The healthcare services center around the wellness, maintenance, and illness management of patients. The delivery of such services heavily depends on up-to-date medical knowledge and experience. However, miscommunication, misinterpretation of diverse forms of information exchanged between clinical settings, inability to be up-to-date with clinical best practices, and accessing rapidly changing and exponentially growing medical knowledge cause errors in healthcare that are leading cause of injuries and deaths. In this thesis, we addressed the research problems related to the interpretation and understanding of medical referral and response letters, exchanged between Specialists and General Practitioners (GPs) for patients care decision making. These research problems are implicitly associated with GPs information needs at point of care, including management of and access to evidence-based medical information and knowledge relevant to medical letters. The interpretation and understanding of referral and response letters and sending alerts for critical situations at point of care require methods for medical information processing, and methodologies for modeling and management of clinical knowledge. It also requires techniques for efficient and context-sensitive retrieval of evidence-based clinical knowledge. We have taken interdisciplinary solution approach along the lines of healthcare knowledge management, contextual information retrieval, and knowledge-based search strategies. We have formulated a knowledge modeling methodology and a computerization technique for clinical practice guidelines, which transform them into computer interpretable segments that are enriched with content-specific meta-information. To link these CPGs segments with online evidence-based medical knowledge, we have developed a technique for automatically generating clinical queries from the segments. These queries are used in a framework to retrieve and link online medical literature with corresponding segments. We have developed a method for computerized processing of referral and response letters. This method analyzes medical information and provides a comprehensive information-view of the letter to help healthcare practitioners formulate customized information specifications to access required knowledge. We also have developed a technique for contextual and statistical analysis of medical concepts and indexing strategy, which are used to retrieve CPGs segments and related medical literature. Finally, we have designed a healthcare-knowledge mediated architecture and implemented a computer system for Clinical Knowledge Assistance (CKA).

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Micr^Hfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).
STATEMENT OF ORIGINALLITY

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ..........................................

Date ........................................... 21/09/2010

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed ..........................................

Date ........................................... 21/09/2010
COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed ..........................................................

Date .......................................................... 21/09/2010

IV
I would have never been able to complete this thesis without the assistance by a multitude of individuals throughout all stages, many of whom through lack of space I will be unable to name. It is my pleasure to acknowledge their efforts, support, love, and guidance throughout the course of my PhD thesis.

First and foremost, I am thankful to Allah for His blessings and love as the eternal guardian for me to give me enough power and perseverance to go through the tough times of the research. All powers and glory goes to Him, now and forever.

I would like to gratefully acknowledge the supervision of Dr. Tatjana Zrimec for many insightful conversations during the development of the ideas in this thesis, for helpful comments, for her dedication, guidance, and for having presented me with the numerous opportunities to broaden my academic and social horizon. I would like to thank heartedly Dr. Andrew Hopkins, for his kind assistance and guidance in the medical field of cardiology, his feedback, discussions, comments and impressive way of conveying thoughts. I am indebted to Dr. Ashesh Mahidadia for his useful thoughts and support during my work.

I wish to express my gratitude to my colleagues and seniors at the Center for Health Informatics for their useful advice, discussions, support and fellowship. Financially, I would like to thank University New South Wales for funding my scholarship. My special thanks go to my friends in Sydney, Shahzad, Haris, Arif, Shoaib, Wissam, Usman, Naweed, and others for providing me humorous moments and constant source of
entertainment during the boring times throughout this journey. I would always remember them in my experience of choosing restaurants for dinners, late night suppers and chilling out.

I am extremely grateful to my parents: Muhammad Iqbal Hashmi and Suriya Iqbal, for their love, care, perseverance and determination for encouraging me always to accept challenges and pursue the journey of PhD. I dedicate this work to them. I am greatly thankful to my brothers and sisters: Azhar Iqbal Hashmi, Faisal Iqbal Hashmi, Saima Iqbal Hahsmi, and Asma Iqbal Hashmi for their love and moral support. Finally, my special gratitude and love to Somaia for her affection, encouragement, moral support, lovely nature and always being there for me.
# TABLE OF CONTENT

THESIS/DISSERTATION SHEET ................................................................................. II
STATEMENT OF ORGINALITY ............................................................................. III
COPYRIGHT STATEMENT ...................................................................................... IV
ACKNOWLEDGMENTS .......................................................................................... V
TABLE OF CONTENT ........................................................................................... VII
LIST OF TABLES AND FIGURES ......................................................................... XII
ABSTRACT ............................................................................................................ XVI
LIST OF ABBREVIATIONS ................................................................................... XVIII
LIST OF PUBLICATIONS ..................................................................................... XX

## CHAPTER 1
INTRODUCTION ..................................................................................................... 1

