

BiRD: A Strategy to Autonomously Supplement Clinical Practice Guidelines with Related Clinical Studies

Syed Sibte Raza Abidi, Micheal Kershaw & Evangelos Milios

*Health Informatics Lab, Faculty of Computer Science, Dalhousie University, Halifax, Canada
sraza@cs.dal.ca*

Abstract

In this paper we introduce a framework to supplement and tag computerized CPG with related best-evidence automatically sourced from on-line medical literature repositories. The idea is to provide CPG users with additional published evidence pertaining to the different sections of a CPG of their interest. We present a web-enabled Best-evidence Retrieval and Delivery (BiRD) system that autonomously retrieves pertinent medical literature with respect to user-specified content from a GEM-encoded CPG. This is achieved via a multi-level literature search strategy that uses the actual CPG content to autonomously generate a search query. The featured search strategy firstly categorizes the search query towards a priori defined clinical query subjects and secondly filters out insignificant medical terms from the search query. The technical architecture comprises existing medical language processing tools and vocabularies, together with newly developed tools to automatically (a) generate optimum search queries; (b) retrieve medical articles from MEDLINE; and (c) embed the retrieved medical articles within XML-based CPG encoded according to the GEM formalism.

1. Introduction

Clinical Practice Guidelines (CPG)—defined as "systematically developed statements to assist practitioners and patient decisions about appropriate health care for specific circumstances" [1]—provide means to improve the quality, effectiveness, and standardization of healthcare practices across different practitioners, departments and even institutions. In this regard, CPG describe the most appropriate diagnostic and treatment plan based on both the best scientific evidence and expert consensus available at the time of its compilation [2]. The origin of CPG is firmly grounded in clinical evidence—to a large extent evidence based medicine—which forms the basis for both formulating and subsequently validating (in practice) a CPG.

The relationship between CPG and best evidence is an on-going and pragmatic one. And, with the emergence of health informatics, in particular the omnipresence of the

Internet, we note with interest a triangular relationship between CPG, online best evidence and health informatics; the role of health informatics being the provider of methods and tools to (a) thematically organize best evidence at online medical literature archives, such as PubMed; (b) represent CPG in computational formalisms so that they can be incorporated into computer based medical systems; (c) to disseminate/retrieve 'relevant' best evidence from online medical literature archives; and (d) to support clinical decision support. It is this functional relationship between CPG, best evidence and health informatics methods for information retrieval that forms the focus of this paper.

Lately, there has been an increase in initiatives to computerize CPG for their usage at the point of care [3, 4]. It is noted that computerized CPG (C-CPG) are able to provide rapid evidence-based recommendations about what medical procedures to perform based on an individual patient's health profile [5,6,7]. To date, a number of C-CPG representation formalisms have been outlined by researchers—such as GLIF, GEM, Asbru, Proforma and so on—and deployed in medical systems.

Linking or tagging C-CPG with related and current best evidence—in terms of online medical literature—is a relatively unexplored idea. The rationale for doing so is based on the observation that medical practitioners when using C-CPG tend to reflect on the current best evidence primarily to validate or to supplement their understanding of the CPG [8]. Their intention usually is to refer to current medical literature to acquire insights into the rationale, past clinical trials, the diversity of opinions, related observations and current clinical evidence with respect to certain recommendations/procedures/treatments specified in the CPG [9]. To achieve the above, health practitioners retrieve medical literature from on-line medical repositories, such as MEDLINE, Cochrane, etc. However, it is noted that most practitioner's information needs are not fully met, for instance due to inexperience and inability to formulate the right search query, unawareness of online literature search facilities and subject specific filters, lack of time, expanding medical literature and differences in indexing schemes [10-13]. This brings to relief the question whether it is possible to automatically generate the most optimal search query for the specified CPG content.

Given the need for timely and consistent best evidence in concert with a computerized CPG, in this paper we investigate (a) the potential for augmenting a GEM-encoded C-CPG [14] with best evidence autonomously sourced from Medline; and (b) the possibility of circumventing the supposed problems experienced by practitioners in searching for the right evidence by automatically generating an ‘optimal’ medical literature search query based on C-CPG content of interest (this may be the title, statements, paragraph, sections) specified by the practitioner. We believe that since a CPG in its entirety deals with a variety of illness related issues (such as etiology, diagnosis, treatment, therapy, etc.) it is best to provide best evidence specific to the different sections of the C-CPG; this is in addition to all-encompassing evidence pertaining to the C-CPG in general. Our approach is to ask the practitioner to specify the C-CPG content for which evidence is sought. Then, in an autonomous and user-transparent manner we formulate an optimal search query based on the specified C-CPG content and finally tag the retrieved evidence (i.e. published articles) with the corresponding C-CPG content. The outcome of the process is a C-CPG supplemented with current and relevant best-evidence sourced from reliable medical literature repositories. We present a web-enabled *Best-evidence Retrieval and Delivery* (BiRD) system that provides the above mentioned functionality working with GEM-encoded C-CPG. BiRD incorporates a multi-level medical literature search strategy that (a) categorizes the search query towards a priori defined clinical query subjects and intentions, (b) filters insignificant (in terms of searching the medical literature in the given context) medical terms from the search query; and (c) improves the search query by removing least frequent terms to optimize the search outcome. The resultant is a highly focused medical literature search query that is classified to a particular query type and comprises contextually significant medical terms. The technical architecture of BiRD integrates existing medical language processing tools, medical vocabularies and PubMed’s E-utilities to query the MEDLINE database, with newly developed tools to generate search queries as per our strategy, to automatically retrieve medical articles from MEDLINE and to systematically embed the retrieved medical articles within GEM-encoded CPG [14].

