O'Sullivan, D., Wilk, S., Michalowski, W. & Farion, K. (2010). Automatic indexing and retrieval of encounter-specific evidence for point-of-care support. Journal of Biomedical Informatics, 43(4), pp. 623-631. doi: 10.1016/j.jbi.2010.03.003

CITY UNIVERSITY LONDON

EST 1894

City Research Online

Original citation: O'Sullivan, D., Wilk, S., Michalowski, W. & Farion, K. (2010). Automatic indexing and retrieval of encounter-specific evidence for point-of-care support. Journal of Biomedical Informatics, 43(4), pp. 623-631. doi: 10.1016/j.jbi.2010.03.003

Permanent City Research Online URL: http://openaccess.city.ac.uk/13211/

Automatic Indexing and Retrieval of Encounter-specific Evidence for Point-of-Care Support

Dympna M. O'Sullivan, PhD[1†], Szymon A. Wilk, PhD[2],

Wojtek J. Michalowski, PhD[3] Ken J. Farion, MD[4]

[1]School of Engineering and Applied Science, Aston University, Birmingham, UK

[2]Institute of Computing Science, Poznan University of Technology, Poznan, Poland

[3] Telfer School of Management, University of Ottawa, Ottawa, Canada

[4] Departments of Pediatrics and Emergency Medicine, University of Ottawa, Ottawa, Canada

[†] Corresponding author:
Dympna O'Sullivan
School of Engineering and Applied Science,
Aston University,
Aston Triangle, Birmingham B4 7ET, U.K
Email: d.m.osullivan@aston.ac.uk
Tel: +44 121 204 3464
Fax: +44 121 204 3681
.

ABSTRACT

Evidence-based medicine relies on repositories of empirical research evidence that can be used to support clinical decision making for improved patient care. However, retrieving evidence from such repositories at local sites presents many challenges. This paper describes a methodological framework for automatically indexing and retrieving empirical research evidence in the form of the systematic reviews and associated studies from The Cochrane Library, where retrieved documents are specific to a patient-physician encounter and thus can be used to support evidence-based decision making at the point of care. Such an encounter is defined by three pertinent groups of concepts - diagnosis, treatment, and patient, and the framework relies on these three groups to steer indexing and retrieval of reviews and associated studies. An evaluation of the indexing and retrieval components of the proposed framework was performed using documents relevant for the pediatric asthma domain. Precision and recall values for automatic indexing of systematic reviews and associated studies were 0.93 and 0.87, and 0.81 and 0.56 respectively. Moreover, precision and recall for the retrieval of relevant systematic reviews and associated studies were 0.89 and 0.81, and 0.92 and 0.89 respectively. With minor modifications, the proposed methodological framework can be customized for other evidence repositories.

1. INTRODUCTION

Evidence-based medicine has been described as "the conscientious, explicit, and judicious use of current best evidence in making medical decisions" [1]. While there is wide acceptance that evidence-based medicine is a necessary component of health care delivery, and many online repositories of clinical evidence are being intensively developed and made available to health care providers, there are few solutions that offer direct support for incorporating it into everyday clinical practice. Numerous systems have been developed to leverage patient-specific data from clinical information systems such as Electronic Health Records (EHR); however, retrieving patient-specific information from diagnosis-driven online evidence repositories presents much greater challenges. In the first instance, the online medical literature is vast (approximately 30,000 scientific articles are published annually [2]), and retrieval of relevant evidence can require significant time and effort on behalf of the physician. Secondly, repositories are characterized by coarsely-grained indexing schemas which fail to adequately describe medical content; in other words they tend to emphasize only diagnostic aspects and often ignore important patient-oriented aspects which are more useful for point-of-care decision support. For example, certain patient characteristics are important for describing a given presentation; e.g. wheezing is an important patient characteristic in management of asthma and emphasizing these contextual characteristics in evidence-based documents should lead to more accurate retrieval. Thirdly, mechanisms for querying repositories are often simplistic and therefore ineffective given the volume of information available; and finally, there exists no standardized formatting among libraries making automatic retrieval of information across repositories difficult.

It has been demonstrated however, that even basic stand-alone facilities for querying medical literature can have a positive impact on physician decision making [3 and 4]. Therefore,

in this work we propose a novel methodology to improve the provision of clinical evidence by developing a framework for the automatic indexing, querying, and retrieving of evidence-based documents that are specific for a patient-physician encounter.

The proposed methodological framework achieves this aim in the following ways: Firstly, we focus on one library, The Cochrane Library [5], and one database from this library, the Cochrane Database of Systematic Reviews (for brevity in the text hereafter we will refer to this database as The Cochrane Library). Systematic reviews are clinically validated, concise summaries (summarized by clinicians) of the best available evidence in a particular clinical domain. They constitute comprehensive and high quality evidence by incorporating the results of many studies usually in the form of randomized controlled trials. In addition, systematic reviews from The Cochrane Library are created using a standard reporting template, and we exploit the structure of this template in our indexing and retrieval methodologies.

Secondly, we propose an enhanced indexing mechanism that better describes medical document content. Currently two main schemes prevail - bibliographic and full-text indexing. In bibliographic databases a human indexer assigns terms from a controlled vocabulary. The method is highly labor intensive as well as tightly coupled with the effectiveness of the controlled vocabulary to describe document concepts [6]. As a result, generated indices can be effective for describing higher-level diagnostic aspects of evidence-based documents but fall short of describing more fine-grained patient-related aspects. In full-text indexing documents are indexed automatically based on all terms which occur in the documents. This technique has often displayed shortcomings when applied to the highly specialized text which composes medical corpuses [7 and 8]. We address these shortcomings by finding UMLS [9] concepts to describe a patient-physician encounter, where an encounter is described in terms of diagnosis-, treatment-,

and patient-related concepts. These concepts are expanded using the ontology of clinical concepts from the UMLS Metathesaurus and are used for indexing by identifying expanded index concepts in evidence-based documents using the MetaMap Transfer (MMTx) system, and by applying well-established methodologies from information retrieval. The created indices are then used to retrieve evidence specific for the current patient-physician encounter.

Thirdly, we improve procedures for querying repositories of evidence-based documents. Currently most online medical literature retrieval systems offer quite simple search capabilities usually using Boolean operators (AND, OR) to combine search terms. Such searches are overly restrictive often resulting in either too few (AND) or too many (OR) retrieved documents. In addition retrieved documents cannot be effectively ranked with respect to a user's query. As a result, Boolean operator searches require users to have some knowledge of the underlying schema in order to carefully craft good queries [10]. We propose a concept-based search methodology that automatically formulates a query that is highly focused on the current patient-physician encounter and that allows for the effective ranking of retrieved documents with respect to a query.