1.1 Preamble ......................................................................................................... 1
1.2 An Overview of Problem Domain ................................................................. 1
1.3 Motivation, Aim and Objectives ..................................................................... 4
1.4 Research Focus ............................................................................................. 6
1.4.1 Problem Statement and Solution Approach ............................................ 6
1.5 Contributions ................................................................................................ 18
1.6 Clinicians’ Behavioral Interaction with Clinical Knowledge Assistance System...
   ... 19
   1.6.1 General Practitioner Interaction Use Cases ........................................... 20
   1.6.2 Specialist Interaction Use Cases ........................................................... 23
1.7 Thesis Outline ............................................................................................... 25

## CHAPTER 2
BACKGROUND: A LITERATURE SURVEY .......................................................... 27

2.1 Introduction .................................................................................................... 27
2.2 Formalisms for Representing Clinical Practice Guidelines .......................... 27
   2.2.1 Overview of Knowledge-base-centric Formalisms ................................. 28
   2.2.1.1 PROforma .................................................................................. 28
   2.2.1.2 GLIF .......................................................................................... 30
   2.2.1.3 GLARE .................................................................................... 31
   2.2.1.4 SAGE ....................................................................................... 32
   2.2.1.5 EON ......................................................................................... 33
   2.2.1.6 Asbru ........................................................................................ 34
   2.2.1.7 Arden Syntax ........................................................................... 35
   2.2.1.8 GUIDE .................................................................................... 35
   2.2.2 Overview of Document-Centric Formalisms ......................................... 36
   2.2.2.1 HGML .................................................................................... 36
   2.2.2.2 GEM ........................................................................................ 36
   2.2.3 Guideline Formalisms Representation Languages .................................... 39
   2.3 Computerized Clinical Practice Guidelines Content Access at Point of Care... 40
CHAPTER 3
RESEARCH METHODOLOGY .......................................................... 69

3.1 Introduction ............................................................................. 69
3.2 Overview of Research Methods ............................................. 69
3.3 Our Research Methodology .................................................... 71
  3.3.1 Knowledge Modeling and Computerization ..................... 73
    3.3.1.1 CPGs Knowledge Representation ............................ 73
    3.3.1.2 CPGs Computerization into Segments .................... 73
  3.3.2 EBM-Literature Linking Strategy ..................................... 74
    3.3.2.1 Key-medical Terms and Query Type Selection .......... 74
    3.3.2.3 Linking CPG with EBM-Literature ....................... 76
  3.3.3 Medical Information Processing ..................................... 76
    3.3.3.3 Information-view for Customized Query ................ 77
  3.3.4 CPGs Content Retrieval .................................................. 78
    3.3.4.1 Computerized CPGs Indexing Strategy .................... 79
    3.3.4.2 Content and Knowledge Retrieval ....................... 79
  3.3.5 Clinical Knowledge Assistance ....................................... 79
    3.3.5.1 Clinical Knowledge Assistance (CKA) Info-Structure .... 80
  3.4 Summary ........................................................................... 80

VIII
## CHAPTER 4
**CLINICAL PRACTICE GUIDELINES’ KNOWLEDGE MODELING AND COMPUTERIZATION**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>81</td>
</tr>
<tr>
<td>4.2 An Overview of CPG Computerization Approaches</td>
<td>82</td>
</tr>
<tr>
<td>4.2.1 Knowledge-Base-Centric Formalism</td>
<td>82</td>
</tr>
<tr>
<td>4.2.2 Document-Centric Formalism</td>
<td>84</td>
</tr>
<tr>
<td>4.2.2.1 Translation of guideline knowledge</td>
<td>85</td>
</tr>
<tr>
<td>4.2.2.2 Workflow Integration</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Proposed Clinical Practice Guideline Computerization Framework</td>
<td>90</td>
</tr>
<tr>
<td>4.3.1 Functional Flow of Our CPG Computerization Framework</td>
<td>92</td>
</tr>
<tr>
<td>4.3.2 Encoding Strategy</td>
<td>94</td>
</tr>
<tr>
<td>4.3.3 Working Example of CPG Modeling Using Our Encoding Strategy</td>
<td>98</td>
</tr>
<tr>
<td>4.3.4 Extended-Knowledge Component Ontology</td>
<td>102</td>
</tr>
<tr>
<td>4.3.5 Extended-Knowledge Component Elements ID Scheme</td>
<td>109</td>
</tr>
<tr>
<td>4.3.6 Contextual Weight</td>
<td>111</td>
</tr>
<tr>
<td>4.3.7 UMLS-MMTX Manger</td>
<td>113</td>
</tr>
<tr>
<td>4.4 Implementation of the CPGs Computerization Framework</td>
<td>114</td>
</tr>
<tr>
<td>4.4.1 Medical Terms Extraction and Meta-Information retrieval Method</td>
<td>114</td>
</tr>
<tr>
<td>4.4.1.1 Experiments for Determining the UMLS Redundant Behavior</td>
<td>118</td>
</tr>
<tr>
<td>4.4.2 Medical Terms Filtration Algorithm</td>
<td>120</td>
</tr>
<tr>
<td>4.4.3 Algorithm for Medical Terms Objects Comparison</td>
<td>123</td>
</tr>
<tr>
<td>4.4.4 Engineering process of the Extended-Knowledge Component Instance Creator</td>
<td>126</td>
</tr>
<tr>
<td>4.5 Summary</td>
<td>132</td>
</tr>
</tbody>
</table>