2. CPG mediated evidence retrieval from online medical archives

To retrieve ‘right’ and ‘relevant’ medical literature vis-à-vis the CPG a practitioner needs to formulate a search query that manifests the CPG context, medical problem and the clinical intention using specialized problem-specific medical terms.

In this regard, the notion of medical query models have been researched leading to various solutions [15]. Cimino identified methods to generate generic queries from the admission information within patient records with the intention of connecting information in patient records with literature at MEDLINE [16]. Co-occurrence of medical concepts in a medical article has been used as the basis for determining a search query [17]. Query categorization has been successfully pursued by Haynes and the results have been incorporated at PubMed as built-in ‘clinical queries’ search filters [18]. The clinical query filters retrieve studies belonging to four study categories—i.e. therapy, diagnosis, etiology and prognosis—and with ‘sensitivity’ and ‘specificity’ options [19]. SUMSearch [20] provides additional filters corresponding to ‘Physical findings’, ‘Adverse treatment effects’ and ‘Screening/prevention’. Query classification for medical literature search has been further pursued by the QueryCat system [21] which automatically categorizes specific queries into ten (10) different query types—i.e. prevention, risk factors/etiology, diagnostic tests, diagnosis, symptoms, treatment, side effects, prognosis, overview & other [22].

In our work we refer to the query classification model proposed by Haynes [18], with the difference that we automatically categorize the query into the four query types based on the MeSH terms derived from the user-specified C-CPG content. This is different from the typical use of Haynes clinical query filters [19], whereby the practitioners determines the intended query type. Once identifying the query type, we select the relevant query type filter available at PubMed to leverage the functionality of PubMed’s E-utilities.

We argue that since a C-CPG is highly focused in their treatment of a clinical problem/procedure, it is important that any best evidence reviewed by practitioners to supplement their understanding of the C-CPG should be directly related to the C-CPG. We contend that the quality of best evidence provided to practitioners can be improved if the origin of the medical literature search query is ‘objectively’ derived from the original content of the C-CPG, given the limitations noted earlier in query formulation [10-12]. Here we are suggesting an alternative, yet interesting and practical, literature search alternative that uses both the original C-CPG content and syntactic structure (i.e. section tags of the C-CPG) to formulate a more focused literature search query. In this case, the structured content contained in a C-CPG becomes the basis for retrieving related literature articles—i.e. the key medical terms present in the C-CPG content together with the context of the C-CPG content realizes an optimal literature search query. This ensures a consistent search query, and in turn consistent published evidence delivery, across different users of a C-CPG.

3. Functional solution design

Designing a functional solution to supplement C-CPGs with relevant medical literature involved the following tasks: (1) Allowing users to identify the C-CPG content for which corresponding medical literature is sought; (2) Determining the medical terms, from the selected C-CPG content, to be used to generate a search query for literature retrieval; (3) Converting the identified medical terms to MeSH compliant terms because medical literature at MEDLINE is indexed based on MeSH terms; (4) Generating a context-sensitive medical literature search query (comprising a set of MeSH terms) to retrieve the most relevant medical articles; (5) Presenting the search query to MEDLINE in compliance with the operational limits enforced by PubMed, to retrieve the most relevant medical articles; (6) Analyzing the search query results to ensure optimum coverage and relevance of the query results; and finally (7) Returning the retrieved medical articles to the corresponding segments within the C-CPG.

For maximum impact it is important that the evidence provided to practitioners is relevant—i.e. the medical articles retrieved from MEDLINE are in line with both the intention and context of the C-CPG segment in question.

The premise of our search strategy is that: (a) The search query should represent the most likely underlying intention/theme of the selected medical content. One way is to categorize the query to a query type based on a set of a priori defined query types [10]. The query type should be objectively determined, as opposed to be subjectively assigned by a user, based on the nature and meaning of candidate MeSH terms derived from the medical content; (b) The search query should be a combination of the query type and a list of candidate MeSH terms. For instance, if the query type is determined to be *etiology* and the candidate terms are {*Kidney, Urinary Tract Infections, Radiation*} then the medical articles retrieved should have a focus on the etiology of as opposed to the diagnosis or treatment of the same; (c) The search query should only include a small number of highly significant MeSH terms. A large number of terms tend to make the query too specific which usually results in no medical articles being retrieved. This requires to filter out less significant MeSH terms from the set of candidate MeSH terms.

Our medical literature search strategy attempts to generate an optimum search query by (a) using MeSH terms only [11]; (b) categorizing the search query to a query type for a more focused search [10]; and (c) reducing the length of the search query by removing insignificant medical terms. Categorization of the search query significantly reduces the search space, thus allowing the search engine to focus on a smaller set of medical articles that are indexed under the search category in

question. Reducing the length of the search query has two implications; (i) the search query becomes more generic, yet more focused, (ii) the search query is less constrained as it does not need to satisfy a larger list of search terms, where some of the search terms were not significant yet impacting the granularity of the search query.

In our work, we have devised a multi-level query generation model that comprises two distinct stages: (1) *Query categorization stage* and (2) *Term filtering stage*. The flowchart of the search strategy is given in figure 1.

