Automatic Query Generation from Computerized Clinical Guidelines

Zafar Hashmi¹, Tatjana Zrimec¹, and Andrew Hopkins²

¹Center of Health Informatics, University New South Wales, Australia
²Cardiology Department, St George Hospital, Sydney, Australia

{Zafarh,Tatjana}@cse.unsw.edu.au, Andrew.Hopkins@sswhs.nsw.gov

تم تطوير إطار عمل يقوم بتوليد استعلامات محددة على حسب السياق آلياً ويقوم بتحديد نوعية الاستعلام من توجيهات المراجعة العينية (Computerized Clinical Guidelines) باستخدام الخدمات الشبكية لجبل وربط المقالات الطبية المتعلقة بمحترفي سي بي جي PubMed. يُستخدم إطار العمل المقدم السابق ودلالات المعلومات الإحصائية وكذلك المعلومات الواسعة للمعلومات الخاصة بلعبارات الطبية في المكونات المعترف بها المتعلقة (Extended knowledge Components) المكونات المعترف بها المتعلقة الحالية التي تم تجليها وربطها عن طريق إطار العمل المقدم لها علاقة بالعامة للسي بي جي الإلكترونية.
Automatic Query Generation from Computerized Clinical Guidelines

Zafar Hashmi¹, Tatjana Zrimec¹, and Andrew Hopkins²
¹Center of Health Informatics, University New South Wales, Australia
²Cardiology Department, St George Hospital, Sydney, Australia
{Zafarh,Tatjana}@cse.unsw.edu.au, Andrew.Hopkins@sswhs.nsw.gov

ABSTRACT: A framework has been developed that automatically generates context-specific query, and determines its query type, from computerized clinical practice guidelines (CPGs). The generated query can be submitted to PubMed using web services to retrieve and link relevant medical literature pertaining to the computerized CPGs content. This framework makes use of contexts, semantics, statistical information and meta-information of the medical phrases in the Extended-Knowledge Components of the computerized CPG content. The medical literature retrieved and linked, by our framework, is found to be relevant to the knowledge of the computerized CPG.

Keywords: Clinical guidelines computerization, Automatic query formulation, Medical literature retrieval.

Received: 12 May 2009, Revised 28 June, Accepted 12 July 2009

1. Introduction
Within the omnipresent internet, a large number of information systems for medical literature have been developed on the web [1, 2]. Such medical information systems, e.g. PubMed, MedlinePlus, Cochrane, provide comprehensive coverage of evidence based medical literature [1]. However, the search interfaces provided by these systems are less than ideal, potentially hindering health practitioners’ ability to fully utilize the data held within them [1, 3]. Most systems require the physician to formulate a focused query consisting of a few keywords, which need to be carefully selected to clearly define the information needs [1, 3-5]. Furthermore, the query may also require the selection of a query type such as “diagnosis”, “etiology”, “prognosis” or “therapies”. Previous studies have shown that searching for clinical information using the web-facilities is labor-intensive, time-consuming and complicated [4, 6-11]. Moreover, it has also been found that the health practitioner’s needs are often not satisfied [4, 8].

Clinical practice guidelines (CPGs) have been developed to provide comprehensive up-to-date interpretation of evidence-based best clinical practices [12]. Yet these complex guidelines are often not understood or, more importantly, fully implemented [13-15]. As a result there have been different methods suggested to transform CPGs into a computer understandable format [16-20]. Such computerized CPGs are used with decision support systems to provide evidence-based assistance [15, 21]. It has been observed that healthcare practitioners, when using CPGs, tend to validate or to supplement their understanding of the CPGs [22, 23]. Practitioners may refer to the current medical literature to acquire deeper insights into the clinical trials that support the clinical guidelines [23, 24].