## CHAPTER 5
**LINKING MEDICAL LITERATURE WITH COMPUTERIZED CPGs SEGMENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>134</td>
</tr>
<tr>
<td>5.2 An Overview of Clinical Query Models, Query Types and Related-Problems</td>
<td>135</td>
</tr>
<tr>
<td>5.3 CQGF: Context Specific Clinical Query Generation Framework</td>
<td>139</td>
</tr>
<tr>
<td>5.3.1 Context Specific Clinical Query Generation Technique</td>
<td>141</td>
</tr>
<tr>
<td>5.3.2 Functional Flow of CQGF Framework</td>
<td>145</td>
</tr>
<tr>
<td>5.3.2.1 Medical Terms Extractor</td>
<td>146</td>
</tr>
<tr>
<td>5.3.2.2 Redundancy Filter</td>
<td>146</td>
</tr>
<tr>
<td>5.3.2.3 Semantic Type (S.T) Filter</td>
<td>147</td>
</tr>
<tr>
<td>5.3.2.4 Info-Manager</td>
<td>148</td>
</tr>
<tr>
<td>5.3.2.5 Diseases Ontology Filter</td>
<td>149</td>
</tr>
<tr>
<td>5.3.2.6 Semantic Relation Score</td>
<td>151</td>
</tr>
<tr>
<td>5.3.2.7 Final query set Creator</td>
<td>152</td>
</tr>
<tr>
<td>5.3.2.8 Query Type Finder</td>
<td>153</td>
</tr>
<tr>
<td>5.3.2.9 PubMed Webservice Connector</td>
<td>155</td>
</tr>
<tr>
<td>5.4 Working Example of CQGF Framework</td>
<td>155</td>
</tr>
</tbody>
</table>
CHAPTER 6
MEDICAL INFORMATION PROCESSING AND QUERY FORMULATION 165

6.1 Introduction ................................................................. 165
6.2 An Overview of Medical Information processing ......................... 166
6.3 AMIPF: Automatic Medical Information Processing Framework ........ 169
   6.3.1 Functional Flow and Techniques of Automatic Medical Information Processing Framework .................. 170
      6.3.1.1 AMIPF Input Pre-processing .................................. 171
      6.3.1.2 Noun-phrase and Concept Recognition ...................... 172
      6.3.1.3 Classification of Noun-Phrases and Concepts ............... 176
         6.3.1.3.1 Meta-Info Extractor ...................................... 176
         6.3.1.3.2 Indexer .................................................... 176
         6.3.1.3.3 Critical Situation Sensor ............................... 176
         6.3.1.3.4 Redundancy Filter ...................................... 177
         6.3.1.3.5 Semantic Type Filer .................................... 177
         6.3.1.3.6 Negation Detector ...................................... 177
         6.3.1.3.7 Phrase and Concepts Classifier ......................... 180
      6.3.1.4 Information Mapping ......................................... 180
         6.3.1.4.1 Diseases Ontology ..................................... 181
         6.3.1.4.2 Info-Mapper ............................................. 181
      6.3.1.5 Query Formulation ........................................... 182
         6.3.1.5.1 Automatic Query Generation ............................ 182
         6.3.1.5.2 Information-View: Customized Query Formulation .... 185
   6.4 A Case Study: Working Example of the AMIPF ....................... 186
   6.5 Summary ............................................................... 192

CHAPTER 7
KNOWLEDGE RETRIEVAL FROM COMPUTERIZED CLINICAL GUIDELINES .... 194

7.1 Introduction.................................................................. 194
7.2 An Overview of Computerized Clinical Practice Guidelines Content and Knowledge Access at Point of Care............................... 195
7.3 Content and Knowledge Retrieval Framework ......................... 199
   7.3.1 Functional Flow of Content and Knowledge Retrieval Framework .... 201
      7.3.2 Index Manager and Indexing Strategy for Extended-Knowledge Components ........................................... 203
         7.3.2.1 Inverted-Ex-KC-Index .................................... 204
         7.3.2.2 Inverted-Path-Index .................................... 205
         7.3.2.3 Meta-Info-Index ........................................... 206
      7.3.3 Representation of Knowledge base for Extended-Knowledge Components and Query ................................................... 207
CHAPTER 8
EVALUATION AND DISCUSSIONS ..................................................... 226