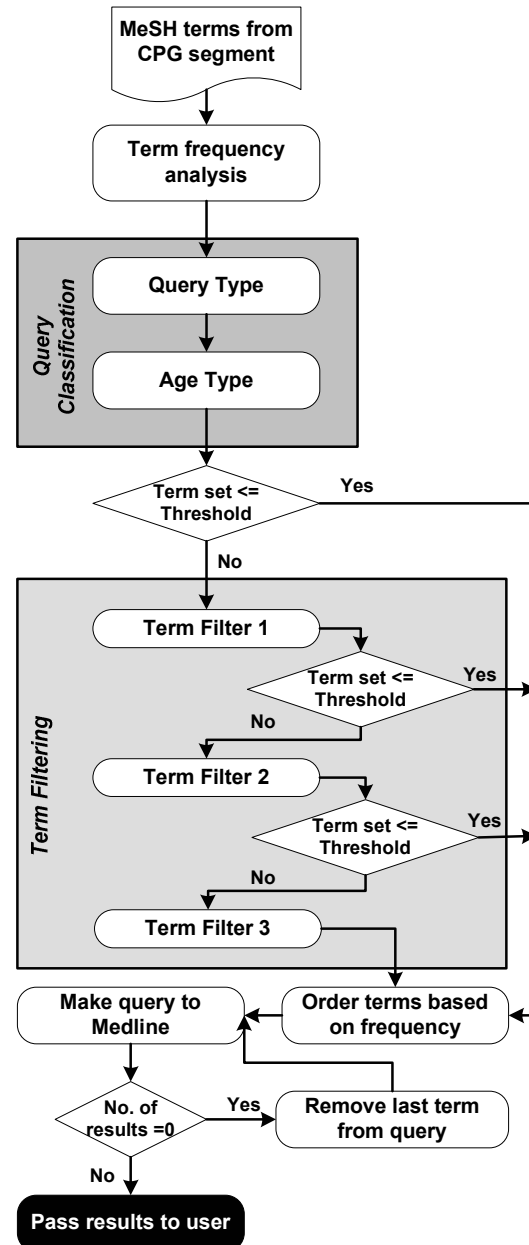


Figure 1. Flow chart of our strategy to generate an optimal literature search query using significant MeSH terms

3.1. Stage 1—Query categorization

We begin by categorizing the query into the four query types supported by PubMed clinical query filters:

1. *Therapy*: Content to be retrieved relates to therapy, prevention or rehabilitation of a disease or condition.
2. *Diagnosis*: Content to be retrieved relates to evaluation of a disease or condition.
3. *Etiology*: Content to be retrieved relates to causation of a disease or condition.
4. *Prognosis*: Content to be retrieved predicts or forecasts the course of a disease or condition.

The above query classification scheme has been suggested as PubMed's Clinical Query filters [19]. Each query type is associated with a set of UMLS semantic types; medical terms belonging to these UMLS semantic types are regarded as *trigger words*—i.e. medical terms that intrinsically reflect the intention of the query type. For instance, the medical terms 'catheterization' belongs to the UMLS semantic type 'therapeutic or preventive procedure' which is related to the query type 'therapy'. Hence, the presence of the medical term 'catheterization' in the selected C-CPG content strengthens the belief that the likely focus of the C-CPG content is 'therapy', and the query type should be set as *therapy*. The set of semantic types representative of each query type have been determined by medical researchers via surveying clinicians, medical librarians and researchers in Canada and the United States [19].

Given a set of candidate medical terms, we determine the most appropriate query type by comparing the candidate medical terms against the set of trigger words for each query type. The query type associated with the most matched trigger words is deemed as the target query type for the candidate medical terms.

3.2. Stage 2—Query Categorization

The second level of categorization is based on the age dimension. Medical studies are specific to age and the results of such studies are applicable to a definitive age-group, hence it is important to retrieve and provide evidence that is pertinent to the age-group in question.

For our purposes, we categorize the query into four age groups: *infant* (0 - 23 Months), *child* (0 - 18 years), *adult* (19+ years), *elderly* (65+ years). The age-groups were derived by subsuming 13 different age-groups defined by PubMed to classify queries. The trigger words for each age group were derived using the UMLS semantic types related to the age dimension—each MeSH term belonging to the Age Group UMLS semantic type was associated with an age-group and it served as the trigger word that age-group. For instance, the MeSH terms (in the Age Group UMLS semantic type) {Infant; Infant, low birth

weight; Infant, newborn; Infant (postmature); Infant (premature); Infant (small for gestational age)} would serve as the trigger word for the age-group *infant*. Whereas, the MeSH terms {Elderly; Centenarians; Nonagenarians; Octogenarians; Frail elderly} would trigger the age-group *elderly*.

3.3. Stage 3—Term Filtering

Categorization of the search query leads to the next stage in which we need to identify and then subsequently filter-out those medical terms that are potentially insignificant and thus unnecessarily constrain the query. We have designed three (3) term filters which are used in sequence; each filter is used only if the number of candidate query terms is greater than a user-defined *query length threshold*.

3.3.1. Term Filter 1. The first term-filter is designed to remove all candidate query terms that belong to a list of filterable UMLS semantic types. These UMLS semantic types are deemed to be of less medical significance [17], for instance the terms listed under the semantic type Reptile are seemingly of nominal significance.

The default list of filterable semantic types is {Physical Object, Amphibian, Plant, Alga, Animal, Vertebrate, Invertebrate, Fish, Bird, Reptile, Research Device, Idea or Concept, Temporal Concept, Qualitative Concept, Quantitative Concept, Functional Concept, Spatial Concept, Geographic Area, Intellectual Product, Classification, Regulation or Law, Language, Occupation or Discipline, Organization, Professional Society, Self Help or Relief Organization, Group, Population Group, Family Group, Professional or Occupational Group}.