2. BACKGROUND

We have investigated clinical information retrieval systems that leverage information from online medical repositories to aid with decision making. A taxonomy of such systems includes those that retrieve information directly from online repositories without additional processing, those that implement methodologies for enhancing queries sent to repositories through query reformulation, and those that rely on creating enhanced indexes for medical documents.

Retrieving information directly from clinical evidence repositories is the approach used by [11, 12 and 13]. In [12 and 13], the authors were interested in testing the relevance of evidence retrieved from MEDLINE using the PubMed engine. Retrieved evidence was termed "successful" if at least one returned citation was relevant for a clinical query where relevance was determined by an expert. Other research has focused on retrieving evidence that is contextual for specific patient characteristics. Relevant documents are pre-selected by clinical experts for predefined clinical topics and/or queries, and evidence (e.g. systematic reviews, peer-reviewed published research papers and clinical practice guidelines (CPG)), is explicitly associated with a particular topic or query in a clinical information system. Such approaches are discussed in [14 and 15] where the authors describe the use of so-called "infobuttons" - a web-based technology linking clinical data to information resources relevant for the current patient context. An "Infobutton Manager" is used to match contextual information against a knowledge base of information needs in order to propose a list of topics that may be of interest, and each topic has a customized link to a resource. The methodology relies on the indexing capabilities of the specific evidence repository (e.g. PubMed), and the information resources must be manually created and maintained. A similar evidence retrieval system is described in [11] where documents from the British Medical Journal's Clinical Evidence publication are preselected by experts and therefore statically associated with treatments that may be prescribed while using a computer physician order entry system.

Examples of systems that rely on enhanced queries sent to online medical repositories include [16, 17, 18 and 19], which act as alternative interfaces to a medical repository and [20], which is implemented as an intermediary layer between the retrieval system and a repository. For example in [16], when a user enters a valid MeSH [21] term as a query term, a list of associated

MeSH subheadings is presented along with a number of other options such as the clinical category (e.g. diagnosis, treatment), and the type of analysis desired. In [19], Zhenya and Wesley propose a knowledge-based query expansion method that exploits the UMLS to augment a query with additional terms that are relevant to a query scenario using a co-occurrence thesaurus-based method. The SAPHIRE system [17 and 18], uses concept extraction and synonym substitution to enhance querying by presenting the user with a list of related query terms from a semantic network. The user enters a free text query and the system automatically converts synonyms to their canonical form in the semantic network. The BiRD system [20], uses computerized CPG to generate evidence-based queries that are sent to PubMed for execution. Users identify a CPG for which literature is sought and the clinical category is determined using UMLS semantic types and a query is generated by converting identified terms to MeSH compliant terms. Contextual domain knowledge and term filtering are applied to refine the generated query and the CPG is annotated with medical literature. The ISAID system [22], uses a domain-dependent query, concept and document models where the query model comprises a set of questions, or generic queries, to be used as templates for document indexes. Candidate indices are proposed using a modified vector-space model (VSM).

From an indexing standpoint, most online medical repositories are indexed using terms from standard nomenclatures (e.g. MeSH terms), and applied by human indexers. The human effort involved in this process has caused researchers to investigate automatic methods for indexing medical content. Early research in this area is exemplified by the probabilistic indexing techniques used in the SAPHIRE system [17 and 18], that uses a variety of synonyms for each concept (term) in a MeSH semantic network and reduces them to the same canonical form, which is automatically applied as an index term. More recent work is exemplified by

IndexFinder [23] , which generates valid UMLS concepts by permuting the set of words in the input text and then filtering out irrelevant concepts via syntactic and semantic filtering. Because of the significant synonymy that exists in the medical vocabulary, a large body of research has investigated natural language processing (NLP) methods for indexing medical literature. NLP methods use part-of-speech tagging to extract noun phrases and lexical mapping to automatically map identified noun phrases to their canonical representation in standard biomedical vocabularies. For example, the Noun Phrase Identification (NPI) module reported by Huang et al in [24], is composed of a sentence boundary detector, a statistical natural language parser, and a noun phrase tagger and has been used for improved identification of noun phrases in clinical radiology reports. MedLEE [25], uses NLP to obtain structured data to automatically encode free-text data with UMLS concepts into noun phrases. Other systems that automatically map clinical concepts to standardized vocabularies have been reported; MetaMap [26], transforms the text in a document by syntactic analysis that recognizes simple noun phrases which are matched to UMLS concepts, Medical Text Indexer (MTI), which combines MetaMap with algorithms for finding related citations and methods for finding the closest MeSH terms to UMLS concepts to discover MeSH headings and more recently subheadings [27], which are used to produce ordered lists of recommended indexing terms, KnowledgeMap [28], is a scored-based algorithm for mapping free text to UMLS and the system developed by Nadkarni et al. [29], uses the UMLS for automated concept matching (and thereby indexing) of medical free text. The system was verified by applying it to analyze discharge summaries and surgical notes.

In many cases automatic indexing using NLP has been shown to be comparable to the accuracy of human indexers, however, generating grammars to achieve high levels of precision in extracting noun phrases is frequently a laborious task. Such approaches are tested by

comparing newly discovered noun phrases and concepts in free text to UMLS Metathesaurus entries, where high numbers of matched concepts indicate a highly performing system. While such methodologies are useful for discovering alternative or synonymous indexing terms, the problem of determining more fine-grained concepts to describe patient-specific aspects which are most useful for point of care support remains unresolved.

3. METHODS

3.1. Design of the Indexing and Retrieval Framework

The proposed methodological framework enables the retrieval of systematic reviews from The Cochrane Library that may be used to support evidence-based decision making at the point-of-care. At the point-of-care, a physician is likely to consult evidence after they have reviewed pertinent patient data, formed a diagnosis, and wish for confirmation of the best treatment to apply. Considering these three distinct activities, we assume that a patient-physician encounter is characterized by concepts from three categories - patient-related (signs and symptoms), diagnosis-related, and treatment-related. Following the idea of concept-based indexing [30], our framework uses concepts from these three categories for automatic indexing and retrieval of evidence.

Diagnosis- and treatment-related concepts refer to information on current diagnosis and proposed treatment(s), while patient-related concepts correspond to important demographic and disease-specific patient characteristics (e.g., a food allergy is an important concept when considering asthma). Diagnosis-, treatment-, and patient-related concepts are recorded in various, broadly understood patient documentation. In this research we identify diagnosis- and treatment-related concepts using CPG outlining patient management for the specific disease. Patient-

related concepts are identified from patient records (paper or electronic), used to store patient data.