8.1 Introduction ............................................................................. 226
8.2 Evaluation of Clinical Practice Guidelines Computerization Framework ........................................... 227
8.3 Evaluation of Context Specific Query Generation Framework (CQGF) ...................................... 231
8.4 Evaluation of Automatic Medical Information Processing Framework (AMIPF) .......................... 234
8.5 Evaluation of the Content and Knowledge Retrieval Framework .............................................. 246
8.6 Summary ............................................................................. 251

CHAPTER 9
CONCLUSION ............................................................................. 252

9.1 Preamble ............................................................................. 252
9.2 Revisiting the Contributions .................................................... 253
9.2.1 Revisiting the First Objective ............................................... 254
9.2.2 Revisiting the Second Objective ........................................... 254
9.2.3 Revisiting the Third Objective ............................................. 255
9.2.4 Revisiting the Fourth Objective .......................................... 255
9.2.5 Revisiting the Fifth Objective ............................................. 255
9.3 Future Directions ................................................................ 256

BIBLIOGRAPHY ............................................................................ 259

APPENDICES ............................................................................. 271
APPENDIX A: GEM-Encoded CPG and Extended-Knowledge Component ........................................... 272
APPENDIX B: Medical Concepts for Heart Disease Ontology ................................................................. 303
APPENDIX C: Semantic Type Filter, Semantic Relations, And Query Types Categories ............................................ 308
APPENDIX D: Associated-With Semantic Relations Between Semantic Types ........................................... 310
APPENDIX E: Medical Information Processing Supplements ..................................................................... 313
APPENDIX F: Clinical Knowledge Assistance System .............................................................................. 320
APPENDIX G: Samples of Medical Letters ............................................................................................ 327
LIST OF TABLES AND FIGURES

Table 4.1 Elements IDs of Ex-KC1 from ‘ACS_06_za.xml’ GEM-Encoded CPG .......................... 111
Table 4.2 Results from the medical Terms identification, extraction and meta-info retrieval processes ............................................................................................................ 118
Table 5.1 Potentially significant extracted medical phrases and medical terms .......................... 158
Table 5.2 Remaining medical phrases and terms after ‘Redundancy Filter’ ................................. 159
Table 5.3 Resulting filtered medical terms and phrases from Semantic Type Filter .......... 160
Table 5.4 Resulting medical terms/phrases from Diseases Ontology Filter .............................. 161
Table 5.5 Medical concepts with SMRscore value .................................................................... 161
Table 5.6 KC set medical concepts ......................................................................................... 161
Table 5.7 Medical articles retrieved from PubMed for “Reperfusion Therapy Ex-KC” .............. 162
Table 6.1 A table with medical concepts to be filtered out based on their semantic types .......................... 188
Medical concepts with light shaded cells are filtered out as their semantic types of not being significant .............................................................................................................. 188
Table 6.2 General classification of medical concepts into five categories. Medical .......................... 189
Concept in dark shaded cell represents negated medical concept ......................................... 189
Table 6.3 Medical concepts after ‘Info-Mapper’ processing ....................................................... 190
Table 6.4 Medical concepts with their ‘Final Weight’ and SMR ‘score’ ........................................ 190
Table 6.5 Medical concepts sorted by Final Weight and SMRsco...
Table B.1 Heart Disease Ontology Medical Concepts ................................................. 303
Table C.1 Semantic Types used for General Classification strategy ...................... 308
Table C.2 Semantic Types for query categories ...................................................... 309
Table E.1 Medical concepts extracted as potential candidates during response letter processing discussed in section 6.4 .................................................. 313
Table E.2 Results for General Classification for each letter where TP stands for true positive, FP for false positive, FN for false negative, Re for Recall, Pr for Precision, and Fm for F-measure .......................................................... 317
Table E.3 Results for Cardiology Classification for each letter where TP stands for true positive, FP for false positive, FN for false negative, Re for Recall, Pr for Precision, and Fm for F-measure ........................................... 318

