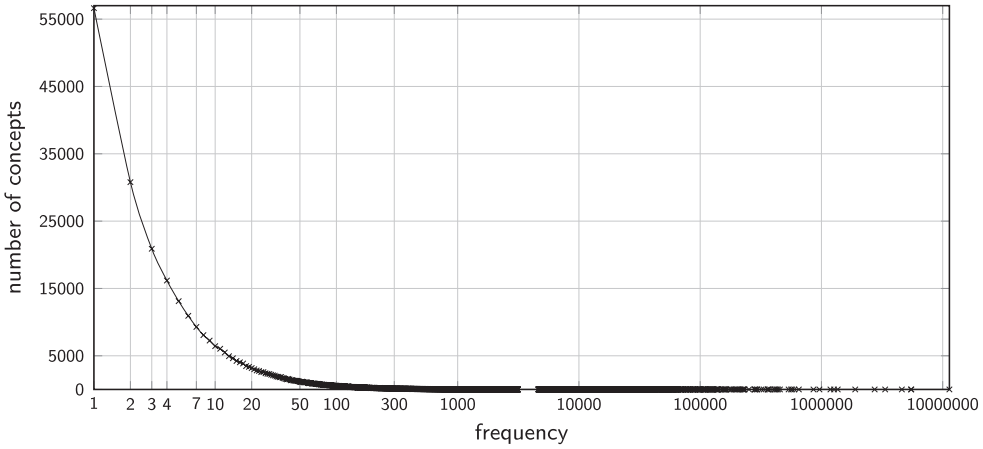

Fig. 4. Distribution of concepts by their frequency, with the horizontal axis in logarithmic scale.

Table 1
Experimental results after processing the 40 guidelines. Recall average after filtering was calculated using 38 elements. Guidelines 11 and 23 were excluded because N_{RS} is 0.

G	Period	N	N_{RS}	F	R_F	%	F_{RS}	$R_{F_{RS}}$	$\%_{F_{RS}}$
1	2002–2010	4	1	4	21,6423	1	1	39,113	1
2	2008–2010	5	3	5	17,4081	1	3	26,320	1
3	1996–2007	7	3	6	88454	0.857	3	15,144	1
4	2006–2011	7	2	7	102,973	1	2	14,327	1
5	2006–2008	9	2	6	130,625	0.667	2	19,649	1
6	2003–2008	10	6	10	68,383	1	6	10,030	1
7	2006–2010	11	1	11	211,279	1	1	33,417	1
8	2006–2011	13	2	13	124,770	1	2	14,089	1
9	2006–2011	14	4	14	71,497	1	4	6877	1
10	2000–2008	15	2	14	74,204	0.933	2	7918	1
11	1996–2010	16	0	16	402,213	1	0	36,343	n.a
12	2005–2009	17	2	17	63,910	1	2	4311	1
13	2002–2009	20	1	13	11,314	0.650	1	1436	1
14	2003–2008	20	10	20	44,756	1	10	7972	1
15	2003–2009	20	2	18	144,705	0.900	2	21,985	1
16	2004–2011	23	4	21	204,353	0.913	3	17,127	0.750
17	2006–2010	26	22	26	174,505	1	22	39,137	1
18	2004–2009	29	24	29	208,812	1	24	32,768	1
19	2002–2008	31	6	25	185,765	0.806	6	46,619	1
20	2000–2009	33	1	29	180,008	0.879	1	27,489	1
21	2003–2008	33	12	27	234,019	0.818	9	21,050	0.750
22	1994–2007	40	1	30	49,081	0.750	1	1805	1
23	2003–2008	40	0	40	238,683	1	0	30,632	n.a
24	2006–2011	44	17	40	221,840	0.909	15	45,562	0.882
25	2006–2009	45	3	41	311,787	0.911	3	28,648	1
26	2004–2010	61	6	59	128,801	0.967	6	23,468	1
27	2005–2008	70	19	64	114,030	0.914	17	23,784	0.895
28	2001–2008	79	48	74	160,378	0.937	47	35,847	0.979
29	2005–2011	86	3	85	42,397	0.988	3	2913	1
30	1999–2008	91	9	78	140,844	0.857	7	18,706	0.778
31	2004–2009	102	8	100	540,165	0.980	7	56,188	0.875
32	2006–2009	104	21	96	318,854	0.923	19	41,363	0.905
33	1995–2009	116	13	115	157,279	0.991	13	13,417	1
34	1999–2008	123	43	119	140,881	0.967	42	18,741	0.977
35	2001–2007	135	25	124	195,897	0.919	24	33,097	0.960
36	1999–2009	142	31	92	167,332	0.648	22	21,356	0.710
37	1998–2007	147	100	132	175,012	0.898	96	39,717	0.960
38	2010–2011	174	24	154	17,349	0.885	22	1293	0.917
39	2002–2009	177	53	173	35,391	0.977	53	4333	1
40	2002–2008	303	136	277	214,597	0.914	125	38,227	0.919
Average Rosenfeld & Shiffman				0.264					
Average				0.919			0.954		

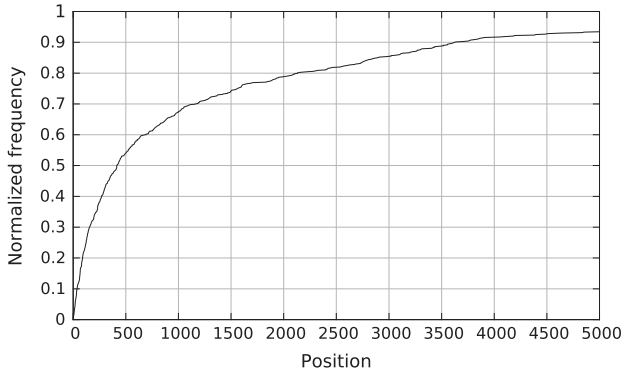

Fig. 5. Positions of results, using a cumulative distribution, after applying Rosenfeld & Shiffman filtering and tf/idf ordering.