There is an ongoing research to find efficient ways to fulfill healthcare practitioners’ information needs and clinical inquiries at point of care [4, 25, 26]. It has been investigated and shown [29] that automatically retrieving and linking relevant online medical evidences to CPGs would alleviate the burden on clinicians, demands on their time and would provide more focused-assistance at point of care [27-28]. As the clinical guidelines deal with variety of illness issues such as diagnosis, etiology, treatment, therapy etc, so, the linked medical literature should be contextually relevant to the CPGs content [23, 30]. Consequently, automatically linking online medical literature to the knowledge content of CPGs would provide potential benefits to healthcare practitioners. In order for this to be effective the linked medical literature must be retrieved by a focused, semantically and contextually correlated clinical query to the CPG content. We have developed an approach for automatically linking current best evidence to particular segments of computerized clinical practice guidelines for providing relevant and more focused information.

In this paper, we present our Context Specific Query Generation Framework (CQGF). This framework automatically generates queries from refined medical terms and phrases derived from computerized CPGs content [31, 32] and determines its query type. This is achieved by using the context, semantics and meta-information of the computerized CPG’s knowledge content. The generated query and its query type are sent to PubMed by using web-services to retrieve relevant medical literature pertaining to the CPGs’ knowledge content. The retrieved literature is linked and embedded along
with its abstract to the corresponding CPG’s knowledge content. The results of applying CQGF in cardiology domain are encouraging.

2. Representing Computerized Clinical Practice Guidelines with Extended Knowledge Components

Clinical practice guidelines aim to comprehensively detail the available evidence and make recommendations. However they can be complex and they lack consistent structure, making them unsuitable for computerized applications. As a result it is necessary to segment the information contained in CPGs to allow further computerized analysis and the generation of effective algorithms to condense the key CPGs knowledge base into effective keywords to be used in a literature search. To this end we have developed a CPGs computerization framework that models the key aspects of clinical practice guidelines and [31, 32] an effective segmentation of CPG data. This framework is based on the document-centric GEM model [18] which is an XML format designed to allow the development of decision support systems based on CPGs. We have extended the GEM model’s Knowledge Components node to represent the knowledge structure of CPG’s content, and we termed these Extended-Knowledge Components. This representation models the clinical practice guidelines content in concise, yet clinically-focused segments of knowledge content. We have enriched the Extended-Knowledge Components with important information including a contextual impact factor, semantics and other meta-information pertaining to CPG content. These ‘Extended Knowledge Components’, created automatically, are computer-interpretable.

This process allows the CPG to be represented by a set of Extended-Knowledge components, which can be formally represented by equation 1. A CPG is represented by ‘C’ and an extended knowledge component is represented by ‘Ex-KC’.

\[ C = \{Ex-KC_1, Ex-KC_2, \ldots, Ex-KC_n\} \] (1)

A segment about Reperfusion therapy from the *Australian clinical practice guidelines for the management of acute coronary syndromes 2006 (ACS_06)*[33] is shown in Figure 1. Figure 2 shows an excerpt of an Extended Knowledge component (Ex-KC ) as a result of the computerization of that segment of the CPG.

3. Context Specific clinical Query Generation Framework (CQGF)

To retrieve ‘relevant’ medical literature related to the CPGs’ knowledge content, medical practitioners need to formulate a search query. Such a search query should manifest CPG-context, semantics, medical problem and clinical intention by using focused medical phrases [1, 3-5]. Multiple studies have investigated healthcare practitioners’ information needs and information seeking behavior [4, 6, 7, 25, 34-38]. Based on their findings, different query models and search strategies as well as techniques to categorize queries into query types have been proposed. However, concepts like scenario specific query, query types, tasks oriented query, query templates, and query categories have almost same meaning. Recent studies reveal that almost 60 % of clinicians queries center on specific query types or scenarios [2, 7, 10, 34-37, 39]. Cimino et al. (2002) showed that clinicians’ questions are predictable and they focus on specific tasks. Therefore, focused and context-sensitive medical queries, either from a healthcare practitioner or system-generated, should appropriately frame the problem and clinical intention. This can be achieved by using a list of specialized problem-specific medical-terms and concepts as the search query [5, 23].