In practice, users are allowed to modify the list to retain the semantic types that they deem as being significant for the problem at hand. Furthermore, two considerations are taken into account before filtering the candidate query term: (1) If the frequency of the term belonging to one of the filterable semantic type is greater than a pre-defined threshold we regard the term to have some significance in the context of the medical content in question. In this case the term is not filtered; (2) If a candidate query term belongs to multiple semantic types, and only one of the semantic types is a filterable semantic type then the term is not filtered.

3.3.2. Term Filter 2. The second term-filter is an extreme case of the first term-filter. It removes a candidate query term if any one of its semantic types belongs to the filterable semantic type group.

3.3.3. Term Filter 3. We know that the candidate query terms are derived from the CPG content in question. In

practice, the text parsing process identifies the noun phrases in the CPG content, and subsequently translates the words within the noun phrase to corresponding MeSH terms. In this process, each noun phrase may map to multiple MeSH terms; where each MeSH term is assigned a *mapping score* (out of 1000) that indicates the strength/confidence of the mapping of the original phrase to a corresponding MeSH term. At times, a single noun phrase may map to multiple MeSH terms, such that each MeSH term having a different mapping score. For instance the noun phrase “kidney cortex” is mapped to the following three MeSH terms, each with a different mapping score (shown in the square brackets): [1000] Kidney Cortex; [755] Cascara; and [694] Kidney.

If a noun phrase produces multiple MeSH terms, in our scheme we are interested in retaining only the highest mapping score MeSH term. The third term-filter is designed to filter-out all terms with a low mapping score that originate from the same noun-phrase. So, for the above example, the third term-filter will only retain the MeSH term “kidney cortex” and filter-out the terms “cascara” and “kidney”.

4. Functionality of BiRD

Our BiRD system features the application of existing medical language processing tools, medical vocabularies and utilities to access PubMed. In addition, we have developed modules to (i) input and visualize a GEM-encoded C-CPG (or text documents), (ii) specify C-CPG content for which evidence is sought, (iii) autonomously generate an optimal search query as per the above-mentioned search methodology, (iv) present the query to PubMed in accordance with PubMed’s querying regulations, (v) retrieve the medical articles, and (vi) incorporate the medical articles within the C-CPG (next to the selected C-CPG content). Figure 2 shows the functional diagram and the constituent modules of BiRD.

In the forthcoming discussion we will discuss the functionality of the various components of the BiRD system using a working example of a GEM-encoded CPG for managing urinary tract infections.

4.1. C-CPG visualizer

The functionality of the C-CPG visualizer is to present a GEM-encoded C-CPG to the user to specify the C-CPG content for which evidence is sought. A GEM document is in XML format, comprising a number of semantically-significant tags that help organize the entire CPG content into meaningful blocks. The C-CPG visualizer leverages the GEM tags to parse the C-CPG into blocks of C-CPG content, each demarcated by a set of GEM tags.

Users can not only view the C-CPG but also simultaneously select those C-CPG content blocks for which supplementary best evidence is required. Selection of a C-CPG content is simply done by clicking on the check-box next to it. Note that each individual C-CPG content within a GEM tag yields a separate search query and is associated with retrieved literature articles.

4.2. MeSH term generator

This module is responsible for generating MeSH terms from the original text of the selected C-CPG content. We use the MetaMap Transfer (MMTx) tool, developed by the National Library of Medicine, that offers the functionality to identify medical terms in a text and translate them to meaningful MeSH terms. We will not be able to provide details of the working of the MMTx tool, however in figure 3 we show a C-CPG content block taken from a UTI CPG available at <http://ycmi.med.yale.edu/GEM>.

We calculate the frequency of each MeSH term identified by MMTx in order to establish its importance within the given C-CPG content and also highlight its UMLS semantic type (as shown in Table 1).

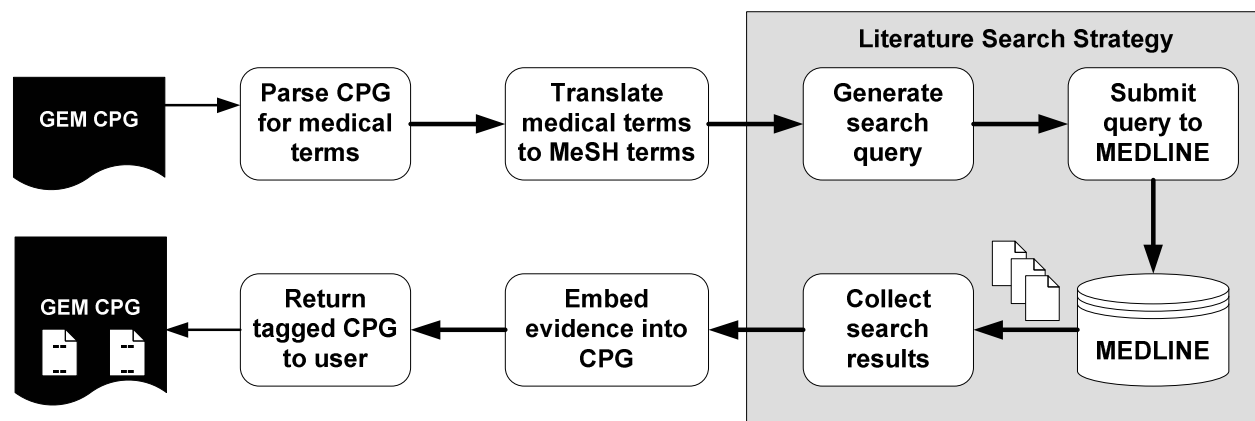


Figure 2. The functional overview of BiRD.