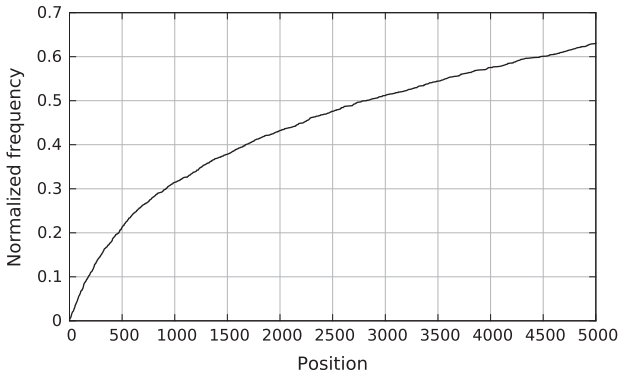

Fig. 6. Positions of result, using a cumulative distribution, without applying Rosenfeld & Shiffman filtering and tf/idf ordering.

$$\begin{aligned}\mathrm{tf/idf}(d,B) &= \sum_{i=1}^{n} \mathrm{score}(t_i,B) \\ &= \sum_{i=1}^{n} \left(\mathrm{freqGuide}(t_i,B)\, \frac{\mathrm{totalConcepts}}{\mathrm{scoreDB}(t_i)} \right) \\ &= \sum_{i=1}^{n} \left(\left(\sum_{z=1}^{|B|} b_{zi} \right) \frac{\sum_{j=1}^{n} \sum_{y=1}^{|P|} p_{yj}}{\sum_{x=1}^{|P|} p_{xi}} \right) \end{aligned} \quad (6)$$

In order to apply tf/idf, it was required to have a collection *P*. To build this collection, we chose the MEDLINE Baseline Distribution from the year 2011. This baseline includes 653 XML files containing 19,569,568 database records of bibliographical references published between the years 1865 and 2010. Within these articles, we identified a total of 212,588,035 MeSH concepts, 430,416 of them unique, and an average concept frequency of 493.91. The

Table A.1
Guidelines used in the experiment (Part 1).

N	Guideline	V_1	V_2	N_1	N_2
1	Screening and treatment for major depressive disorder in children and adolescents: U.S. Preventive Services Task Force recommendation statement	[a]	[b]	7	4
2	Assessment and management of chronic pain	[c]	[d]	146	5
3	Screening for Peripheral Arterial Disease: Recommendation Statement	[e]	[f]	14	7
4	Evaluation and management of patients with heart failure and preserved left ventricular ejection fraction: HFSA 2010 comprehensive heart failure practice guideline	[g]	[h]	53	7
5	2008 focused update incorporated into the ACC/AHA 2006 guidelines for the management of patients with valvular heart disease. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1998 Guidelines for the Management of Patients with Valvular Heart Disease)	[i]	[j]	904	9
6	Treatment of non-small cell lung cancer-stage IIIA: ACCP evidence-based clinical practice guidelines. (2nd edition)	[k]	[l]	89	10
7	Mood, memory, and cognition. In: Menopause and osteoporosis update 2009	[m]	[n]	93	11
8	EFNS guideline on the management of status epilepticus in adults	[o]	[p]	39	13
9	Surgical approaches to the treatment of heart failure: HFSA 2010 comprehensive heart failure practice guideline	[q]	[r]	27	14
10	Clinical practice guidelines for the management of sporotrichosis: 2007 update by the Infectious Diseases Society of America (IDSA)	[s]	[t]	27	15
11	Folic acid for the prevention of neural tube defects: U.S. Preventive Services Task Force recommendation statement	[u]	[v]	97	16
12	ACR Appropriateness Criteria® Crohn's disease	[w]	[x]	45	17
13	Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement	[y]	[z]	4	20
14	Treatment of non-small cell lung cancer, stage IIIB: ACCP evidence-based clinical practice guidelines. (2nd edition)	[aa]	[ab]	20	20
15	Antibiotic prophylaxis for GI endoscopy	[ac]	[ad]	113	20
16	Adapting your practice: general recommendations for the care of homeless patients	[ae]	[af]	19	23
17	EFNS guideline on the drug treatment of migraine revised report of an EFNS task force	[ag]	[ah]	177	26
18	Antithrombotic and thrombolytic therapy for ischemic stroke. American College of Chest Physicians evidence-based clinical practice guidelines (8th edition)	[ai]	[aj]	194	29
19	American Society of Clinical Oncology 2007 clinical practice guideline update on the role of bisphosphonates in multiple myeloma	[ak]	[al]	65	31
20	Clinical practice guidelines for the management of blastomycosis: 2008 update by the Infectious Diseases Society of America	[am]	[an]	21	33

Table A.2
Guidelines used in the experiment (Part 2).