![Management of patients with ST-segment-elevation myocardial infarction](image)

Reperfusion therapy

Reperfusion may be obtained with fibrinolytic therapy or PCI. A combination of fibrinolysis and PCI may also be used (facilitated or rescue PCI). Coronary artery bypass graft (CABG) surgery may occasionally be more appropriate — particularly in patients who have suitable anatomy and are not candidates for fibrinolysis or PCI.

Figure 1. A segment of content from ‘ACS 06’ CPG entitled reperfusion therapy
In devising our framework for query generation the following objectives were considered:

1. Retrieved literature should be focused and relevant to Extended-Knowledge Components since they are highly focused for specific clinical problems/procedures.
2. The search query should be objectively and automatically derived from the content of computerized CPGs (i.e. from Ex-KCs).
3. To formulate a more focused search query by exploiting the Ex-KC syntactic structure (elements tags), contextual, semantic and other meta-information.
4. To use a query model that captures the scope and intentions of Ex-KC. As most of the clinicians’ search queries center on specific scenarios or query types.

To achieve the above defined objectives, we have developed a technique that (a) extracts potentially important medical terms and phrases based on the contextual importance, (b) filters out insignificant medical candidates, (c) exploits the information about the semantic types to find association between medical phrases and (d) categorizes search query based on defined clinical query intention defined in the query model. The resulting context-specific query is used to query MEDLINE using the "web service" provided by PubMed’s E-utilities. After retrieving relevant medical articles, they are systematically linked and embedded within the Ex-KC. In the following section we describe our algorithm that carries out above defined tasks.

3.1 Algorithm for Context-Specific Clinical Query Generation Technique

The heuristic for our technique of context-specific query generation from Extended-Knowledge Components is shown in the flowcharts in Figures 3, 4 and 5. Specifically (Figure 3), potentially significant medical term candidates are extracted from Ex-KCs. These medical term candidates are passed through a redundancy filter to eliminate redundant medical term candidates. Remaining medical term candidates are evaluated based on the significance of their semantic types. This is done by applying a semantic type filter to remove those medical term candidates that have semantically less significance. The resulting medical term candidates are then analyzed and additional meta-information is attached to each candidate by using Term frequency analysis and by Total weight calculation.

Figure 2. An Excerpt of an extended knowledge component (EX-KC) resulted from the computerization of ACS_06 CPG segment of reperfusion therapy procedures
A disease ontology filter is applied to each potentially significant medical term candidate to further specify the meaning of each medical term candidate. Those medical term candidates that do not conform to disease ontology are filtered out.

In the next step, three sets of potentially significant medical term candidates are created:

- **Total weight (TW) set** - In this set, all filtered medical term candidates that are potentially significant are sorted based on their total weight. The details for the total weight technique are contained in 3.2.4 below.
- **Semantic relation score (SMR) set** - This set has the same medical term candidates as in the TW set, but the only difference is that the medical term candidates are sorted based on a Semantic Relation Score (SMR score). We have quantified the mutual-relation of each medical term candidate based on their semantic types. The details of this technique will be discussed in 3.2.6.
- **Knowledge component (KC) set** - This set consists of all those significant medical term candidates which are found under the KC element of an Ex-KC.

To formulate the final query set, a threshold is applied that is used to determine the number of medical term candidates for a final query set. If the number of potentially significant medical candidates is less than or equal to the threshold then all candidates are members of the final query set. The medical terms and phrases are used to find the query type (details in 3.2.8). If the number of medical term candidates is greater than the threshold then a **Common Term Analysis** is applied to the above defined three sets to reduce the number of candidates.

Figure 4 shows the steps for the **Common Term Analysis**. With this analysis, common medical terms and phrases that are within certain threshold values from the three sets of medical term candidates are identified (details in 3.2.7). The threshold values are based on the highest value of the **total weight** and the highest value of the **SMR score** of the candidates in **TW set** and **SMR set**. The goal here is to find a candidate that is common in at least two sets and has highest TW score and SMR score. This candidate becomes a member of the final query set. The threshold values vary between 100% to 25% of the highest values of the **total weight and SMR score**.

---

**Figure 3. Flowchart of search query generation from Ex-KC – Generation of potentially significant medical candidates**
The **Common Term Analysis** is continued until the number of candidates in the final query set reaches the threshold. Once the final query set is created the medical candidates are used to find the query type (see Figure 5).