<directive.benefit source="explicit">
 Urinary tract infections (UTIs) are important because they cause acute morbidity and may result in long term medical problems, including hypertension and reduced renal function. Management of children with UTI involves repeated patient visits, use of antimicrobials, exposure to radiation, and cost. Accurate diagnosis is extremely important for two reasons: to permit identification, treatment, and evaluation of the children who are at risk for kidney damage and to avoid unnecessary treatment and evaluation of children who are not at risk, for whom interventions are costly and potentially harmful but provide no benefit. Infants and young children with UTI are of particular concern because the risk of renal damage is greatest in this age group and because the diagnosis is frequently challenging: the clinical presentation tends to be nonspecific and valid urine specimens cannot be obtained without invasive methods (suprapubic aspiration [SPA], transurethral catheterization).
 </directive.benefit>

Figure 3: UTI CPG content.

Table 1. The MeSH terms derived from the C-CPG content given in figure 2. Also, shown is the potential of a term as a trigger word for a particular query type and its age type.

MeSH Term	Frq	UMLS Semantic Type	Triggers	
			Sem	Age
Urinary Tract Infections	1	Disease or Syndrome	-	-
Urinary Tract	1	Body System, Body Part, Organ, or Organ Component	-	-
Urine	2	Body Substance, Functional Concept	-	-
Infections	1	Disease or Syndrome	-	-
Causes	3	Functional Concept	Etio	-
Morbidity	1	Quantitative Concept, Intellectual Product	Prog	-
Hypertension	1	Disease or Syndrome, Sign or Symptom	-	-
Kidney	3	Body Part, Organ, or Organ Component	-	-
Function	1	Biomedical Occupation or Discipline, Physiologic Function	-	-
Children	4	Age Group	-	Chi
Patient	1	Patient or Disabled Group	-	-
Use	1	Quantitative Concept	-	-
Radiation	1	Natural Phenomenon or Process	-	-
Cost	1	Quantitative Concept	-	-
Extremity	1	Body Location or Region	-	-
Diagnosis	2	Diagnostic Procedure, Functional Concept	Diag	-
Rat	1	Mammal	-	-

Sons	1	Family Group	-	-
Permit	1	Manufactured Object, Regulation or Law	-	-
Identification	1	Mental Process	-	-
Treatment	2	Therapeutic or Preventive Procedure	Tpy	-
Evaluation	2	Intellectual Product, Research Activity	-	-
Risk	3	Qualitative Concept, Quantitative Concept	Etio	-
DAME	2	Amino Acid, Peptide, or Protein	-	-
Infants	1	Age Group	-	Inf
Age Group	1	Age Group	-	-
Agar	1	Indicator, Reagent, or Diagnostic Aid, Pharmacologic Substance	-	-
Ageism	1	Individual Behavior, Social Behavior	-	-
Frequency	1	Temporal Concept	-	-
Methods	1	Intellectual Product, Organism Attribute, Intellectual Product	-	-
Aspiration	1	Organ or Tissue Function	-	-
Catheterization	1	Therapeutic or Preventive Procedure	Tpy	-

4.3. Search query generator

The literature search query builds on the MeSH terms provided by the C-CPG parser, and uses our multi-level literature search methodology to generate the most optimum literature search query.

4.3.1. Level 1: Determining query type. We first categorize the query into the following query types: *therapy, diagnosis, etiology and prognosis*. Our query classification strategy is as follows:

- Check whether the semantic type of the MeSH term is the trigger word for a query type.
- If yes, increment the count for the query type by the frequency of the MeSH term.
- Remove the MeSH term that has served as trigger word from the list of candidate MeSH terms.
- Perform the above for all the candidate MeSH terms.
- Finally, the query type with the highest count is deemed to be the winner and hence the most appropriate query type for the given set of candidate MeSH terms.

In case, two query types have the same count then the search query is assigned both the winning query types. However, if more than two query types have the same count then the search query is deemed as 'generic' and no query type is assigned to it.

Continuing with the exemplar C-CPG content shown in figure 3 and the MeSH terms derived from it (see table 1) we determine the query type as follows:

[Therapy:3, Diagnosis:2, **Etiology:6**, Prognosis:1]

Given that the query type *etiology* has the highest count (i.e. 6) we determine that the most appropriate query type for the selected C-CPG content is *etiology*. As per our approach, the trigger words for the designated query type are subsequently removed from the list of candidate query terms.

4.3.2. Level 2: Determining age group. At the second step of the query categorization stage, we categorize the query along the age dimension using the same strategy as used earlier, however here we search for trigger words for the age dimension.

According to our strategy we remove the MeSH terms that have served as a trigger word for determining the query type. In this case, the six (6) terms that had the query type ‘etiology’ have been removed from the list of available MeSH terms, leaving behind 24 terms.

Given the MeSH terms in table 1, the count for the four different age-groups is:

[Infant: 1, **Child:4**, Adult:0, Elderly:0]

Hence, the query is classified to target the *child* age-group. The candidate MeSH terms are reduced to 24 and are listed in the first column of table 2.

4.3.3. Level 3: Term filtering. In order to generate an optimal search query, the term filtering stage attempts to further reduce the dimension of the search query by removing insignificant candidate MeSH terms.