N	Guideline	V_1	V_2	N_1	N_2
21	Lung cancer chemoprevention: ACCP evidence-based clinical practice guidelines. (2nd edition)	[ao]	[ap]	74	33
22	Reducing the risk of HIV infection associated with illicit drug use	[aq]	[ar]	13	40
23	(1) ACS guidelines for breast cancer screening: update 2003. (2) American Cancer Society Guideline for breast screening with MRI as an adjunct to mammography (2007)	[as]	[at]	170	40
24	Testosterone therapy in adult men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline	[au]	[av]	107	44
25	Long-term follow-up guidelines for survivors of childhood, adolescent, and young adult cancers. Sections 38–91: radiation	[aw]	[ax]	283	45
26	Reducing the risk of thrombosis and embolism during pregnancy and the puerperium	[ay]	[az]	45	61
27	Guidelines for the early management of adults with ischemic stroke. A guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups	[ba]	[bb]	68	70
28	Dementia	[bc]	[bd]	31	79
29	Updated guidelines for using interferon gamma release assays to detect Mycobacterium tuberculosis infection: United States 2010	[be]	[bf]	18	86
30	National Academy of Clinical Biochemistry laboratory medicine practice guidelines: Use of cardiac troponin and B-type natriuretic peptide or N-terminal proB-type Natriuretic Peptide for Etiologies other than acute coronary syndromes and heart failure	[bg]	[bh]	105	91
31	Antithrombotic therapy in neonates and children. American College of Chest Physicians evidence-based clinical practice guidelines (8th edition)	[bi]	[bj]	455	102
32	Opioids in the management of chronic non-cancer pain: an update of American Society of the Interventional Pain Physicians' (ASIPP) guidelines	[bk]	[bl]	279	104
33	Infant/toddler pulmonary function tests-2008 revision & update	[bm]	[bn]	30	116
34	National Academy of Clinical Biochemistry laboratory medicine practice guidelines: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes	[bg]	[bo]	105	123
35	NKF-K/DOQI clinical practice guidelines for hemodialysis adequacy: update 2006	[bp]	[bq]	218	135
36	National Academy of Clinical Biochemistry laboratory medicine practice guidelines: Clinical utilization of cardiac biomarker testing in heart failure	[bg]	[br]	105	142
37	Management of patients with dementia. A national clinical guideline	[bs]	[bt]	57	147
38	SAGES guidelines for the clinical application of laparoscopic biliary tract surgery	[bu]	[bv]	8	174
39	Guidelines for topical photodynamic therapy: update	[bw]	[bx]	103	177
40	Guidelines for the diagnosis and treatment of non-ST-segment elevation acute coronary syndromes	[by]	[bz]	211	303

most common concepts that this baseline contains are "humans", "male", and "female', whose occurrences are 11,359,319, 5,514,369 and 5,475,527 respectively. Fig. 4 depicts the distribution of these concepts by their frequency, where it can be appreciated how most concepts had a frequency smaller than approximately 100, while there were a few concepts with much higher frequencies, for example up to 10,000,000. These concepts were imported into a relational database to support the calculation of the tf/idf values by the ordering algorithm.

3.5. Algorithm

Using the data available and the proposed approach for scoring concepts, details on how the implementation was performed are presented in Appendix B. The main aspects of this implementation are detailed as pseudocode in Algorithm 1 and the auxiliary algorithms are detailed in Algorithms 2–5.

Firstly, we retrieve the related articles using ELink service. This process was performed twice to retrieve all related articles. Then,

Table A.3
Table with the references to the articles used as material in the experiment. Part 1.

Ref	Article	
a	U.S Preventive Services Task Force. Screening for Depression: Recommendations and Rationale. Ann Intern Med 2002; 136(10):760–4	
b	U.S Preventive Services Task Force. Screening and Treatment for Major Depressive Disorder in Children and Adolescents: US Preventive Services Task Force Recommendation Statement. Pediatrics 2009;123(4):1223–8	
c	Institute for clinical systems Improvement. Health Care Guideline: Assessment and Management of Chronic Pain. Assessment and Management of Chronic Pain; 3 ed.; 2008	
d	Institute for clinical systems Improvement. Health Care Guideline: Assessment and Management of Chronic Pain. Assessment and Management of Chronic Pain; 4 ed.; 2009	
e	U.S. Preventive Services Task Force. Guide to Clinical Preventive Services: Report of the U S Preventive Services Task Force; chap. 5. Lippincott Williams	Wilkin (1996)
f	U.S. Preventive Services Task Force. Screening for Peripheral Arterial Disease: Recommendation Statement. Am Fam Physician 2006;73(3):497–500	
g	Heart Failure Society of America. Section 11: Evaluation and Management of Patients With Heart Failure and Preserved Left Ventricular Ejection Fraction. J Card Fail 2006;12(1):e80–5	
h	Heart Failure Society of America. Section 11: Evaluation and Manage-ment of Patients with Heart Failure and Preserved Left Ventricular Ejection Fraction. J Card Fail 2010;16(6):e126–33	
i	Bonow RO, Carabello BA, Chatterjee K, de Leon Jr AC, Faxon DP, Freed MD, et al. ACC/AHA 2006 Guidelines for the Management of Patients With Valvular Heart Disease: A Report of the American College of Car-diology/American Heart Association Task Force on Practice Guideline (Writing Committee to Revise the 1998 Guidelines for the Management of Patients With Valvular Heart Disease) Developed in Collaboration With the Society of Cardiovascular Anesthesiologists Endorsed by the Society for Cardiovascular Angiography and Interventions and the Society of Thoracic Surgeons. J Am Coll Cardiol 2006;48(3):e1–148	
j	Bonow RO, Carabello BA, Chatterjee K, de Leon Jr AC, Faxon DP, Freed MD, et al. 2008 Focused Update Incorporated Into the ACC/AHA 2006 Guidelines for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1998 Guidelines for the Management of Patients With Valvular Heart Disease) Endorsed by the Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2008;52(13):e1–142	
k	Robinson LA, Wagner H, Ruckdeschel JC. Treatment of Stage IIIA Non-small Cell Lung Cancer. Chest 2003;123(1 suppl):202S-20	
l	Robinson LA, Ruckdeschel JC, Wagner H, Stevens CW. Treatment of Non-small Cell Lung Cancer-Stage IIIA. Chest 2007;132(3 suppl):243S-65	
m	The society of obstetricians and gynaecologist of Canada, editor. Canadian Consensus Conference on Menopause, 2006 Update; vol. 28; chap. 5. Journal of Obstetrics and Gynaecology Canada; 2006, p. S53–9	