### 3.2 Functional Flow of the Context Specific clinical Query Generation Framework (CQGF)

In this section, we describe the functional flow of our CQGF framework and explain its modules. The CQGF architecture with its ‘functional-overview’ is shown in Figure 6. As described earlier (section 2), after computerization of CPG, its content is transformed into ‘Extended-Knowledge Components’ (Ex-KCs) that constitute the ‘Ex-KC knowledge base’. Each Ex-KC is

---

**Figure 4.** Flowchart of the Common Term Analysis process

---

**Figure 5.** Determining the Query Type of a query set
taken one by one to be processed by CQGF framework so that a ‘Context-specific’ query from the Ex-KCs can be generated. The resulting queries are submitted to retrieve relevant medical articles pertaining to the Ex-KCs. These medical articles are linked and their abstracts are embedded within the corresponding Ex-KC.

3.2.1 Medical Terms Extractor

The ‘Medical Phrase Extractor’ module extracts medical phrases from Ex-KCs. This module extracts only those medical phrases/terms from Ex-KCs that have specific ‘contextual impact factors’, as defined by the GEM model, within the Ex-KCs. The significant contextual segments within the Ex-KCs are (i) ‘Knowledge Component’ segment, (ii) ‘Recommendation’ segment, (iii) ‘Decision Variable’ segment, (iv) ‘Action’ segment and (v) ‘Imperative’ segment. The medical phrases within these five segments in the Ex-KC have relatively higher impact factors than other elements, so that, these have greater potential to be representative of the Ex-KCs. The ‘Medical Phrase Extractor’ sends the extracted medical phrases and terms to a ‘Redundancy Filter’

3.2.2 Redundancy Filter

There is a strong likelihood of redundancy in a set of extracted medical term candidates due to the same term being present in several different segments of the Ex-KC. The ‘Redundancy Filter’ removes all the redundant medical term candidates from the potential query set. The set of remaining medical term candidates is sent to ‘ST Filter Module’.

3.2.3 Semantic Type (ST) Filter

The Semantic Type (ST) Filter module filters out those medical candidates whose semantic types are not considered to be significant. This module retrieves the ULMS (Unified Medical Language System) semantic types for all the medical terms / phrases. Then it uses a list of “permissible” semantic types to select those medical term candidates that are relevant for the domain. We have generated a list of semantic types relevant for Cardiology after consulting with specialists in the cardiology domain and after a qualitative analysis of the usage of these semantic types. With additional experiments conducted with clinical practice guidelines (CPGs) for heart disease we generated a list of semantic types best represented CPGs in Cardiology.
3.2.4 Info-Manager

The ‘Info-Manager’ module for each potentially significant medical term candidate adds the following weights: ‘Term Frequency’, ‘Total Contextual Weight’ and ‘Total weight’.

‘Term Frequency’ is calculated by a frequency analysis of a medical term candidate to find the number of times this medical term candidate appeared in the corresponding Ex-KC. After frequency analysis a “Total Contextual Weight” of each medical candidate is calculated using the Contextual Weight. Every element in Ex-KC refers to a specific context e.g. Recommendation, Decision Variable etc. Such contexts of different elements have been quantified based on their impact factor, into contextual weights (represented as $\eta$) [40]. A medical term candidate belonging to a certain element is given the contextual weight of that element. Since the medical term candidate may appear in different elements within the Ex-KC, the aggregation of its contextual weights is calculated, which is named Total Contextual Weight.