Systematic reviews are population-based and summarize results of many studies, thus they underline common diagnosis- and treatment-related aspects (e.g., use of inhaled steroids in mild asthma), without describing in detail patients who participated in specific underlying studies. Detailed descriptions of participants in terms of patient-related information are part of individual study characteristics.

Our framework extensively exploits the standardized structure of systematic reviews from The Cochrane Library. Specifically we focus on three sections: Abstract, Plain Language Summary, and Characteristics of Studies. The first two sections present and summarize pertinent review information, while the last section characterizes individual studies included in a review.

In the proposed indexing and retrieval framework, for each relevant systematic review, the Abstract and Plain Language Summary sections are extracted as well as the Characteristics of Studies section, which is split into separate studies that compose the review. This results in a hierarchical structure, where each review is limited to the two extracted sections (Abstract and Plain Language Summary), and is associated with a set of underlying studies. For simplicity, in the text hereafter we will use the term "review" to refer to a document containing the Abstract and Plain Language Summary sections of a systematic review, and the term "study" to refer to a document from the Characteristics of Studies section which describes characteristics of a specific underlying study.

Considering the focus of the two types of documents, we use diagnosis- and treatment-related concepts for indexing and retrieving reviews, and patient-related concepts for indexing

and retrieving studies. In order to properly identify concepts in evidence-based documents, and to apply them to index and retrieve relevant documents we have to address the following issues:

1. A single concept may have multiple names; CPG and patient documentation introduce concepts under names that are often specific to a local setting (therefore we call them local names). The same concepts may be referenced by different names in evidence-based documents which render simple keyword-based indexing and retrieval useless, and calls for a method for mapping local names to concepts.

2. Concepts defining an encounter may not be exactly the same as concepts appearing in evidence-based documents, although they may be related. It is likely that evidence-based documents will include concepts that are more general (hypernymous), more specific (hyponymous), or similar (synonymous), to the concepts defining the encounter. This may pose a problem for basic concept-based indexing and calls for a means to properly identify such concepts when indexing evidence-based documents.

3. Evidence-based documents form a hierarchy with reviews at the top level and linked underlying studies at a lower level. Documents from both levels should be retrieved, combined, and ordered according to their relevance (similarity), to the patient-physician encounter. Typical approaches for document retrieval deal with only one level of documents and are based on some form of similarity score resulting from VSM [31]. However such an approach is not sufficient for our retrieval task which requires a customized methodology to retrieve documents from a hierarchical structure. Therefore we extend VSM so it can be used at multiple levels by introducing an aggregation layer that combines results obtained at different levels of the hierarchy. Finally, different categories of concepts (diagnosis,

treatment and patient), may have different importance for the physician and these differences should be accounted for in the retrieval process.

We address these issues by proposing a methodological framework that includes three execution phases:

1. In the first phase, with the help of a domain expert, we manually identify a set of concepts on the basis of their local names (there are several approaches for identifying local concept names ranging from directly asking a physician to identify concept names to automatically parsing patient records). The identified local names are then automatically mapped to standard concepts used by the UMLS Metathesaurus. Mapped standard concepts for identified concepts define the patient-physician encounter and are applied as indices to evidence-based documents; therefore we call them core index concepts.

2. In the second phase, the set of core index concepts is expanded to include related index concepts (similar, more general, and more specific), using the UMLS Metathesaurus ontology. The expanded index concepts are then identified in evidence-based documents using the MMTx system [32], and documents are indexed with core index concepts corresponding to located expanded index concepts, as well as with weights calculated using the term frequency – inverse document frequency (TF–IDF) schema [33].

3. Finally in the third phase, mappings between local names and core index concepts are applied to find concepts characterizing the current patient-physician encounter and to generate a concept-based query. The VSM is applied to retrieve relevant evidence-based documents by executing three queries corresponding to the three categories of core index concepts and combining their results.

Thus, the first two phases are concerned with preparing a repository of indexed documents, while the third phase involves querying and retrieving evidence at the point-of-care.

3.2. Identification of Core Index Concepts

For each identified local concept name (diagnosis-, treatment-, and patient-related), a core index concept is located in the UMLS Metathesaurus using normalized string matching. The mappings used to identify core index concepts are realized as translation lookup tables where keys correspond to the local names, and values correspond to subsets of core index concepts.

The first phase produces three sets of core index concepts from the UMLS Metathesaurus - diagnosis-, treatment-, and patient-related indices, and three mappings between local names and concepts for these categories. A detailed description, including pseudo-code, of the procedure for identifying core index concepts is outlined in Appendix A.

3.3. Indexing of Evidence-based Documents

In order to improve the coverage of core index concepts we expand them with related concepts. Related concepts are found by exploring ISA (hyponym), INVERSE ISA (hypernym), and SAME AS (synonym) relationships between concepts in the UMLS Metathesaurus. Expanded index concepts are used to construct index lookup tables by mapping expanded index concepts to core index concepts (keys correspond to expanded index concepts, and values to core index concepts). The index lookup tables are used when tagging evidence-based documents: Firstly, the MMTx system is used to find expanded index concepts in the documents by finding possible mappings between phrases and Metathesaurus concepts. Secondly, core index concepts that correspond to any expanded index concepts located in the text are found in the index lookup tables and used to tag documents. Finally, the tagged document content (i.e., lists of assigned

core index concepts), is used to compute TF–IDF weights in order to characterize the relevance of index concepts for individual reviews and studies from the corpus. Therefore, the second phase results in the generation of a set of indices for each document (review and linked studies), where indices are characterized by weighted core index concepts (thus only concepts that have been found in the document content have weights greater than 0). There are three types of weighted indices, – diagnosis and treatment-related indices for reviews, and patient-related indices for studies. A detailed description of the procedure for indexing evidence-based documents, including pseudo-code, is outlined in Appendix B.

3.4. Retrieving Evidence-based Documents

A specific patient-physician encounter is described by the subset of local names of core index concepts that characterize the encounter, and we consider three components of such a description – diagnosis-, treatment-, and patient-related. These components of the encounter description are used to construct queries for retrieving evidence-based documents relevant to the encounter. Since documents are indexed with core index concepts, the queries need to be expressed using the same concepts. This is achieved by automatic application of the translation lookup tables (created in the first phase) to the components of the encounter description, and results in three concept-based queries -- diagnosis-, treatment-, and patient-related.