Table A.4
Table with the references to the articles used as material in the experiment. Part 2.

Ref.	Article
n	The society of obstetricians and gynaecologist of Canada, editor. Menopause and Osteoporsis Update 2009; vol. 31; chap. 6. Journal of Obstetrics and Gynaecology Canada; 2009, p. S31–3
o	Meierkord H, Boon P, Engelsen B, Gacke K, Shorvon S, Tinuper P, et al. EFNS guideline on the management of status epilepticus. Eur J Neurol 2006;13(5):445–50
p	Meierkord H, Boon P, Engelsen B, Gacke K, Shorvon S, Tinuper P, et al. EFNS guideline on the management of status epilepticus in adults. Eur J Neurol 2010;17(3):348–55
q	Heart Failure Society of America. Section 10: Surgical Approaches to the Treatment of Heart Failure. J Card Fail 2006;12(1):e76–9
r	Heart Failure Society of America. HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16(6):e1–2
s	Kauffman CA, Hajjeh R, Chapman SW, Group MS. Practice Guidelines for the Management of Patients with Sporotrichosis. Clin Infect Dis 2000;30(4):684–7
t	Kauffman CA, Bustamante B, Chapman SW, Pappas PG. Clinical Practice Guidelines for the Management of Sporotrichosis: 2007 Update by the Infectious Diseases Society of America. Clin Infect Dis 2007;45(10):1255–65
u	U.S. Preventive Services Task Force. Guide to clinical preventive services; chap. 42. Williams & Wilkins; 2 ed.; 1996, p. 467–84
v	U.S. Preventive Services Task Force. Folic acid for the prevention of neural tube defects: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;150(9):626–31
w	Huprich J, Rosen M, Fidler J, Gay S, Grant T, Greene F, et al. ACR Appropriateness Criteria Crohn's disease. Reston (VA): American College of Radiology (ACR); 2005
x	Huprich J, Rosen M, Fidler J, Gay S, Grant T, Greene F, et al. ACR Appropriateness Criteria Crohn's disease. Reston (VA): American College of Radiology (ACR); 2008
y	U.S. Preventive Services Task Force. Screening for Colorectal Cancer: Recommendation and Rationale. Ann Intern Med 2002;137(2):129–31
z	U.S. Preventive Services Task Force. Screening for Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2008;149(9):627–37
aa	Jett JR, Scott WJ, Rivera MP, Sause WT. Guidelines on Treatment of Stage IIIB Non-small Cell Lung Cancer. Chest 2003;123(1 suppl):221S-5
ab	Jett JR, Schild SE, Keith RL, Kesler KA. Treatment of Non-small Cell Lung Cancer, Stage IIIB. Chest 2007;132(3 suppl):266S-76
ac	Hirota WK, Petersen K, Baron TH, Goldstein JL, Jacobson BC, Leighton JA, et al. Guidelines for Antibiotic Prophylaxis for GI Endoscopy. Gastrointest Endosc 2003;58(4):475–82
ad	Banerjee S, Shen B, Baron TH, Nelson DB, Anderson MA, Cash BD, et al. Antibiotic prophylaxis for GI endoscopy. Gastrointest Endosc 2008;67(6):791–8

we obtained the information corresponding to each specific article by means of the *Esummary* and *EFetch* Web Services. After that, we filtered the articles according to the method detailed in Section 3. Finally, the articles were ordered using the function tf/ idf specifically designed for this purpose and defined by Eq. (6).

4. Experiments and discussion

In order to validate our algorithm we proposed the following set of experiments: we randomly selected 40 guidelines from the National Guideline Clearinghouse before manually identifying the PubmedIds. Next, we applied the algorithm to each guideline, using the list of PUbMedIds that composes the first version of the guideline as input. As a result, the algorithm produced a list of PubmedIds from articles that had been published within a certain period of time. Finally, we compared these results with the PubmedIds from the second version, in order to measure the precision and recall of the algorithm.