Figure 1.1 General Practitioner interactions with CKA system .......................... 20
Figure 1.2 Specialist interactions with CKA system .............................................. 23
Figure 2.1 PROforma task ontology (Fox J, et al., 1998) ........................................ 29
Figure 2.2 GLARE ontological entities in the representation of clinical guidelines (Terenziani, et al., 2001) ................................................................. 32
Figure 2.3 GEM elements hierarchy (Yale School of Med, 2006) ....................... 37
Figure 2.4 Knowledge component element’s hierarchy in GEM II (Yale School of Med, 2006) ................................................................................ 38
Figure 2.5 Loading an XML summary tree into index structure (Kotsakis, 2002) .... 48
Figure 2.6 Search engine architecture (Abbaci, et al., 2006) .............................. 49
Figure 2.7 System architecture for context-sensitive keyword search (Botev and Shanmugasundaram, 2005) ......................................................... 50
Figure 2.8 Node Path Index (NPI)(Xin-Ye, et al., 2006) ....................................... 51
Figure 2.9 Inverted tag node index (Xin-Ye, et al., 2006) ...................................... 52
Figure 2.10 Inverted semantic keyword index (Xin-Ye, et al., 2006) .................... 52
Figure 3.1 Research Methodology, UMLS = Unified Medical Language System, CPG = Clinical Practice Guidelines, Ex-KC = Extended-Knowledge Component, EBM literature = Evidence-based medical literature ........................................ 75
Figure 4.1 CPGs Computerization Framework ...................................................... 93
Figure 4.2 Excerpt from Australian CPG for Management of acute coronary syndrome(Heart Foundation Australia, 2006) ............................................. 99
Figure 4.3 Excerpt of XML output for Knowledge Components mark up .......... 99
Figure 4.4 Marking up of Recommendation element ........................................ 100
Figure 4.5 Marking up of Conditional element .................................................. 100
Figure 4.6 Marking up of ‘Decision variable’, ‘Value’, ‘Decision Variable Description’ elements ................................................................. 101
Figure 4.7 Mark up output of ‘Action’, ‘Action Description’, ‘Recommendation Strength’, and ‘Reason’ elements .................................................. 101
Figure 4.8 Mark up of ‘Linkage’ element ......................................................... 102
Figure 4.9 ‘Extended-Knowledge Component Ontology’ elements structure and relationships .................................................................................. 104
Figure 8.3 AMIPF overall average F-measure for General classification indicated by thick straight line and average F-measure for General classification of each letter indicated by thin curve........................................240
Figure 8.4 Comparison of AMIPF average Recall, average Precision, and average F-measure for General classification.................................................................241
Figure 8.5 AMIPF overall average Recall for Cardiology classification indicated by thick straight line and average Recall for Cardiology classification of each letter indicated by thin curve........................................................................243
Figure 8.6 AMIPF overall average Precision for Cardiology classification indicated by thick straight line and average precision for Cardiology classification of each letter indicated by thin curve..................................................................................243
Figure 8.7 AMIPF overall average F-measure for Cardiology classification indicated by thick straight line and average F-measure for Cardiology classification of each letter indicated by thin curve..................................................................................244
Figure 8.8 Comparison of AMIPF average Recall, average Precision, and average F-measure for Cardiology classification .................................................................244
Figure 8.9 Results of Recall and Precision of retrieved Ex-KCs for each query.............249
Figure 8.10 Results of overall average Recall and Precision of CKR framework for retrieved Ex-KCs ........................................................................................................249
Figure 8.11 Results of relevancy in terms of average precision for ranked Ex-KCs retrieved for a set of queries........................................................................................................250
Figure F.1 Main Interface for CKA adminstration and navigation to its different functions ..........................................................321
Figure F.2 Showing GEM-encoded CPG to create Ex-KCs.............................................321
Figure F.3 Showing the indexing process of EX-KC ..........................................................................................................................322
Figure F.4 Showing the process of generating query from Ex-KCs...............................322
Figure F.5 Showing the generated query for Ex-KC corresponding retrieved medical articles from Pubmed ..........................................................................................323
Figure F.6 Showing the interface to select the letter that will be processed by the CKA ..............................................................................................................................323
Figure F.7 Showing the information-view of processed letter ....................................324
Figure F.8 Information-view showing general classification and cardiology classification .......................................................................................................................324
Figure F.9 Information-view showing potential query terms and interface to formulate customized query .........................................................................................................325
Figure F.10 Showing the retrieved CPGs segments for the query presented in Figure F.9 ..................................................................................................................................................325
Figure F.11 Showing an interface of submitting user query without processing a letter 326
Figure F.12 Showing detailed view of a retrieved CPG segment..................................326
Figure G.1 Medical letter 1 ...........................................................................................................................327
Figure G.2 Medical letter 2 ...........................................................................................................................328
Figure G.3 Medical letter 3 ...........................................................................................................................329
ABSTRACT

Healthcare is a complex domain and distributed in nature. Some of its major characteristics are shared and distributed decision making and management of care. Such characteristics require the communication of complex and diverse forms of information between clinical and other settings. The healthcare services center around the wellness, maintenance, and illness management of patients. The delivery of such services heavily depends on up-to-date medical knowledge and experience. However, miscommunication, misinterpretation of diverse forms of information exchanged between clinical settings, inability to be up-to-date with clinical best practices, and accessing rapidly changing and exponentially growing medical knowledge cause errors in healthcare that are leading cause of injuries and deaths.