These queries are executed by matching them against the appropriate document indices created in the second phase, and computing similarity scores according to the VSM. The first two queries are executed on the review indices, and the patient-related query is executed on the study index. This produces three lists of documents (documents are ordered according to their similarity scores), which are finally combined into two ordered list of reviews and studies that are ranked according to their relevance (overall similarity) to the patient-physician encounter. In

addition, the retrieval results can be customized to the preferences of a particular physician by specifying weights for each component of the patient-physician encounter (diagnosis, treatment, and patient-related). For example, one physician may be interested in documents that emphasize treatment and diagnosis aspects, while another may prefer documents with an emphasis on patient characteristics. Therefore the third phase generates two lists (reviews and studies) that are ordered according to their relevance (similarity) to the patient-physician encounter and any weights assigned by user. The retrieval procedure is presented in detail, including pseudo-code, in Appendix C.

4. RESULTS

4.1. Application of the Methodological Framework in a Sample Clinical Domain - Pediatric Asthma Exacerbations

In this section we discuss an implementation of the indexing and retrieval framework for providing evidence to support management of pediatric asthma patients in an emergency department. (Please note that throughout this section we refer to terminology and notation introduced in the pseudo-code from Appendices A, B and C.)

#### 4.1.1. Identification of Core Index Components

Diagnosis-, treatment-, and patient-related concepts for a patient-physician encounter are disease specific and we define them for pediatric asthma in the following way: The local names of diagnosis- and treatment-related concepts (LN_{Dx} and LN_{Tx}), were taken from the CAEP (Canadian Association of Emergency Physicians [34]) pediatric asthma CPG, which prescribes treatments for different levels of asthma exacerbation severity (mild, moderate and severe). An

emergency physician identified and supplied the local names of patient-related concepts (LN_{Pt})

from patient records (patient charts) at a collaborating teaching hospital.
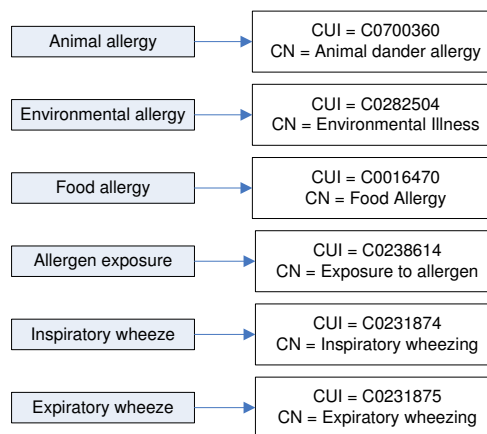
For each local name, a corresponding core concept defined in the UMLS Metathesaurus

was identified (search was limited to SNOMED CT, MeSH and MedDRA). The translation

lookup tables mapping local names to identified core index concepts (TLT_{Dx}, TLT_{Tx} and TLT_{Pt})

are shown in Fig. 1 (grey boxes correspond to local names and white boxes correspond to core

index concepts). We rely on concept unique identifiers (CUIs) for identification of core index

concepts, however for better readability we also present concept names (CN) as given by the

UMLS Metathesaurus.
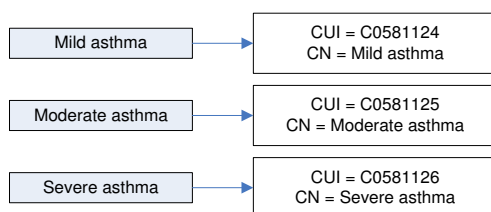
a) Diagnosis-related (TLT_{Dx}: $LN_{Dx} \rightarrow CIC_{Dx}$)
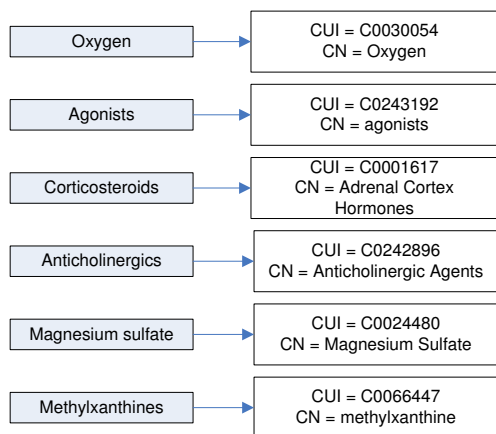
b) Treatment-related (TLT_{Tx}: $LN_{Tx} \rightarrow CIC_{Tx}$)　　c) Patient-related ($TLT_{Pt}$: $LN_{Pt} \rightarrow CIC_{Pt}$)

Fig.1. Translation lookup tables

4.1.2. Indexing of Evidence-based Documents

We began by retrieving systematic reviews for pediatric asthma from The Cochrane Library using a MeSH search. The query consisted of MeSH descriptor "asthma" and MeSH check word "child" and resulted in the retrieval of 56 systematic reviews. As described in "3.1. Design of the Indexing and Retrieval Framework", they were processed and separated into two types of evidence-based documents – 56 reviews with Abstract and Plain Language Summary sections and 423 studies with characteristics of individual studies underlying reviews. The documents were exported to a local database for easier indexing and retrieval. The set of core index concepts was then expanded with synonymous, hypernymous, and hyponymous concepts (see selected expanded concepts from Fig. 2 with white nodes corresponding to core concepts).
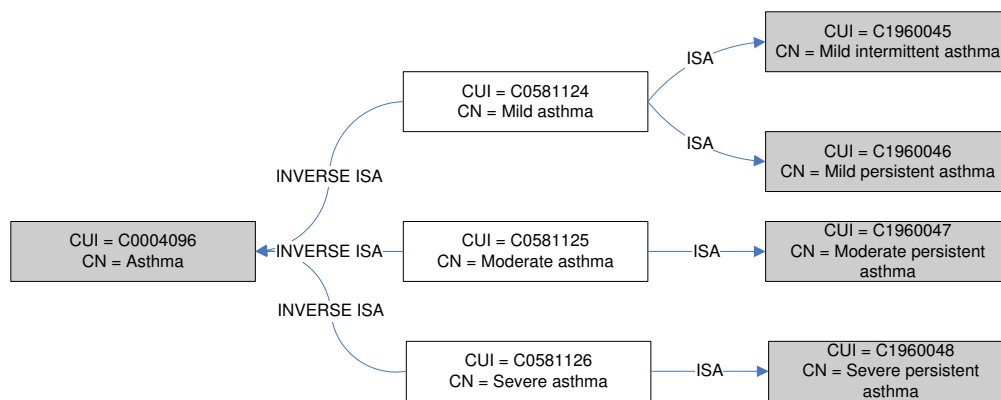
a) For "mild asthma", "moderate asthma" and "severe asthma"
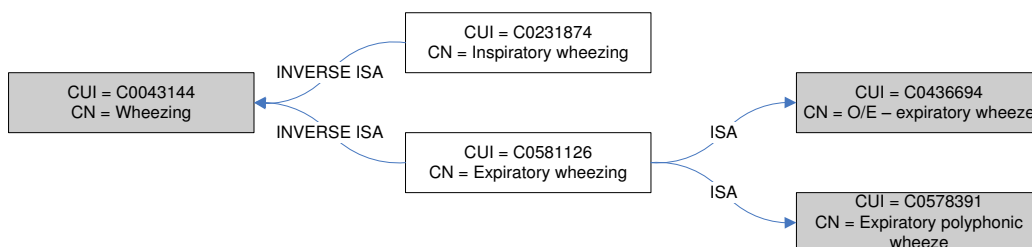
b) For "inspiratory wheezing" and "expiratory wheezing"