From a practical perspective term filtering is introduced because it was noted that a search query comprising more than 15 search terms is not handled well by MEDLINE. A search query with too many search terms is considered too specific by MEDLINE and as a consequence no medical articles are retrieved. It is hard to decide the optimal length of the search query. For our purposes, based on empirical results, we decided to set the default length of the search query to six (6) terms. If after the term filtering stage the length of the search query is greater than the pre-defined length then we remove terms on a term frequency basis—i.e. terms with a smaller frequency are removed. This process is characterized as *query rollback* and it involves the removal of one term (with the lowest frequency) at a time and then the submission of the relatively shortened search query to PubMed; if meaningful results are obtained then the search process is concluded, else the next lowest frequency term is removed and the process is repeated. We will like to point out that repetitive submissions of

search queries is not a practical solution considering the search submission constraints imposed by PubMed—i.e. PubMed accepts queries from a user every three seconds, hence BiRD waits for at least three seconds before presenting the next search query.

Term filtering is achieved via three (3) term filters (details are described earlier). Table 2 shows the outcome of the term filtering stage given the list of candidate MeSH terms retained from the query classification stage

At the conclusion of both query generation stages the following output is provided to the next module:

Query type = Etiology

Age-Group = Child

Search Query: {Kidney:3, Hypertension:1, Urinary Tract Infections:1, Patient:1, Radiation:1, Extremity:1}

Table 2. The outcome of the term filtering stage. The last column indicates the filter number responsible for removing the term. NF means ‘term not filtered’

Term	UMLS Semantic Type	Filter
Urinary Tract Infections	Disease or Syndrome	NF
Urinary Tract	Body System, Body Part, Organ, or Organ Component	3
Urine	Body Substance, Functional Concept	2
Infections	Disease or Syndrome	3
Hypertension	Disease or Syndrome, Sign or Symptom	NF
Kidney	Body Part, Organ, or Organ Component	NF
Function	Biomedical Occupation or Discipline, Physiologic Function	3
Patient	Patient or Disabled Group	NF
Use	Quantitative Concept	1
Radiation	Natural Phenomenon or Process	NF
Cost	Quantitative Concept	1
Extremity	Body Location or Region	NF
Rat	Mammal	1
Sons	Family Group	1
Permit	Manufactured Object, Regulation or Law	2
Identification	Mental Process	3
Evaluation	Intellectual Product, Research Activity	2
DAME	Amino Acid, Peptide, or Protein	3
Age Group	Age Group	3
Agar	Indicator, Reagent, or Diagnostic Aid, Pharmacologic Substance	3
Ageism	Individual Behavior, Social Behavior	3
Frequency	Temporal Concept	1
Methods	Intellectual Product, Organism Attribute, Intellectual Product	2
Aspiration	Organ or Tissue Function	3

4.4. PubMed query facilitator

This module is responsible for submitting the search query to PubMed and then collecting the results of the query. PubMed's E-utilities are incorporated within BiRD to facilitate the process. PubMed's E-utilities allow scripts or servlets to easily query the MEDLINE database.

The query rollback feature is called into action only when no results are obtained from the search query. This typically may happen when the presence of too many terms in the search query makes it too constrained. The query rollback feature removes one term at a time to reduce the query length. This process is repeated until results are returned. For instance, we had to make three queries for MeSH terms corresponding to our example.

Search Query 1: {Kidney:3, Hypertension:1, Urinary Tract Infections:1, Patient:1, Radiation:1, Extremity:1 }

Results: None

Search Query 2: {Kidney:3, Hypertension:1, Urinary Tract Infections:1, Patient:1, Radiation:1 }

Results: None

Search Query 3: {Kidney:3, Hypertension:1, Urinary Tract Infections:1, Patient:1 }

Results: 8 articles retrieved as listed below:

1. Rizvi SA, Naqvi SA, Hussain Z, Hashmi A, Akhtar F, Zafar MN, Hussain M, Ahmed E, Kazi JI, Hasan AS, Khalid R, Aziz S, Sultan. Living-related pediatric renal transplants: a single-center experience from a developing country. *Pediatr Transplant.* 2002 Apr;6(2):101-10.
2. Wennerstrom M, Hansson S, Hedner T, Himmelmann A, Jodal U. Ambulatory blood pressure 16-26 years after the first urinary tract infection in childhood. *J Hypertens.* 2000 Apr;18(4):485-91.
3. Roy S, Dillon MJ, Trompeter RS, Barratt TM. Autosomal recessive polycystic kidney disease: long-term outcome of neonatal survivors. *Pediatr Nephrol.* 1997 Jun;11(3):302-6. Erratum in: *Pediatr Nephrol* 1997 Oct;11(5):664.

4. Gschwend JE, Paiss T, Gottfried HW, Hautmann RE. Extracorporeal shockwave lithotripsy in children. Complications and long-term results. *Urologe A.* 1995 July;34(4):324-8.
5. Delaney VB, Adler S, Bruns FJ, Licinia M, Segel DP, Fraley DS. Autosomal dominant polycystic kidney disease: presentation, complications, and prognosis. *Am J Kidney Dis.* 1985 Feb;5(2):104-11.
6. Cheigh JS, Chami J, Stenzel KH, Riggio RR, Saal S, Mouradian JA, Fotino M, Stubenbord WT, Rubin AL. Renal transplantation between HLA identical siblings. Comparison with transplants from HLA semi-identical related donors. *N Engl J Med.* 1977 May 5;296(18):1030-4.
7. Johnston JH, Mix LW. The Ask-Upmark kidney: a form of ascending pyelonephritis? *Br J Urol.* 1976 Dec;48(6):393-8.
8. Gower PE. A prospective study of patients with radiological pyelonephritis, papillary necrosis and obstructive atrophy. *Q J Med.* 1976; 45(178):315-49.