In this thesis, we addressed the research problems related to the interpretation and understanding of medical referral and response letters, exchanged between Specialists and General Practitioners (GPs) for patients care decision making. These research problems are implicitly associated with GPs information needs at point of care, including management of and access to evidence-based medical information and knowledge relevant to medical referral and response letters. The interpretation and understanding of referral and response letters and sending alerts for critical situations at point of care require methods for medical information processing, and methodologies for modeling and management of clinical knowledge. It also requires mechanisms to access concise, relevant, evidence-based clinical knowledge and techniques for efficient and context-sensitive retrieval. The goal of the thesis has been divided into five objectives, therefore,
five-phased multi-steps research methodology was devised. We have taken interdisciplinary solution approach along the lines of healthcare knowledge management, contextual information retrieval, and knowledge-based search strategies. We have formulated a knowledge modeling methodology and a computerization technique for clinical practice guidelines, which transform them into computer interpretable segments that are enriched with content-specific meta-information. To link segments of computerized clinical practice guidelines with online evidence-based medical knowledge, we have developed a technique for automatically generating clinical queries from these segments. These queries are used by our Context Specific Query Generation framework to retrieve relevant medical literature from online evidence-based knowledge sources and to link them with corresponding computerized CPGs segments.

We have developed a method for computerized processing of referral and response letters. This method analyzes medical information and provides a comprehensive information-view of the letter to help healthcare practitioners formulate customized and focused information specifications to access required knowledge. We also have developed a technique for contextual and statistical analysis of medical concepts and indexing strategy, which are used to retrieve CPGs segments and related medical literature relevant to information needs of healthcare practitioners. Finally, we have designed a healthcare-knowledge mediated architecture and implemented a computer system for Clinical Knowledge Assistance (CKA) to provide better interpretation and understanding of referral and response letters.
<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Phrases and Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC.MedTerm</td>
<td>Action Medical Term</td>
</tr>
<tr>
<td>AC.text</td>
<td>Action Element Text Value</td>
</tr>
<tr>
<td>AQG</td>
<td>Automatic Query Generation</td>
</tr>
<tr>
<td>ASTM</td>
<td>American Society For Testing And Materials</td>
</tr>
<tr>
<td>BNF</td>
<td>Backus-Naur Form</td>
</tr>
<tr>
<td>C.C</td>
<td>Cardiology Classification</td>
</tr>
<tr>
<td>C.Dig</td>
<td>Cardiology Classification Of Diagnosis</td>
</tr>
<tr>
<td>C.Dis</td>
<td>Cardiology Classification Of Diseases</td>
</tr>
<tr>
<td>C.Med</td>
<td>Cardiology Classification Of Medications</td>
</tr>
<tr>
<td>C.MedTerm</td>
<td>Conditional Medical Term</td>
</tr>
<tr>
<td>C.Sym</td>
<td>Cardiology Classification Of Symptoms</td>
</tr>
<tr>
<td>C.The</td>
<td>Cardiology Classification Of Therapies</td>
</tr>
<tr>
<td>C-CPGs</td>
<td>Computerized Clinical Practice Guidelines</td>
</tr>
<tr>
<td>CKA</td>
<td>Clinical Knowledge Assistance</td>
</tr>
<tr>
<td>CKR framework</td>
<td>Content And Knowledge Retrieval Framework</td>
</tr>
<tr>
<td>CPGs</td>
<td>Clinical Practice Guidelines</td>
</tr>
<tr>
<td>CTH</td>
<td>Common Terms Threshold</td>
</tr>
<tr>
<td>CW</td>
<td>Contextual Weight</td>
</tr>
<tr>
<td>DeGeL</td>
<td>Digital Electronic Guideline Library</td>
</tr>
<tr>
<td>DV</td>
<td>Decision Variable</td>
</tr>
<tr>
<td>DV.MedTerm</td>
<td>Decision Variable Medical Term</td>
</tr>
<tr>
<td>Dvalue.MedTerm</td>
<td>Decision Variable Value Medical Term</td>
</tr>
<tr>
<td>EBM</td>
<td>Evidence-Based Medicine</td>
</tr>
<tr>
<td>EFS</td>
<td>Earliest Finishing Shift</td>
</tr>
<tr>
<td>ESS</td>
<td>Earliest Starting Shift</td>
</tr>
<tr>
<td>Ex-KC-O</td>
<td>Extended-Knowledge Component Ontology</td>
</tr>
<tr>
<td>Ex-KCs</td>
<td>Extended-Knowledge Components</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>G.C</td>