Fig.2. Expanded index concepts

The examples in Fig. 2 demonstrate some implications of the automated indexing approach. For example, consider the phrase "for persistent asthma", where MMTx locates the concept of "asthma" (CUI = C0004096). This concept is linked to three core index concepts, "mild asthma" (CUI = C0581124), "moderate asthma" (CUI = C0581125), and "severe asthma" (CUI = C0581126), as their enhanced parent (shown in Fig. 2a). Thus, whenever the enhanced concept (asthma) is located, the text will be tagged with all three core concepts (mild, moderate and severe asthma), resulting in an 'oversampling' or 'greedy indexing' of documents. However, in other situations the same issue can be beneficial. For example, consider the patient-related local concept 'wheezing'. As shown in Fig. 2b, core concepts "expiratory wheeze" and "inspiratory wheeze" have a common enhanced parent. Thus whenever the "wheeze" concept appears in a document, the text is tagged with "expiratory wheeze" and "inspiratory wheeze" concepts. Obviously, if either of the two specific concepts ("expiratory wheeze" or "inspiratory wheeze") is found, they will be identified by MMTx and the associated documents will be tagged more precisely.

Once the core index terms were applied to reviews and linked studies, they were weighted using the TF–IDF weighing schema. In calculating TF–IDF weights we used the functionality implemented by the Lucene Search engine [35]. The result of applying the indexing process was a hierarchy of weighted index vectors with top level index vectors corresponding to reviews (weighted diagnosis and treatment concepts), and lower level index vectors corresponding to underlying studies (weighted patient concepts). In Fig. 3 we have presented the index vectors computed for a sample review ('Intravenous aminophylline for acute severe asthma in children over two years receiving inhaled bronchodilators', by Mitra et al[‡]) and its

---

[‡] http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD001276/frame.html

included studies (patient-based concepts were located in 6 of 7 linked studies). For better

readability only the most pertinent index information (i.e., core index terms with TF–IDF

weights greater than 0) is shown. A detailed evaluation of the indexing phase is outlined in the
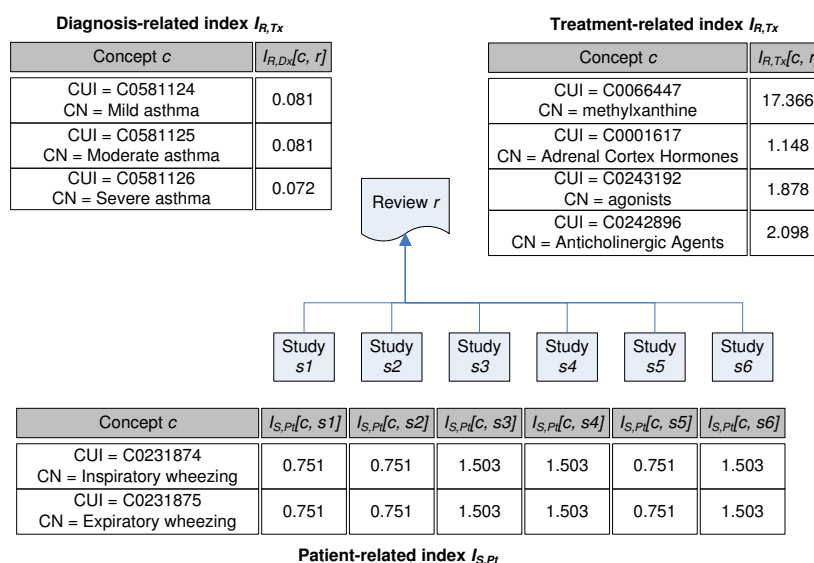
section 4.2.

Diagnosis-related index $I_{R,Tx}$

Concept c	$I_{R,Dx}[c, r]$
CUI = C0581124 CN = Mild asthma	0.081
CUI = C0581125 CN = Moderate asthma	0.081
CUI = C0581126 CN = Severe asthma	0.072

Review r

Treatment-related index $I_{R,Tx}$

Concept c	$I_{R,Tx}[c, r]$
CUI = C0066447 CN = methylxanthine	17.366
CUI = C0001617 CN = Adrenal Cortex Hormones	1.148
CUI = C0243192 CN = agonists	1.878
CUI = C0242896 CN = Anticholinergic Agents	2.098

Study s1 | Study s2 | Study s3 | Study s4 | Study s5 | Study s6

Concept c	$I_{S,Pt}[c, s1]$	$I_{S,Pt}[c, s2]$	$I_{S,Pt}[c, s3]$	$I_{S,Pt}[c, s4]$	$I_{S,Pt}[c, s5]$	$I_{S,Pt}[c, s6]$
CUI = C0231874 CN = Inspiratory wheezing	0.751	0.751	1.503	1.503	0.751	1.503
CUI = C0231875 CN = Expiratory wheezing	0.751	0.751	1.503	1.503	0.751	1.503

Patient-related index $I_{S,Pt}$

Fig.3.Weighed index vectors for the sample review and underlying studies

4.1.3. Retrieving Evidence-based Documents

A patient-physician encounter involving a pediatric asthma patient begins with a

physician collecting a set of signs and symptoms to obtain a complete clinical picture of the

patient. This may take several iterations and when completed a diagnostic decision is made

regarding the severity of the asthma exacerbation. Knowledge of a patient's status together with

a diagnosis allows a physician to choose the most appropriate treatment. The proposed evidence

retrieval framework can be used to support a physician in choosing a treatment by issuing a set

of three queries which work on the indices created in the second (indexing) phase. For example,

Fig. 4a illustrates a sample patient-physician encounter, its diagnosis, treatment, and patient-

related components, and the corresponding queries created by applying the translation lookup

tables from Fig. 1. Similarity scores calculated for these queries and documents from Fig. 3 (1 review and 6 linked studies), are given in Fig. 4b.