The Total Contextual Weight of any medical phrase is a summation of contextual weights of the medical phrase within Ex-KC that is defined in equation (2). In this equation, $T_{\eta_{\text{mp}_i}}$ is a Total Contextual Weight of a medical candidate ($\text{mp}_i$) in ‘ith’ number of Ex-KC ($\alpha$). The “$\eta_{\text{mp}_j}$” is defined as the Contextual Weight of ‘jth’ mp in ‘ith’ number of Ex-KC ($\alpha$) at context-structure (CTS). ‘N’ represent the number of time that $\text{mp}$ is found in $\alpha$.

\[ T_{\eta_{\text{mp}_i}} = \sum_{j=1}^{N} \eta_{\text{mp}_j} \]  \hspace{1cm} (2)

After calculating the Total Contextual Weight of a medical term candidate, the Total Weight for the medical term candidate is calculated by two factors (a) Term-Frequency and (b) Total Contextual Weight. So that the Total weight is a summation of these two factors. If the Total Weight is represented by ($\psi$) and the Term Frequency is represented by ($\lambda$) then the Total Weight can be formally defined in equation (3), where $T_{\eta_{\text{mp}_i}}$ is the “Total Contextual Weight of ‘medical candidate’” as defined in equation (2).

\[ \psi = \lambda + T_{\eta_{\text{mp}_i}} \]

3.2.5 Diseases Ontology Filter

‘Diseases Ontology’ represents domain knowledge that provides relationship information concerning different entities such as diseases, symptoms, diagnosis, procedures, etc, pertaining to the corresponding disease. Such domain knowledge plays an important role in identifying the essence of different medical terms. We have developed a prototype of diseases ontology for our framework. This Diseases Ontology can be extended to be applied to any domain in general.

For cardiology domain, we developed a “Heart Diseases Ontology” (HDO) that defines the relationship among heart diseases, symptoms, therapies, diagnosis procedures, and diagnostic clues. In Heart Diseases Ontology, we use ‘Has a’ relationship between diseases, symptoms, therapies, diagnosis, and diagnostic procedures. For example ST segment elevation myocardial infarction has the following symptoms: chest pain, arm pain etc. We used ICD10, SNOMEDCT, MeSH and the information provided by cardiology specialists.

The “HDO-Filter” module uses this ontology and filters out those medical term candidates that do not conform to the HDO.

3.2.6 Semantic Relation Score Module

The UMLS provides relationships between different semantic types. We used the “associated-with” relation that determines whether one semantic type is associated with another semantic type. Since every medical term candidate has a semantic type, analyzing the ‘associated with’ relation of semantic types, we were able to determine whether the medical term candidate is associated-with ‘none’, ‘one’, or ‘many’ medical phrases. The rationale behind this is the more associated a ‘medical candidate’ is with other medical term candidates, the more potential it has to be included in the final query set.

We quantified the ‘associated-with’ relation by giving score of “1”, if medical candidate is found ‘associated-with’ other medical candidate. The score of “0” is given if ‘associated-with’ relation is not found with other medical phrase. It is formally represented by equation 4 where the ‘associated with relation’ between two medical candidates ($\text{mp}_i$ and $\text{mp}_j$) in Ex-KC ($\alpha$) is represented by ($\text{SMR}_{\text{rel}} = \xi$).

---

International Journal of Information Studies  Volume 1 Issue 4  December 2009  287
Upon receiving the medical term candidates, the Final Query Set Creator module creates the three sets of medical term candidates as described in section 3.1: a Total Weight (TW set), with medical term candidates sorted by the Total Weight, an ‘SMR score’ (SMR set) where medical candidates are sorted by the SMR score and a ‘KC’ element set of medical candidates found under the of the Ex-KC (KC set). The rationale behind the third set is that the ‘KC’ element within Ex-KC represents Ex-KC title so the medical term candidates appearing in the KC element have additional significance. If the KC element medical term candidate is also found in the TW set or the SMR set, it increases the likelihood that the corresponding medical term candidate will be a member of final query set.

The number of medical term candidates in the final query set is then adjusted. The Final Query Set Creator evaluates the number of received medical candidates against the threshold value (5 terms) for final query set. If the number of received medical candidates is less than the threshold value, it tries to find medical term candidates from the ‘KC set’ which do not appear in the TW set or the SMR set. The selection of medical term candidates from the KC set is based on the ‘Total weight’, as assigned by the “Info-manager”. Those candidates are used to complete the number of medical terms of the final query set. Otherwise the final query set contains the received elements. The resulting final query set is sent to the Query Type Finder.