Table 1 presents the experimental results, where column G identifies the guideline used to perform the experiment and col-

Table A.5
Table with the references to the articles used as material in the experiment. Part 3

Ref.	Article
ae	Bonin E, Brehove T, Kline S, Misgen M, Post P, Strehlow A, et al. Adapting your Practice: General Recommendations for the Care of Homeless Patients. Health Care for the Homeless Clinicians' Network, National Health Care for the Homeless Council; 2004
af	Bonin E, Brehove T, Carlson C, Downing M, Hoeft J, Kalinowski A, et al. Adapting your Practice: General Recommendations for the Care of Homeless Patients. Health Care for the Homeless Clinicians' Network, National Health Care for the Homeless Council; 2010
ag	Evers S, Afra J, Frese A, Goadsby PJ, Linde M, May A, et al. EFNS guideline on the drug treatment of migraine - report of an EFNS task force. Eur J Neurol 2006;13(6):560–72
ah	Evers S, Afra J, Frese A, Goadsby PJ, Linde M, May A, et al. EFNS guideline on the drug treatment of migraine - revised report of an EFNS task force. Eur J Neurol 2009;16(9):968–81
ai	Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest 2004;126(3 suppl):483S-512S
aj	Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest 2008;133(6 suppl):630S-69
ak	Berenson JR, Hillner BE, Kyle RA, Anderson K, Lipton A, Yee GC, et al. American Society of Clinical Oncology Clinical Practice Guidelines: The Role of Bisphosphonates in Multiple Myeloma. J Clin Oncol 2002;20(17):3719–36
al	Kyle RA, Yee GC, Somerffeld MR, Flynn PJ, Halabi S, Jagannath S, et al. American Society of Clinical Oncology 2007 Clinical Practice Guideline Update on the Role of Bisphosphonates in Multiple Myeloma. J Clin Oncol 2007;25(17):2464–72
am	Chapman SW, Bradsher RW, Campbell GD, Pappas PG, Kauffman CA. Practice Guidelines for the Management of Patients with Blastomycosis. Clin Infect Dis 2000;30(4):679–83
an	Chapman SW, Dismukes WE, Proia LA, Bradsher RW, Pappas PG, Threlkeld MG, et al. Clinical Practice Guidelines for the Management of Blastomycosis: 2008 Update by the Infectious Diseases Society of America. Clin Infect Dis 2008;46(12):1801–12
ao	Dragnev KH, Stover D, Dmitrovsky E. Lung Cancer Prevention. Chest 2003;123(1 suppl):60S-71S
ap	Gray J, Mao JT, Szabo E, Kelley M, Kurie J, Bepler G. Lung Cancer Chemoprevention. Chest 2007;132(3 suppl):56S-68S
aq	Provisional Committee on Pediatric AIDS. Reducing the Risk of Human Immunodeficiency Virus Infection Associated With Illicit Drug Use. Pediatrics 1994;94(6):945–7
ar	Committee on Pediatric AIDS. Reducing the Risk of HIV Infection Associated With Illicit Drug Use. Pediatrics 2006;117(2):566–71
as	Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WP, et al. American Cancer Society Guidelines for Breast Cancer Screening: Update 2003. CA Cancer J Clin 2003;53(3):141–69
at	Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. CA Cancer J Clin 2007;57(2):75–89

Table A.6
Table with the references to the articles used as material in the experiment. Part 4.

Ref.	Article
au	Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Testosterone Therapy in Adult Men with Androgen Deficiency Syndromes: An Endocrine Society Clinical Practice Guideline. J Clin Endocr MeTable 2006;91(6):1995–2010
av	Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Testosterone Therapy in Men with Androgen Deficiency Syndromes: An Endocrine Society Clinical Practice Guideline. J Clin Endocr MeTable 2010;95(6):2536–59
aw	Children's Oncology Group. Long-term follow-up guidelines for survivors of childhood, adolescent, and young adult cancers; chap. 38–91: radiation. Bethesda (MD): Children's Oncology Group; 2006, p. 45–101
ax	Children's Oncology Group. Long-term follow-up guidelines for survivors of childhood, adolescent, and young adult cancers; chap. 38–91: radiation. Bethesda (MD): Children's Oncology Group; 2008, p. 47–159
ay	Nelson-Piercy C. Guideline No. 37. Thromboprophylaxis during pregnancy, labour and after vaginal delivery. Royal College of Obstetricians and Gynaecologists; 2004
az	Nelson-Piercy C, MacCallum P, Mackillop L. Guideline No. 37a. Reducing the risk of thrombosis and embolism during pregnancy and the puerperium. Royal College of Obstetricians and Gynaecologists; 2009.
ba	Adams H, Adams R, Del Zoppo G, Goldstein LB. Guidelines for the Early Management of Patients With Ischemic Stroke. Stroke 2005;36(4):916–23
bb	Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the Early Management of Adults With Ischemic Stroke. Stroke 2007;38(5):1655–711
bc	Singapore. Ministry of Health and National Medical Research Council and National Committee on Neuroscience (Singapore). Dementia. MOH clini-cal practice guidelines; Ministry of Health; 2001
bd	Singapore. Ministry of Health. Clinical practice guidelines: dementia. MOH clinical practice guidelines; Ministry of Health; 2007
be	Mazurek GH, Jereb J, Lobue P, Iademarco MF, Metchock B, Vernon A. Guidelines for using the QuantiFERON-TB Gold test for detecting Mycobacterium tuberculosis infection, United States. MMWR Recomm Rep 2005;54(RR-15):49–55
bf	Mazurek GH, Jereb J, Vernon A, LoBue P, Goldberg S, Castro K. Updated guidelines for using Interferon Gamma Release Assays to detect Mycobacterium tuberculosis infection - United States, 2010. MMWR Recomm Rep 2010;59(RR-5):1–25
bg	Wu AH, Apple FS, Gibler WB, Jesse RL, Warshaw MM, Valdes R. National Academy of Clinical Biochemistry Standards of Laboratory Practice: Recommendations for the Use of Cardiac Markers in Coronary Artery Diseases. Clin Chem 1999;45(7):1104–21
bh	NACB Writing Group Members, Wu AH, Jaffe AS, Apple FS, Jesse RL, Francis GL, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Use of Cardiac Troponin and B-Type Natriuretic Peptide or N-Terminal proB-Type Natriuretic Peptide for Etiologies Other than Acute Coronary Syndromes and Heart Failure. Clin Chem 2007;53(12):2086–96