Patient-physician encounter		
Patient suffering from a moderate asthma exacerbation, experiencing both inspiratory and expiratory wheeze, and possibly treated with beta-agonists		

	Encounter component	Query
Diagnosis-related	CT_{Dx} = {"moderate asthma"}	Q_{Dx} = {CUI = C0581125}
Treatment-related	CT_{Tx} = {"agonists"}	Q_{Tx} = {CUI = C0234192}
Patient-related	CT_{Pt} = {"inspiratory wheeze", "expiratory wheeze"}	Q_{Pt} = {CUI = C0231874, CUI = C0231875}

a) Patient-physician encounter and derived queries
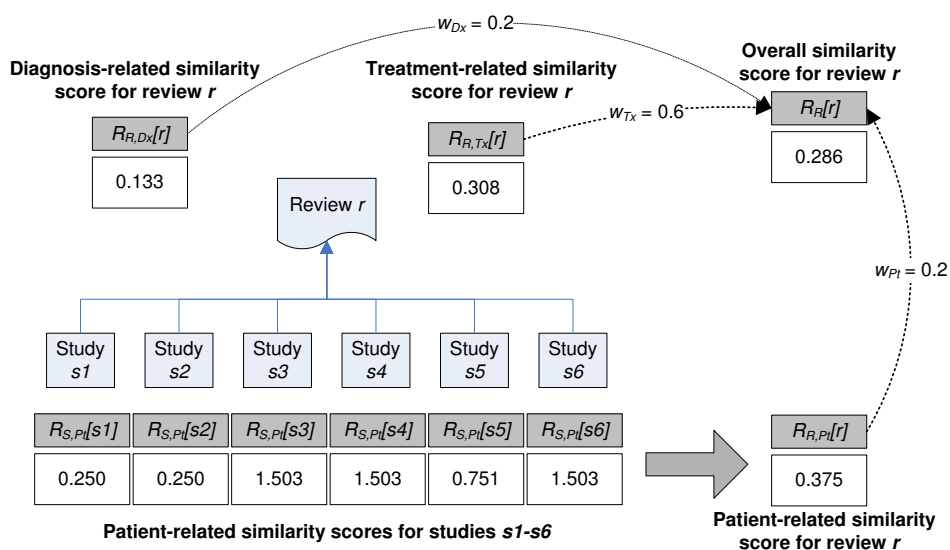
b) Similarity scores for a sample review and linked studies

Fig. 4. Sample patient-physician encounter and corresponding similarity scores

As shown in Fig. 4b, the overall similarity score for a review is obtained as the weighted sum of the three constituent similarity scores (for diagnosis-, treatment-, and patient-related queries), where each constituent score is scaled by a weighting factor supplied by a physician

who in this particular example is most interested in treatment-related aspects of clinical evidence (w_{Dx} = 0.2, w_{Tx} = 0.6, w_{Pt} = 0.2). Therefore the overall percentage similarity score between the characteristics of patient-physician encounter and the review is given by: (0.133*0.2) + (0.308*0.6) + (0.375*0.2) = 0.286*100 = 28.6%, indicating that in this instance the relevance of the review and underlying studies for the given patient-physician encounter is limited.

4.2. Testing of Automatic Indexing and Retrieval of Evidence-based Documents

The evaluation was performed on the repository of evidence-based documents for pediatric asthma already described in section 4.1 "Application of the Methodological Framework to a Sample Clinical Domain - Pediatric Asthma Exacerbations". The evaluation has two parts: firstly, we compared the proposed automated indexing approach with indices created by an expert, and secondly, we compared the automated retrieval approach with documents retrieved by an expert.

In evaluating the quality of automatic indexing, an expert physician analyzed the documents using treatment- and patient-related concepts from Fig. 1 (to prevent bias in the evaluation, diagnosis-related concepts for pediatric asthma were excluded due to the 'oversampling' issue discussed in the previous subsection), and assigned appropriate core index concepts for any identified treatment- and patient-related concepts. The index concepts applied by the expert were then used as a gold standard to which index concepts assigned by our automatic indexing methodology were compared.

In line with [36 and 37], we used precision and recall measures to evaluate the performance of our approach. The results were calculated using the following formulae:

$$\text{Precision} = \frac{|\text{CIC}(\text{E}) \cap \text{CIC}(\text{A})|}{|\text{CIC}(\text{A})|} \tag{1}$$

$$\text{Recall} = \frac{|CIC(E) \cap CIC(A)|}{|CIC(E)|} \qquad (2)$$

where $CIC(E)$ is a set of core index concepts assigned by the expert, and $CIC(A)$ is a set of core index concepts assigned by our automatic approach. Therefore, in this evaluation, precision corresponds to the number of correctly assigned core index over the total number of concepts assigned by the automatic approach. Recall corresponds to the number of correctly assigned core index concepts over the total number of concepts assigned by the expert.

We conducted these computations for two distinct groups of documents - reviews and studies, and the results are presented in Table 1

Table 1

Indexing evaluation results

Measure	Reviews	Studies
Precision	0.93	0.81
Recall	0.87	0.56

For both types of documents, the reported precision is high, demonstrating that the automatic approach does not assign many redundant (noisy or incorrect) index concepts. Recall is lower, especially for studies, indicating that the automatic approach is not able to capture all concepts that are assigned by the expert. The results for reviews are good and also compare favorably with other automatic indexing approaches described in [36, 38 and 39], which report precision and recall 0.78 and 0.68, 0.78 and 0.21, and 0.46 and 0.71 respectively. However, we note that in these studies texts were indexed using the entire MeSH vocabulary, hence a mismatch was more likely, while in our analysis the possible vocabulary (the set of core index concepts) was significantly smaller.

These less satisfactory results for studies can be explained by the following factors:

- Study descriptions are very concise and seldom formulated with full sentences, for example text describing participants are often of the form: "Participants: Male: 6, Female: 3", making it difficult for MMTx to identify core index concepts.

- Reporting modalities used in the studies are not restricted so the expert had to extensively use clinical knowledge to correctly index some studies. For example, studies reporting patients recruited in allergy clinic were indexed by an expert with the "exposure to allergen" core index concept while the allergy concept was missed by automatic approach.