If the initial number of received medical term candidates is more than 5, then Common Term Analysis is applied to select the first 5 best medical term candidates. The heuristic used in the Common Term Analysis is based on three different threshold values: 100%, 50% and 25% of the highest value (as shown in Figure 4) in the TW set and the SMR set. All candidates that satisfy the specified conditions are moved to the Final Query Set which is then sent to the Query Type Finder.

3.2.8 Query Type Finder

Different models have been proposed to categorize the intentions of clinicians for their information needs. In our work, we used PubMed to find relevant medical literature. PubMed has adopted a query model proposed by Haynes [36, 38] for clinician query classification. In line with the query classification at PubMed, we categorized our search query into four types: (i) Diagnosis: literature related to evaluation of disease process, (ii) Etiology: literature related to causation of disease or condition, (iii) Prognosis: literature related to prediction or forecast of a course of disease or condition, (iv) Therapy: literature related to therapy, prevention, or rehabilitation. If the medical terms / phrases belong to two query types, both query types are assigned [29] to the final search query. If the search query falls into more than two categories or does not fall into any above defined category, it is assigned as a “General” type.

The final search query set was automatically categorized by these four categories. Figure 5 shows the process for determining a query type. A set of semantic types is assigned to each query type. These sets were generated in consultation with cardiologists following the methods provided in the literature [2, 36, 41-43]. Each medical term candidate is associated with at least one semantic type in UMLS. The query type for a search query is determined by finding the maximum number of medical terms / phrases that belong to a specific query type.
3.2.9 PubMed Webservice

The ‘PubMed Webservice Module’ is used to connect to PubMed, to send a search query and to receive the retrieved medical articles. It uses the Pubmed E-utilities features and implements the specification defined by PubMed. This module transforms a final search query into the format defined by PubMed and implements our heuristic to retrieve, link and embed information in the corresponding Ex-KC. If medical literature is not retrieved for the submitted query, this module removes the least significant medical term / phrase form the query based on the Total Weight and SMRscore value and re-submits the new query to PubMed. This process is repeated until a result set of medical literature is retrieved or the query terms are exhausted.

4. Working Example of CQGF Framework

This section presents a working example of the CQGF framework functionalities. First, a section of knowledge content is taken from the clinical practice guideline (CPG) “Management of acute coronary syndromes 2006” (ACS 06) [33]. In this example the section entitled “Reperfusion Therapy” was used and details the management of patients with ST-segment-elevation myocardial infarction with thrombolytic agents, and is shown in Figure 1. Using the GEM Cutter tool and our CPG computerisation framework this section was transformed into an Ex-KC. Figure 2. shows an excerpt of the resulting “Reperfusion Therapy” Ex-KC.

Second, the CQGF analyses the “Reperfusion Therapy Ex-KC” to find potentially significant medical term candidates as described in section 3.2.1. Next, redundant medical terms are filtered out from the extracted medical terms and medical phrases by the Redundancy Filter. The resulting set of remaining medical terms and phrases are listed, along with their semantic types, in Table 1. Further filtering is applied to this set of medical terms by the Semantic Type Filter that removes insignificant terms based on their semantic types (as described in section 3.2.3). The highlighted rows in Table 1 indicate those medical terms that remained after the application of the Semantic Type Filter. This subset of medical terms is then processed by the Info-Manager to add weighting information to the medical terms and phrases (see section 3.2.4). At this stage there were eleven potentially significant medical terms. These eleven terms were further processed by the Heart Diseases Ontology Filter to further stratify the relative importance of the medical terms.