<td>General Classification</td>
</tr>
<tr>
<td>G.Dig</td>
<td>General Classification Of Diagnosis</td>
</tr>
<tr>
<td>G.Dis</td>
<td>General Classification Of Diseases</td>
</tr>
<tr>
<td>G.Med</td>
<td>General Classification Of Medications</td>
</tr>
<tr>
<td>G.Sym</td>
<td>General Classification Of Symptoms</td>
</tr>
<tr>
<td>G.The</td>
<td>General Classification Of Therapies</td>
</tr>
<tr>
<td>GEM</td>
<td>Guideline Element Model</td>
</tr>
<tr>
<td>GLARE</td>
<td>Guideline Acquisition, Representation, And Execution</td>
</tr>
<tr>
<td>GLIF</td>
<td>Guideline Interchange Format</td>
</tr>
<tr>
<td>HGML</td>
<td>Hypertext Guideline Markup Language</td>
</tr>
<tr>
<td>HKB</td>
<td>Healthcare-Knowledge Base</td>
</tr>
<tr>
<td>HKM</td>
<td>Healthcare Knowledge Management</td>
</tr>
<tr>
<td>IM.MedTerm</td>
<td>Imperative Medical Term</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>KC</td>
<td>Knowledge Component</td>
</tr>
<tr>
<td>KC.MedTerm</td>
<td>Knowledge Component Medical Term'</td>
</tr>
<tr>
<td>KC-Extractor</td>
<td>Knowledge-Component-Extractor</td>
</tr>
<tr>
<td>LFS</td>
<td>Latest Finishing Shift</td>
</tr>
<tr>
<td>LSS</td>
<td>Latest Starting Shift</td>
</tr>
<tr>
<td>M-Cand</td>
<td>Medical Concepts</td>
</tr>
<tr>
<td>MEDLINE</td>
<td>Medical Literature Analysis and Retrieval System Online</td>
</tr>
<tr>
<td>MeSH</td>
<td>Medical Subject Headings</td>
</tr>
<tr>
<td>MLM</td>
<td>Medical Logic Modules</td>
</tr>
<tr>
<td>ON</td>
<td>Original Name</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Format</td>
</tr>
<tr>
<td>Rec</td>
<td>Recommendation</td>
</tr>
<tr>
<td>Rec.MedTerm</td>
<td>Recommendation Medical Term</td>
</tr>
<tr>
<td>RP</td>
<td>Research Problem</td>
</tr>
<tr>
<td>SAGE</td>
<td>Standards-Based Sharable Active Guideline Execution</td>
</tr>
<tr>
<td>Scor</td>
<td>UMLS Score</td>
</tr>
<tr>
<td>SMR score</td>
<td>Semantic Relation Score</td>
</tr>
<tr>
<td>ST</td>
<td>Semantic Type</td>
</tr>
<tr>
<td>TCW</td>
<td>Total Contextual Weight</td>
</tr>
<tr>
<td>THFQS</td>
<td>Threshold For A Final query set</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TW</td>
<td>Total Weight</td>
</tr>
<tr>
<td>TW set</td>
<td>Total Weight Set</td>
</tr>
<tr>
<td>VS</td>
<td>Vocabulary Source</td>
</tr>
<tr>
<td>XHTML</td>
<td>Extensible Hypertext Markup Language</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Mark Up Language</td>
</tr>
</tbody>
</table>
LIST OF PUBLICATIONS


CHAPTER 1
INTRODUCTION

1.1 Preamble

The work presented in this thesis falls under the domain of health informatics. This thesis addresses the research problems associated with the representation, management, and delivery of evidence-based clinical knowledge for better interpretation, and understanding of patient-centered medical documents. It may improve decision-making for patients at point of care. The techniques, methods, and strategies developed in this thesis take approaches from healthcare knowledge management, contextual information retrieval, and knowledge-based search strategies. The experiment domain of this thesis is cardiology. In this chapter, we present first a general overview of the problem domain in section 1.2; proceeding to the section 1.3, we define the motivation, aims, and objectives of the thesis; in section 1.4, we describe the research problems and our solution approaches; in section 1.5, we state the contributions of this thesis, towards the end of this chapter, we present thesis outline.

1.2 An Overview of Problem Domain

Healthcare is regarded as “a knowledge and experience dependent” and people-centered enterprise (Abidi and Yu-N, 1999). The delivery of healthcare service involves an active interplay between medical knowledge and experience whereby business of healthcare addresses both the wellness, maintenance and illness management of
individuals (Abidi and Yu-N, 1999). It is a complex and knowledge-intensive domain that is applied in everyday patient treatment. In healthcare, nature of medical decision-making is quite complex. Clinicians, not only, must know the facts about the problems e.g. diseases, their frequencies, signs, symptoms but also must know how all these factors are affected by a patient characteristics e.g. age, sex, family history, risk factor, and other diseases, etc (Alime, 2007; Eddv, 1986). Furthermore, the number of medical facts, information and correlation among these facts that a healthcare practitioner has to take into account are overwhelming (Alime, 2007).