umn Period identifies the range of dates used to find articles. For example, if the period considered is 2002–2009, it implies that only articles published from 1 January 2002 to 1 January 2009 were retrieved. N indicates the number of new articles in V_2 that were published both during the period in question and contained by MEDLINE, while N_{RS} corresponds to the number of articles from N that fulfil the Rosenfeld & Shiffman criteria. F is the number of found articles (true positives) without using Rosenfeld & Shiffman criteria and R_F is the number of false positives after this first filtering. Column % is the percentage of elements that have been found (F/N). F_{RS} corresponds to the number of found articles (true positives) after Rosenfeld & Shiffman filtering. Finally, $R_{F_{RS}}$ is the number of false positives after Rosenfeld & Shiffman filtering and $\%_{F_{RS}}$ the percentage of elements that have been found (F_{RS}/N_{RS}).

After analysing the results, we conclude that the attempt to infer the bibliography for an updated clinical guideline from its previous version has been successful. As such, our method had a 91.9% finding average (or recall), and this percentage increased to 95.4% if

Table A.7
Table with the references to the articles used as material in the experiment. Part 5.

Ref.	Article
bi	Monagle P, Chan A, Massicotte P, Chalmers E, Michelson AD. Antithrom-botic Therapy in Children. Chest 2004;126(3 suppl):645S-87
bj	Monagle P, Chalmers E, Chan A, deVeber G, Kirkham F, Massicotte P, et al. Antithrombotic Therapy in Neonates and Children. Chest 2008;133(6 suppl):887S-968S
bk	Trescot AM, Boswell MV, Atluri SL, Hansen HC, Deer TR, Abdi S, et al. Opioid guidelines in the management of chronic non-cancer pain. Pain Physician 2006;9(1):1–39
bl	Trescot AM, Helm S, Hansen H, Glaser SE, Adlaka R, Patel Sa. Opioids in the Management of Chronic Non-Cancer Pain: An Update of American Society of the Interventional Pain Physicians' (ASIPP) Guidelines. Pain Physician 2008;11(2-SI):S5–62
bm	American Association for Respiratory Care (AARC). AARC (American Association for Respiratory Care) clinical practice guideline. Infant/toddler pulmonary function tests. Resp Care 1995;40(7):761–8
bn	American Association for Respiratory Care (AARC). Infant/toddler pulmonary function tests-2008 revision & update. Resp Care 2008;53(7):929–45
bo	NACB Writing Group Members, Morrow DA, Cannon CP, Jesse RL, Newby LK, Ravkilde J, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical Characteristics and Utilization of Biochemical Markers in Acute Coronary Syndromes. Circulation 2007;115(13):e356–75
bp	I. NKF-K/DOQI Clinical Practice Guidelines for Hemodialysis Adequacy: Update 2000. Am J Kidney Dis 2001;37(1, Supplement):S7–64
bq	Peritoneal Dialysis Adequacy 2006 Work Group. Clinical Practice Guidelines for Hemodialysis Adequacy, Update 2006. Am J Kidney Dis 2006;48, Supplement 1(0):S2–90
br	Tang WW, Francis GS, Morrow DA, Newby LK, Cannon CP, Jesse RL, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical Utilization of Cardiac Biomarker Testing in Heart Failure. Clin Biochem 2008;41(45):210–21
bs	Scottish Intercollegiate Guidelines Network. Interventions in the Management of Behavioural and Psychological Aspects of Dementia. No. 22 in SIGN publication; Edinburgh: SIGN; 1998
bt	Scottish Intercollegiate Guidelines Network. Management of patients with dementia. A national clinical guideline. No. 86 in SIGN publication; Edinburgh: SIGN; 2006
bu	Guidelines for the clinical application of laparoscopic biliary tract surgery. Society of American Gastrointestinal Endoscopic Surgeons. Surg Endosc 2000;14(8):771–2
bv	Overby D, Apelgren K, Richardson W, Fanelli R. SAGES guidelines for the clinical application of laparoscopic biliary tract surgery. Surg Endosc 2010;24:2368–86
bw	Morton C, Brown S, Collins S, Ibbotson S, Jenkinson H, Kurwa H, et al. Guidelines for topical photodynamic therapy: report of a workshop of the British Photodermatology Group. Brit J Dermatol 2002;146(4):552–67
bx	Morton C, McKenna K, Rhodes L, on behalf of the British Association of Dermatologists Therapy Guidelines, Subcommittee A, the British Photodermatology Group. Guidelines for topical photodynamic therapy: update. Brit J Dermatol 2008;159(6):1245–66
by	Bertrand ME, Simoons ML, Fox KA, Wallentin LC, Hamm CW, McFadden E, et al. Management of acute coronary syndromes in patients presenting withoutpersistent STsegment elevation. Eur Heart J 2002;23(23):1809–40
bz	Bassand JP, Hamm CW, Ardissino D, Boersma E, Budaj A, Fernndez-Avils F, et al. Guidelines for the diagnosis and treatment of non-ST-segment elevationacute coronary syndromes. Eur Heart J 2007;28(13):1598–660

we focused on randomized control trials, which are considered A-level evidence (Scottish Intercollegiate Guidelines Network, 2008).