4.5. Search result visualizer

This module is responsible for incorporating the search results within the GEM-encoded C-CPG. Recall that BiRD offers the functionality to select multiple C-CPG segments, where each selected C-CPG is treated as a separate query and the search results are therefore the search results need to be associated with the respective C-CPG section. The result visualizer module collects the individual search results—i.e. list of medical articles—for each selected C-CPG and embeds the search results within the C-CPG next to the respective C-CPG section (as shown in figure 4). The result visualizer presents only the title, authors, publication information and abstract of the retrieved article, with a hyperlink to the entire paper. This allows us to conserve the visualization space and yet allows practitioners to access the full article if they find the abstract interesting and relevant.

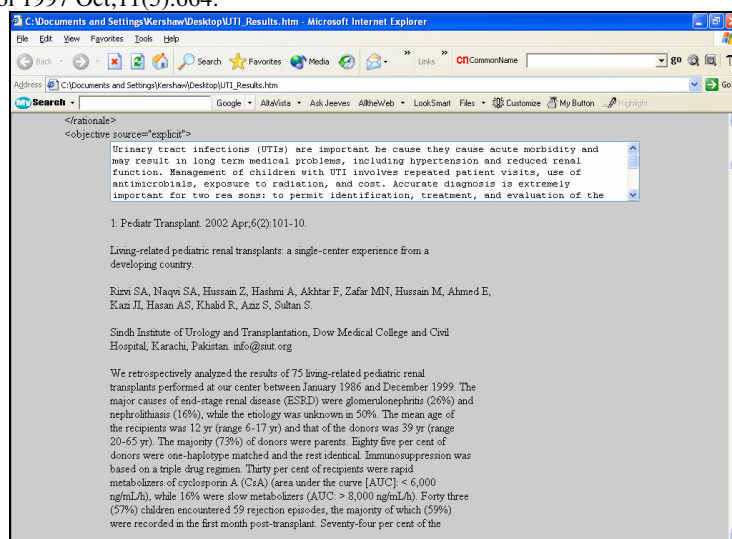


Figure 4. Screen shot of the search results embedded within the GEM C-CPG below the respective CPG content.

5. Evaluation

We evaluated the efficacy of the search strategy—i.e. the query-type and age-group classification—with the assistance of a urologist who is a domain expert. Based on the above-mentioned UTI_guideline 46 non-empty tags were identified as to have meaningful content that could be used to generate search queries. The evaluation set-up was as follows:

Step 1: The domain expert was provided the 46 different CPG content segments and was asked to manually determine the query-type and age-group of each CPG segment based on its content. Note that the domain expert was unaware of how BiRD had classified the query type and age-group for these CPG segments.

Step 2: The domain expert was then shown the classification performed by BiRD and asked to rate BiRD's output using a Likert scale.

5.1. Evaluation of query-type classification

For step1 evaluation, table 3 shows the extent to which the domain expert's classification compared with the classification performed by BiRD. Note that a CPG content can be simultaneously classified into two query-types. The label 'partial match' reflects this situation and implies that either the domain expert or BiRD suggested multiple query-types for the said CPG-content, however there is only a single common element between these two suggested outputs. Three partial match examples are given to illustrate this notion: (i) BiRD classifies query as *Diagnosis* and the domain expert classifies query as both *Diagnosis/Etiology*; (ii) BiRD classifies query as *Diagnosis/Therapy* and the domain expert classifies the query as *Prognosis/Therapy*; Domain expert classifies query as *Diagnosis* and the BiRD classifies the query as both *Diagnosis/Etiology*. For step 2 evaluation, the evaluation results are presented in table 4.

Table 3. Evaluation results for the query-type classification; comparing domain expert's classification with BiRD's classification of query type for 46 CPG segments.

Exact Match		Partial Match		No Match	
Number	%	Number	%	Number	%
27	59%	9	19%	10	22%

Table 4. Evaluation results for the query-type classification; asking the domain expert to rate BiRD's classification results for each of the 46 CPG segments.

Strongly Agree	Agree	Agree with Reservation	Disagree	Strongly Disagree
33	3	4	0	6

5.2. Evaluation of Age-Group Classification

The performance age-group classification was evaluated in a manner similar to that of query-type classification evaluation. In this case 26 CPG segments contained age-specific content and hence were accordingly classified by BiRD. Table 5 shows the degree to which domain expert agrees with BiRD's output (i.e. step 1 evaluation). Table 6 shows a rating by the domain expert to BiRD's output.

Table 5. Evaluation results for the query-type classification; comparing domain expert's classification with BiRD's query type classification of 46 CPG segments.

Exact Match		No Match	
Number	%	Number	%
22	85%	4	15%

Table 6. Evaluation results for the query-type classification; asking the domain expert to rate BiRD's classification results for each of the 46 CPG segments.

Strongly Agree	Agree	Agree with Reservation	Disagree	Strongly Disagree
25	1	0	0	0

6. Concluding Remarks

In this paper we have presented a medical literature search strategy that aims to generate an optimal search query for retrieving medical literature from MEDLINE. The featured search strategy builds a search query in an objective manner, as opposed to the traditional subjective manner of query specification by users, by using the original medical content as the basis for deriving the search query. We believe that such an approach would have implications in (a) easing the query specification process for health practitioners; and (b) improving the quality of best evidence retrieved.