<table>
<thead>
<tr>
<th>Medical Phrase/Term</th>
<th>Semantic Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>reperfusion therapy</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>science of anatomy</td>
<td>Biomedical Occupation or Discipline</td>
</tr>
<tr>
<td>thrombolytic therapy</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>assay of fibrinolysis</td>
<td>Laboratory Procedure</td>
</tr>
<tr>
<td>transplantation</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>operative surgical procedures</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>Physiologic Function</td>
</tr>
<tr>
<td>using</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>patients</td>
<td>Patient or Disabled Group</td>
</tr>
<tr>
<td>combined</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>methodology</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>myocardial infarction</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>surgery specialty</td>
<td>Biomedical Occupation or Discipline</td>
</tr>
<tr>
<td>anatomy aspects</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>percutaneous coronary intervention</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>coronary artery bypass surgery</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>anatomy</td>
<td>Anatomical Structure</td>
</tr>
<tr>
<td>protein c inhibitor</td>
<td>Biologically Active Substance</td>
</tr>
<tr>
<td>transplanted tissue</td>
<td>Tissue</td>
</tr>
<tr>
<td>surgical aspects</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>procedures</td>
<td>Health Care Activity</td>
</tr>
<tr>
<td>st segment elevation</td>
<td>Finding</td>
</tr>
<tr>
<td>utilization qualifier</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>graft material</td>
<td>Biomedical or Dental Material</td>
</tr>
</tbody>
</table>

Table 1. Remaining medical phrases and terms after ‘Redundancy Filter’. Highlighted rows indicate those terms that are filtered by Semantic Type Filter.
After HDO filtration, seven medical phrases are left which are further processed by the SMR score module (as described in section 3.2.6). These medical terms are shown along with their Total weight and their SMR score in Table 2. In Table 2 the medical candidate “fibrinolysis” has a SMR score of “0” because, its semantic type is “Physiologic Function” that does not have an “associated-with” relation with “Therapeutic or Preventive Procedure”, “Disease or Syndrome”, or “Finding” semantic types in UMLS. The number of the remaining medical terms and phrases for the final query set is greater than the pre-defined threshold for a final query set (THFQS), that is ‘5’ terms. To further reduce the number of terms in the query set, three sets were created. The Common Term Analysis was performed using the 3 sets (see section 3.2.7). TW set and SMR set terms and their values can be found in Table 2. The KC set medical terms were “reperfusion therapy, methodology, and procedures”.

After the Common Term Analysis five medical terms / phrases remained in the final query set.

FinalQuerySet = {reperfusion therapy, percutaneous coronary intervention, myocardial infarction, ST segment elevation, thrombolytic therapy}.

The query type for the FinalQuerySet was determined (section 3.2.6) to be “Therapy”. The final query submitted to the PubMed was {reperfusion therapy, percutaneous coronary intervention, myocardial infarction, st segment elevation, thrombolytic therapy} with query type “Therapy”. The medical articles retrieved and linked to Reperfusion Therapy Ex-KC are shown in Table 3.

<table>
<thead>
<tr>
<th>Medical Phrase/Term</th>
<th>TW</th>
<th>Semantic Type</th>
<th>SMR score</th>
</tr>
</thead>
<tbody>
<tr>
<td>reperfusion therapy</td>
<td>4.75</td>
<td>Therapeutic or Preventive Procedure</td>
<td>1.0</td>
</tr>
<tr>
<td>percutaneous coronary intervention</td>
<td>2.5</td>
<td>Therapeutic or Preventive Procedure</td>
<td>1.0</td>
</tr>
<tr>
<td>ST segment elevation</td>
<td>1.75</td>
<td>Finding</td>
<td>1.0</td>
</tr>
<tr>
<td>myocardial infarction</td>
<td>1.75</td>
<td>Disease or Syndrome</td>
<td>5.0</td>
</tr>
<tr>
<td>thrombolytic therapy</td>
<td>1.5</td>
<td>Therapeutic or Preventive Procedure</td>
<td>1.0</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>1.5</td>
<td>Physiologic Function</td>
<td>0.0</td>
</tr>
<tr>
<td>coronary artery bypass surgery</td>
<td>1.5</td>
<td>Therapeutic or Preventive Procedure</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 2. Resulting medical terms/phrases from Heart Diseases Ontology Filter

Retrieved and Linked Medical Articles for Reperfusion Therapy Ex-KC


5. Discussion and Evaluation

The clinical guideline, *Australian clinical practice guidelines for the management of acute coronary syndromes 2006*, was computerized using our CPG computerization framework. This guideline is widely used in the evaluation and management of patients presenting to Emergency Departments with acute coronary syndromes in Australia. The computerized CPG in form of EX-KCs was processed with CQGF framework to generate clinical queries, and to retrieve medical literature from MEDLINE. Domain experts (specialists from the Cardiology domain) were asked to analyze and evaluate the results in the following settings.