The number of false positives is an aspect that needs further study. Our method returns a high percentage of them, thus producing a low precision that averaged 0.07% and 0.13% respectively. However, by applying the scoring function tf/idf (see Eq. (6)), we are able to return more than 65% of the target articles on the first 1000 positions (see Fig. 5). Figs. 5 and 6 use a cumulative distribution function to show how the found articles are positioned in relation to the total of articles retrieved, when Rosenfeld & Shiffman filtering is applied or not, respectively. For instance, Fig. 5 shows how almost 60% of the articles to be found (i.e. true positives) are positioned within the 500 first positions of retrieved articles. In the case of Fig. 6, only about 20% of found articles fell within the first 500 positions of the total. Comparing the results offered in Figs. 5 and 6 side by side, we conclude that the distribution of results was better when using Rosenfeld & Shiffman filtering.

Finally, it is necessary to remark that the evaluation criteria used in this study might have been too restrictive. We skimmed trough the results and realised that those articles classified as false positives were indeed about the disease being analysed in the corresponding guideline. We can only suppose that these articles might have been analysed during the actual guideline update process but eventually not included due to a number of different reasons. This could explain why they could be classified as interesting articles for guideline update purpose.

5. Conclusions

In this paper, we proposed a comprehensive approach to help medical committees in their decision process when updating clinical guidelines. We demonstrated how it was possible to infer a sensible bibliographic collection for a new version of an existing guideline, based on the existing bibliography of the latter. In that way, we were able to automatically suggest a set of articles to the medical committee as a starting point in their evaluation task. It is then the decision of those medical experts whether or not to accept the proposed articles as candidates for the new version of the clinical guideline being updated.

We consider that future work on filtering criteria could result in better general results. Specifically, incorporating to the current framework measures provided by the Journal Citation Reports (JCR) might contribute to better identify the most relevant authors in a particular field. JCR is the primary marker currently used when evaluating scientific publications by means of measuring journal relevance and publication quality. We believe that, by employing this ranking, it might be possible to enhance the scoring or filtering methods in order to improve the precision of the algorithm.

Acknowledgements

The authors thank J. López-Cuadrado for his comments on the manuscript and doctors J. J. Elorz Lambarri, E. González Perez-Yarza, and J. J. Korta Murua for the fruitful discussions. This work was supported by funding received from the Department of Education, Universities and Research of the Basque Government (Grant No. BFI-09-270), the UPV/EHU [GIU08/27, INF10/58, GIU11/28 and UFI11/19], Gipuzkoa Regional Council [OF53/2011], the Department of Industry, Commerce and Tourism — Basque Government [S-PE09UN60 and S-PE11UN115], and the Spanish Ministry of Science and Innovation [TIN2009-14 159-C05-03].

Appendix A. Bibliographic material

This appendix includes all the bibliographic materials that have been used to evaluate the algorithm. Tables A.1 and A.2 enumerate the 40 different guidelines that have been used to test the algorithm. Each guideline is composed by 2 bibliographic entries or

versions, that are listed in Tables A.3, A.4, A.5, A.6, A.7 in order to enable the reproducibility of the experiment.

Appendix B. Pseudocode

Algorithm 1. Guideline update algorithm

```
Input: oldPMIDs. List of PMIDs from first version articles.
Input: minDate
Input: maxDate
                                   ▷ Get related articles PMID
  newPMIDs ← relatedArticles(oldPMIDs,minDate,maxDate);
  resultPMIDs ← newPMIDs;
  resultPMIDs ← (resultPMIDs) ∪ (relatedArticles
  (newPMIDs,minDate,maxDate));
                                   ▷Get article information
  articles ← getInformation(resultPMIDs);
                                                ▷Filters
  articles ← filter(articles,oldPMIDs);
  articles ← filterRosenfeld&Shiffman(articles);
                                   ▷ Rank articles from max to min
  Scores
  articles ← rankArticlesMaxToMin(articles);
  printResults(articles);
```

Algorithm 2. Related Articles algorithm

```
Input: oldPMIDs. List of PMIDs from first version articles.
Input: minDate
Input: maxDate
Output: related. List of related PMIDs.
    for all pmid ∈ oldPMIDs do
                                   ▷ Get information from ELink
    Web Service
      related ← related ∪ ELinkWebService(pmid,
    minDate,maxDate)
    end for
  return related
```

Algorithm 3. Get Information algorithm

```
Input: PMIDs. List of PMIDs to collect information.
Output: articles. List of articles with complete information.
    for all pmid ∈ PMIDs
                                   ▷ Get information from
    ESummary and EFetch Web Services
      art ← EsummaryWebService(pmid) ∪
    EFetchWebService(pmid)
      articles ← articles ∪ art
    end for
    return articles
```

Algorithm 4. Filter Rosenfeld & Shiffman algorithm

```
Input: articles. List of articles to filter.
Output: articles. List of articles that fulfil Rosenfeld &
  Shiffman criteria.
    for all art ∈ articles do
    if "randomized controlled trial" ∈ art.pubtype then
        c1 ← True
    else if "controlled clinical trial" ∈ art.pubtype then
```