Our work is a first step in realizing the need for supplementing C-CPG with best evidence—evidence that is autonomously sourced from medical literature archives. We have presented a working *proof-of-concept* that clearly demonstrates how to retrieve related evidence based on C-CPG content and how to link retrieved evidence vis-à-vis medical articles to respective C-CPG sections. We believe that this concept will be appealing to health practitioners and will improve the acceptance and efficacy of C-CPG in clinical settings. Our novel BiRD system, incorporating our multi-level search strategy—provides the basis for a testbed to extend the functionality of linking medical literature with other C-CPG formalisms. We have identified a number of issues that need to be addressed in the realm of future work, for instance investigating the possibility to use the semantically-explicit tags in the GEM document to assist

in query-type categorization. Furthermore, in the light of empirical evaluations we intend to fine tune some of the initial assumptions regarding the filterable semantic types, optimal query length and query rollback scheme.

We will like to point out that our query classification results are positive and vindicate our search query generation model. Indeed, the retrieval results are directly related to the query categorization accuracy. Having said that, query types are not necessarily formal, rather abstractions of medical queries typically posed by health practitioners. To formalize an objective set of query types is a major research endeavor and a task for future research.

7. References

- [1] Field MJ, Lohr KN (Eds). "Guidelines for clinical practice: from development to use", Institute of Medicine, Washington, D.C. National Academy Press, 1992.
- [2] Institute of Medicine. Guidelines for clinical practice: from development to use. Washington, DC, National Academy Press, 1992.
- [3] Shiffman RN, Liaw Y, Brandt CA, Corb GJ, "Computer-based guideline implementation systems: a systematic review of functionality and effectiveness", *Journal American Medical Informatics Association*, 1999, pp. 104–114.
- [4] Davis DA, Taylor-Vaisey A, "Translating guidelines into practice: A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines", *Canadian Medical Association Journal*, 1997, pp. 408–416.
- [5] Tierney WM, Overhage JM, McDonald CJ, "Computerizing guidelines to improve care and patient outcomes: The example of heart failure", *Journal American Medical Informatics Association*, 1995, pp. :316–322.
- [6] Lobach DF, Hammond WE, "Computerized decision support based on a clinical practice guideline improves compliance with care standards", *Journal American Medical Informatics Association*, 1997, pp. 89–98.
- [7] Maviglia SM, Zielstroof RD, Paterno M, Teich JM, Bates DW, Kuperman GJ. "Automating complex guidelines for chronic disease: lessons learned". *Journal American Medical Informatics Association*, 2003, pp. 154–165.
- [8] Chueh H, Barnett GO, "Just-in-time clinical information", *Acad Med*, 1997, pp. 512–517.
- [9] Chambliss ML, Conley J, "Answering clinical questions." *Journal of Family Practice*, 1996 (43), pp. 140–144.
- [10] Covell DG, Uman GC, Manning PR, "Information needs in office practice: are they being met?", *Annals of Internal Medicine*, 1985 (103), pp. 596–599.
- [11] Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH, Miller RA, "Physicians' information needs: analysis of questions posed during clinical teaching", *Annals of Internal Medicine*, 1991(114), pp. 576–581.
- [12] Timpka T, Ekstrom M, Bjurulf P, "Information needs and information seeking behaviour in primary health care", *Scandinavian Journal of Primary Health Care*, 1989 (7), pp. 105-109.
- [13] Gorman PN, Helfand M, "Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered", *Medical Decision Making*, 1995 (15), pp. 113-119.
- [14] Shiffman RN, Karras BT, Nath S, "GEM: A proposal for a more comprehensive guideline document model using XML", *Journal American Medical Informatics Association*, 2000 (7), pp. 488–498.
- [15] Berrios DC, Cucina RJ, Fagan LM, "Methods for Semi-automated Indexing for High Precision Information Retrieval", *Journal of the American Medical Informatics Association*, 2002 (9), pp. 637-652.
- [16] Cimino JJ, Aguirre A, Johson SB, Peng P, "Generic queries for meeting clinical information needs. Bulletin of the Medical Library Association. 1993 (81), pp. 195–206.
- [17] Cimino J, Mendonca E. "Automated knowledge extraction from MEDLINE citations", *Proc AMIA Symposium*, 2000, pp. 575-9.
- [18] Haynes B, Wilczynski, McKibbon A, Walker C, Sinclair J, "Developing optimal search strategies for detecting clinically sound studies in MEDLINE", *Journal American Medical Informatics Association*, 1994, pp. 447-458.
- [19] Clinical Queries (PubMed), NLM, USA (WWW page), <http://www.ncbi.nlm.nih.gov/PubMed/clinical.html>, Accessed June 9, 2004.
- [20] SUMSearch, Society of General Internal Medicine, University of Texas, USA (WWW page), <http://sumsearch.uthscsa.edu/>, Accessed June 09 2004.
- [21] Pratt W, Fagan L, "The usefulness of dynamically categorizing search results", *Journal of the American Medical Informatics Association*, 2000 (7) pp. 605-617.
- [22] Pratt W, Wasserman H, "QueryCat: automatic categorization of MEDLINE queries". *Proceedings of the American Medical Informatics Association (AMIA) Fall Symposium*, 2000. pp. 655-659.