In the second part of our evaluation, we analyzed the performance of the proposed framework in retrieving evidence-based documents. Queries corresponding to 15 encounters were outlined (the diagnosis-related component was again omitted), and relevant reviews and studies from the repository were identified by the expert physician. The documents identified by the expert were then used as a gold standard to which documents retrieved by our automatic retrieval methodology were compared. Precision and recall were calculated according to the following formulae:

$$\text{Precision} = \frac{|R(E) \cap R(A)|}{|R(A)|} \tag{3}$$

$$\text{Recall} = \frac{|R(E) \cap R(A)|}{|R(E)|} \tag{4}$$

where $R(E)$ is a set of documents identified as relevant by an expert, and $R(A)$ is the set of documents retrieved by our framework. Therefore, precision is the number of relevant documents retrieved over the total number of documents retrieved by our approach, and recall is the number of documents retrieved over the total number of expert-deemed relevant documents.

The average precision and recall values recorded for the 15 queries are presented in Table 2. The precision and recall is high for both types of documents, demonstrating that the automatic approach is capable of retrieving relevant evidence-based documents useful for point-of-care support. They also compare favorably with other automatic retrieval approaches described in [40 and 41], which report precision and recall of 0.85 and 0.79, and 0.86 and 0.78 respectively, for retrieving Medline abstracts. We intend to carry out a comprehensive evaluation of proposed retrieval metrics involving multiple physicians in the near future.

Table 2

Retrieval evaluation results

Measure	Reviews	Studies
Precision	0.89	0.92
Recall	0.81	0.89

5. DISCUSSION

The proposed methodological framework shares characteristics with other methods and approaches; however it expands them in a number of different ways. Similarly to "infobuttons" and the system described in [11], we provide evidence that is relevant for the current patient-physician encounter, however, the focus is on creating an enhanced, automatic, and scalable indexing framework so that relevant evidence-based documents need not be manually pre-selected. This in turn allows for easier system maintenance.

In a manner similar to the BiRD system [20], our framework uses concepts from a CPG to guide indexing and retrieval. However, instead of using only coarse-grained MeSH terms we use the ontology of concepts available in the UMLS Metathesaurus to map local names to standard concepts from different dictionaries as well as to explore relationships between

concepts. In this sense our methodology annotates medical literature with concepts from CPG and hospital documentation for enhanced indexing and retrieval whereas the BiRD system annotates CPG with medical literature for enhanced referencing. As in the SAPHIRE system [17], we recognize synonymous concepts and map them to core index concepts by utilizing the 'SAME AS' relationship defined in the Metathesaurus. However, we expand this idea by employing 'IS A' and 'INVERSE IS A' relationships to locate other hypernymous and homonymous concepts. Our methodology also relates to research into creating indices using methods from NLP, for example, from free text at the phrase and sentence level [42 and 43]. It uses NLP methods supplied by the MMTx system and combines these with pre-selected specialized sets of concepts from other information sources which are used to preprocess documents (these sources are CPG for diagnosis- and treatment-related concepts and critical pathways, assessment records and EHR for patient-related concepts). In that sense our work goes beyond lexical mapping to biomedical vocabularies described by [26, 28 and 29] to create more focused indices using specialized sets of concepts; in particular we can create indices that emphasize patient-oriented aspects of biomedical evidence.

The proposed framework has a number of shortcomings that we plan to address in future work. The first is the undesirable effects of 'oversampling', as discussed in "Application of the Methodological Framework to a Sample Clinical Domain - Pediatric Asthma Exacerbations". This could be addressed by applying expert knowledge to exclude certain extended indexing concepts after their identification. A related issue is the inability to distinguish between general and specific concepts. This issue could be also solved by manual processing where specific phrases are assigned greater importance by an expert.

Another shortcoming relates to the inability to use associated values for concepts. For example, using the UMLS Metathesaurus it is possible to identify a concept 'age' (understood as a property of organism), but it is not possible to capture exact age, as there is no concept of age expressed as a number or range (an additional issue here is that the Metathesaurus defines concepts with numbers only from 1 to 12). Capturing concepts and their values (especially numerical ones), requires more sophisticated analysis and we are developing a parser that uses NLP to post-process original documents after identifying index concepts with UMLS to identify potential associated values.

In related future work we are planning a comprehensive user evaluation to measure the document retrieval capabilities of the framework across multiple clinical presentations, multiple users, and multiple expert indexers. We are interested in incorporating a mechanism for relevance feedback to determine which returned evidence-based documents are most useful for point-of-care support. Such feedback could be obtained explicitly, by asking physicians to rate the relevance of retrieved documents, or implicitly, by monitoring, recording, and analyzing physician interactions with returned documents. While implicit relevance feedback may not be as accurate as explicit feedback, it may be more realistic at the point-of-care as it would not place a significant additional burden on physicians. From the perspective of maintainability, we intend to use MetaMap rather than MMTx, as MMTx is no longer supported by the National Library of Medicine.

6. CONCLUSIONS

In this paper we have proposed a methodological framework for the automatic indexing and querying of documents to retrieve medical evidence to support physician decision making at the point-of-care. The aim of the research is to enable indexing and retrieval of medical evidence

that is contextually relevant for a patient-physician encounter. This is achieved through a methodology that uses ontologies of diagnosis-, treatment-, and patient-related concepts which may be initially identified in a number of ways, for example, asking an expert, analyzing patient data or parsing patient records. These concepts are then automatically mapped to associated UMLS Metathesaurus concepts to construct extended indices so that documents are directly associated with local (and often limited) vocabularies used for describing a patient-physician encounter. As the framework is developed using functionality provided by the UMLS Metathesaurus, the MMTx system, and standard information retrieval methodologies (VSM and TF–IDF), it may be easily customized for any clinical problem. The proposed approach can be implemented as a component of a clinical decision support or computer physician order entry system and we have implemented it for the former [44]. In the paper we have illustrated our approach for evidence retrieval in the pediatric asthma domain and documents from The Cochrane Library. It is possible to adapt the proposed methodological framework for other repositories of medical evidence. However as the current approach uses the hierarchal structure of The Cochrane Library, it would be necessary to slightly modify the document retrieval procedure to accommodate libraries with different structures.

REFERENCES

1.   Sackett DL, Rosenberg W, Gray JM, Haynes R, Richardson W. Evidence based medicine: what it is and what it isn't. Br Med J 1996;312:71-72.

2.   Choi BCK. Understanding the basic principles of knowledge translation. J Epidemiol Community Health 2005;59(2):93.

3.   Chambliss ML, Conley J. Answering clinical questions. J Fam Pract 1996;43:140-144.

4.   Westbrook JI, Gosling AS, Coiera E. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321.

5.   The Cochrane Library [Internet]. Wiley InterScience; c1999-2009 [updated 2009 Apr 15; cited 2009 Apr 21]. Available from http://www.thecochranelibrary.org/.