Algorithm 4 (*continued*)

Algorithm 4. Filter Rosenfeld & Shiffman algorithm

```
        c1 ← True
      end if
      if "randomized" or "randomly" or "placebo" ∈
    art.abstract then
        c2 ← True
      end if
      if "trial as topic" ∈ art.conceptList then
        c3 ← True
      end if
      if "trial" ∈ art.title then
        c4 ← True
      end if
      if "animals" ∈ art.conceptList then
        c5 ← True
      end if
      if (c5 or (not (c1 or c2 or c3 or c4))) then
        articles ← articles − art;
      end if
    end for
    return articles;
```

Algorithm 5. Filter algorithm

```
Input: articles. List of articles to filter.
Input: oldPMIDs. List of PMIDs from first version articles.
Output: articles. List of articles after filtering.
  oldArticles ← getInformation(oldPMIDs);
                                   ▷Get all journal subsets
  jSubsets ← oldArticles.JournalSubsets;
                                              ▷Filter Subset
    for all art ∈ articles do
      remove ← false;
        for p = 0 → art.JournalSubsetList.length do
          if art.JournalSubsetList[p] ∉ jSubsets then
            articles ← articles − art;
          end if
        end for
    end for
                                   ▷Calculate score for each article
  using Eq. (6)
                                                   ▷Filter
  Concepts
    maxConcepts ← articles.getMaxScores();
    threshold ← maxConcepts∗0.05;
    for all art ∈ articles do
      if art.score < threshold then
        articles ← articles − art
      end if
    end for
    return articles;
```

References

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Wokingham, UK: Addison-Wesley.

Field, M., & Lohr, K. (Eds.). (1990). *Clinical practice guidelines: Directions for a new program*. Washington, DC: National Academies Press.

Graham, R., Mancher, M., Miller, D., Greenfied, S., & Steinberg, E. (2011). *Clinical practice guidelines we can trust*. Washington, D.C: The National Academies Press. Chapter 6, (pp. 146).

Grol, R., Dalhuijsen, J., Thomas, S., Veld, C. I., Rutten, G., & Mokkink, H. (1998). Attributes of clinical guidelines that influence use of guidelines in general practice: Observational study. *BMJ, 317*, 858–861.

Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is ''quality of evidence'' and why is it important to clinicians? *BMJ, 336*, 995–998.

Ramampiaro, H. (2010). Biomedical information retrieval: the BioTracer approach. In *Proceedings of the first international conference on Information technology in bio- and medical informatics ITBAM'10* (pp. 143–157). Berlin, Heidelberg: Springer-Verlag.

Robertson, S. E., & Walker, S. (1999). Okapi/keenbow at TREC-8. In *Proceedings of the 8th text retrieval conference* (pp. 151–161).

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science, 27*, 129–146.

Rosenfeld, R. M., & Shiffman, R. N. (2009). Clinical practice guideline development manual. *Otolaryngol Head Neck Surg, 140*, S1–43.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Infomation Processing and Management, 24*, 513–523.

Scottish Intercollegiate Guidelines Network (2008). SIGN 50: A guideline developer's handbook. Edinburgh: SIGN.

Shapiro, D. W., Lasker, R. D., Bindman, A. B., & Lee, P. R. (1993). Containing costs while improving quality of care: The role of profiling and practice guidelines. *Annual Review of Public Health, 14*, 219–241.

Shekelle, P. G., Woolf, S. H., Eccles, M., & Grimshaw, J. (1999). Developing guidelines. *BMJ, 318*, 593–596.

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine, 30*, 7–18.

Swanson, D. R. (1988). Migraine and magnesium: Eleven neglected connections. *Perspectives in Biology and Medicine, 31*, 526–557.

Swanson, D. R. (1990). Somatomedin C and arginine: Implicit connections between mutually isolated literatures. *Perspectives in Biology and Medicine, 33*, 157–186.

Swanson, D. R. (1991). Complementary structures in disjoint science literatures. In *Proceedings of the 14 annual international acm sigir conference on research and development in information retrieval sigir '91* (pp. 280–289). NY, USA: ACM.

Swanson, D. R. (2008). Running, esophageal acid reflux, and atrial fibrillation: A chain of events linked by evidence from separate medical literatures. *Med Hypotheses, 71*, 178–185.

Timmermans, S., & Mauck, A. (2005). The promises and pitfalls of evidence-based medicine. *Health Affairs, 24*, 18–28.

van der Weijden, T., Boivin, A., Burgers, J., Schnnemann, H. J., & Elwyn, G. (2012). Clinical practice guidelines and patient decision aids an inevitable relationship. *Journal of Clinical Epidemiology, 65*, 584–589.

Weeber, M., Vos, R., Klein, H., De Jong-Van Den Berg, L., Aronson, A., & Molema, G. (2003). Generating hypotheses by discovering implicit associations in the literature: A case report of a search for new potential therapeutic uses for thalidomide. *Journal of the American Medical Informatics Association, 10*, 252–259.

Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Potential benefits, limitations, and harms of clinical guidelines. *BMJ, 318*, 527–530.

Wren, J. D., Bekeredjian, R., Stewart, J. A., Shohet, R. V., & Garner, H. R. (2004). Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics, 20*, 389–398.