First, the generated clinical queries were evaluated based on their relevancy: (i) Relevant and (ii) Not relevant. ‘Relevant’ indicates that clinical query is relevant to the corresponding segment of CPG and ‘Not relevant’ indicates that clinical query does not represent the corresponding content. CQGF generated query for this evaluation were marked by domain experts. Table 4 shows the results of this evaluation.

Another evaluation of the generated query was conducted along the lines of three categories: (i) Relevant, (ii) Partially Relevant, and (iii) Not-relevant. For this evaluation ‘Relevant’ indicates that generated query exactly represents the CPG content, ‘Partially relevant’ indicates that generated query is relevant but medical terms may be added or deleted, and ‘Not relevant’ indicates that generated query does not represent CPG content at all. Table 5 shows the results for this evaluation.

To evaluate the ‘Query Type’ generated by the CQGF for clinical query, domain experts were asked to provide their feedback along the lines of two measures: (i) Agree, and (ii) Disagree. Table 6 shows the results for this evaluation.

We conducted another evaluation of the ‘Query Type’ along the lines of the following measures: (i) Agree, (ii) Agree with reservation, and (ii) Disagree. In this evaluation, ‘Agree’ indicates that query type is correct, ‘Agree with reservation’ indicates that Query Type is marginally correct but it could be different as well, and ‘Disagree’ indicates that Query Type is not correct. Table 7 shows the results for this evaluation. CQGF assigned some of the queries two query types e.g. Diagnosis and Therapy. In this case, domain expert classified the Query Type under “Agree with reservation” category.

CQFG depends on the PubMed retrieval technique to retrieve the medical literature using the generated query and query type. We also asked the domain expert to analyze the retrieved articles for segments of CPG content. For this analysis maximum twelve retrieved medical articles were analyzed. Retrieved medical articles were evaluated for two categories (i) Relevant and (ii) Not-relevant. In this evaluation ‘Relevant’ indicates that retrieved articles are relevant with the possibility that some...
of the articles are not directly relevant to CPG content, and ‘Not-relevant’ indicates that retrieved articles are not directly relevant to CPG content. Table 8 shows the results of this evaluation.

<table>
<thead>
<tr>
<th>Relevant</th>
<th>Not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>88.8 %</td>
<td>11.2 %</td>
</tr>
</tbody>
</table>

Table 8. Results for relevancy of retrieved medical articles

On reviewing the evaluation by domain experts, we found that queries and query types generated for those CPG segments that have: medications information, quantity of medication dose, and usage of those medications, were not good match. Such information also does not appear quite often in medical articles titles and abstract, which affects the relevancy of retrieved articles as well.

6. Concluding Remarks and Future Work

In this paper, we have presented our technique that automatically generates clinical query from semi-structured, information-rich segments of clinical practice guideline’s content. This technique exploits the semantic tags, contextual information, semantic relation, and relevant meta-information of CPG content to formulate focused, context-specific query and determines its query type. This technique is implemented in our ‘Context-specific Query Generation Framework’. This generated clinical query is used to query MEDLINE database. By using online PubMed facilities the system retrieves and links medical articles with the corresponding segments of a CPG content. Automatically generating clinical query to provide articles pertaining to CPG content can improve the use of computerized CPGs in clinical settings. It also would help alleviate some of the burden in seeking relevant information at point of care.

We have identified number of issues that need to be addressed in future work for improving the automatic delivery of medical evidence pertaining to computerized CPG content. This includes: investigating the ways to enhance the accuracy of the semantic filter, to develop more detailed domain ontology, investigating other semantic tags in computerized CPG to increase the precision of clinical query formulation, finding more optimal query length and investigating other techniques to improve query categorization.

References