6.   Stuckenschmidt H, van Harmelen F, Serafini L, Bouquet P, Giunchiglia F. Using C-OWL for the alignment and merging of medical ontologies. In: Hahn U, editor. Proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (KRMed'04); 2004. p. 88-101.

7.   Friedman C, Hripcsak G. Evaluating natural language processors in the clinical domain. Methods Inf Med 1997;37(4-5):334-344.

8.   Ciravegna F. Challenges in information extraction from text for knowledge management. IEEE Intell Syst App 2001;16(6):88-90.

9.   Unified Medical Language System [Internet]. National Library of Medicine; c1999-2009 [updated 2009 May 06; cited 2009 May 08]. Available from http://www.nlm.nih.gov/research/umls/.

10.   Frants V, Shapiro J, Taksa I, Voiskunskii V. Boolean search: Current state and perspectives. JASIST 1999;50(1):86-95.

11.   King JW, Le Saux N, Sampson M, Gaboury I, Norris M, Moher D. Effect of point of care information on inpatient management of bronchiolitis. BMC Pediatr 2007;7:4.

12.   Hauser SE, Demner-Fushman D, Jacobs JL, Humphrey SM, Ford G, Thoma GR. Using wireless handheld computers to seek information at the point of care: an evaluation by clinicians. J Am Med Inform Assoc 2007;14(6):807-815.

13.   Hauser SE, Demner-Fushman D, Ford GM, Jacobs JL, Thoma GR. Preliminary comparison of three search engines for point of care access to MEDLINE citations. AMIA Annu Symp Proc 2006:945.

14.   Demner-Fushman D, Seckman C, Fisher C, Hauser SE, Clayton J, Thoma GR. A prototype system to support evidence-based practice. AMIA Annu Symp Proc 2008:151-155.

15.   Bakkena S, Currie LM, Lee N-J, Roberts WD, Collins SA, Cimino JJ. Integrating evidence into clinical information systems for nursing decision support. Int J Med Inform 2008;77(6):413-420.

16.   Abasolo JM, Gomez M. MELISA: An ontology-based agent for information retrieval in medicine. ECDL Workshop on the Semantic Web 2000.

17.   Hersh W, Hickam D. Information retrieval in medicine: the SAPHIRE experience. Medinfo 1995;8(Pt 2):1433-1437.

18.   Hersh W, Mailhota M, Arnott-Smith C, Lowe HJ. Selective Automated Indexingnext term of Findings and Diagnoses in Radiology Reports. J Biomed Inform 2001;34(4):262-273.

19.   Zhenyu L, Wesley WC. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. Proceedings of the 2005 ACM symposium on Applied computing 2005:1076 - 1083.

20.   Abidi SSR, Kershaw M, Milios E. BiRD: a strategy to autonomously supplement clinical practice guidelines with related clinical studies. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05); 2005. p. 141b.

21.   Medical Subject Headings [Internet]. National Library of Medicine; c1999-2009 [updated 15 December 2008; cited 1 October 2009]. Available from http://www.nlm.nih.gov/mesh/.

22.   Berrios DC, Cucina RJ, Fagan LM. Methods for Semi-automated Indexing for High Precision Information Retrieval J Am Med Inform Assoc 2002;9:637-652.

23.   Zou Q CW, Morioka C, Leazer GH, Kangarloo H. Index-Finder: a method of extracting key concepts from clinical texts for indexing. AMIA Annu Symp Proc 2003:763–767.

24.   Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved Identification of Noun Phrases in Clinical Radiology Reports Using a High-Performance Statistical Natural Language Parser Augmented with the UMLS Specialist Lexicon. J Am Med Inform Assoc 2005;12(3):275-285.

25.   Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on Natural Language Processing. J Am Med Inform Assoc 2004.

26.   Aronson A. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. AMIA Annu Symp Proc 2001:17-21.

27.   Névéol A, Shooshan SE, Humphrey SM, Mork JG, Aronson AR. A recent advance in the automatic indexing of the biomedical literature. J Biomed Inform 2009;42(5):814-823.

28.   Denny JCS, Anderson, Miller, Randolph A, Schildcrout, Jonathan , Darbar, Dawood, Rosenbloom, S. Trent, Peterson, Josh F. Identifying QT prolongation from ECG impressions using a general-purpose Natural Language Processor. AMIA Annu Symp Proc 2005:196–200.

29.   Nadkarni P, Chen D, Brandt D. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc 2001;8(1):81-90.

30.   Elkin PL, Ruggieri A, Bergstrom L, Bauer BA, Lee M, Ogren PV, et al. A randomized controlled trial of concept based indexing of Web page content. AMIA Annu Symp Proc 2000:220-224.

31.   Salton G. Introduction to Modern Information Retrieval: McGraw-Hill; 1983.

32.   MetaMap. MetaMap Transfer (MMTx). c2009.

33.   Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inform Process Manag 1988;24(5):513-523.

34.   Welcome to CAEP [Internet]. Canadian Association of Emergency Physicians; c2009 [cited 2009 May 11]. Available from http://www.caep.ca/.

35.   The Apache Lucene [Internet]. The Apache Software Foundation; c2009 [cited 2009 April 22]. Available from http://lucene.apache.org/.

36.   Névéol A, Mork J, Aronson A, Shooshan S, Humphrey S. A recent advance in the automatic indexing of the biomedical literature. J Biomed Inform 2009;42:814-823.

37.   Rindflesch T, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform 2003;36:462–477.

38.   Névéol Aurélie, Mork James G., Aronson Alan R., J. D. Evaluation of French and English MeSH Indexing Systems with a Parallel Corpus. AMIA Annu Symp Proc 2005:565–569.

39.   Gay CW, Kayaalp M, Aronson AR. Semi-Automatic Indexing of Full Text Biomedical Articles. AMIA Annu Symp Proc 2005:271-275.

40.   Yu H, Lee M, Kaufman D, Ely J, Osheroff J, Hripcsak G, et al. Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. J Biomed Inform 2007;40(3):236-251.

41.   McDonald R, Pereira F. Identifying gene and protein mentions in text using conditional random fields. BMC Bioinformatics 2005;6(1).

42.   Mao W, Chu WW. Free-text medical document retrieval via phrase-based vector space model. AMIA Annu Symp Proc 2002:489-493.

43.   Berrios DC. Automated indexing for full text information retrieval. AMIA Annu Symp Proc 2000:71-75.

44.   Wilk S, Michalowski W, O'Sullivan D, Farion K, Matwin S. Engineering of a clinical decision support framework for the point of care use. AMIA Annu Symp Proc 2008:814-818.