Moreover, medical knowledge is not only changing rapidly rather growing exponentially (Ammon, et al., 2009). An inability to be up-to-date with changing and exponentially growing medical information causes errors in healthcare that are the leading cause of injury and death (Katharina, 2005). Kohn et al. (2000) stated in their study that one of the important sources of leading cause of death is lack of knowledge of current best clinical practices.

Such a knowledge gap has demanded the paradigm shift to the practice of “Evidence-Based-Medicine” (EBM) (Parry, 2005). Evidence-Based-Medicine is defined as “the integration of best research evidence with clinical expertise and patient value” (Sackett, et al., 1996).

However, healthcare practitioners face many hurdles in seeking evidences and relevant information at point of care. Some of the hurdles are time constraint, inexperience and inability to formulate the right and focused search query that can result in overwhelming number of documents (Gorman and Helfand, 1995; Jimmy and Dina, 2007; Osheroff, et al., 1991). In addition, factors like unawareness of online literature
search facilities and subject specific filters for locating relevant articles, expanding medical literature, etc add extra pressure and cognitive load (Abidi, et al., 2005 a; Covell, et al., 1985; Gorman and Helfand, 1995; Jimmy and Dina, 2007; Osheroff, et al., 1991; Wanda and Henry, 2000).

The challenges and issues, pertaining to the delivery of evidence-based medicine care and keeping clinicians up-to-date with new advances and evidences in medical practice require interdisciplinary approaches based on healthcare knowledge management, contextual information retrieval, and knowledge-based search strategies (Abidi, 2008; Jackson, 2000; Jeremy, 2001; Shepherd, 2007). Such interdisciplinary approaches are necessary in providing (Abidi, 2008; Jackson, 2000; Jeremy, 2001; Shepherd, 2007):

- focused and contextually-sensitive clinical knowledge assistance,
- assistance in formulating clinical queries and search strategies for accessing clinical knowledge,
- better interpretation and understanding of patients’ medical documents,
- knowledge-medicated solutions and their integration in institutional workflow,
- improvement of quality, efficiency and efficacy of health delivery systems.

Healthcare knowledge management (HKM) can be defined as “systematic creation, modeling, sharing, operationalization and translation of healthcare knowledge to improve the quality of patient care” (Abidi, 2008). The aim of HKM is to provide relevant, optimal, effective, and concise healthcare knowledge to healthcare professionals for their
information needs to assist them in making well-informed, evidence-based, cost-effective patient care decisions (Abidi, 2008; Shepherd, 2007).

Healthcare knowledge management addresses the problem issues pertaining to healthcare stakeholders through: (a) technical framework that deals with knowledge management strategies, knowledge representation, knowledge processing methods to develop and deploy knowledge-driven solutions and (b) operational info-structure that deals with operational issues, strategies, information workflows to help incorporate knowledge management solutions in clinical settings (Abidi, 2008).

Information retrieval (IR) deals with structuring, organizing, analyzing, searching, and retrieving of data, information, and knowledge (Salton, 1968; Salton and McGill, 1983). Contextual information retrieval makes use of search technologies, knowledge-based search strategies, context-sensitivity, and meta-information of content into a single framework in order to provide relevant and most optimal answer to user’s information needs (Bose, 2003; Shepherd, 2007).

1.3 Motivation, Aim and Objectives

The motivation of the work presented in this thesis emerged from the research problems, in a clinical setting, related to the interpretation and understanding, of medical referral and response letters, exchanged between ‘Specialists’ and ‘General Practitioners’ (GP), and sending alert messages for critical situations that can help improve decision-making at point care. These research problems are implicitly associated with GPs information needs at point of care, including management of and access to evidence-based medical information and knowledge relevant to medical referral and response letters. We take interdisciplinary solution approach along the lines of healthcare
knowledge management, contextual information retrieval and knowledge-based search strategies to address the research problems.

In general, the aim of this research is to develop methods and techniques that support the management and the delivery of evidence-based clinical information and knowledge, relevant to patient-medical documents, for better interpretation, understanding and decision-making at point of care. These methods and techniques can be used to develop a computer system such as clinical decision support system, clinical information systems etc.

Specifically, this thesis addresses the designing and development of methods and techniques for a computer system that manages heterogeneous clinical knowledge, provides evidence-based information and knowledge to assist healthcare practitioners for better interpretation and understanding of clinical referral and response letters, and sending alert messages for critical situations at point of care.

The goal of this thesis has been divided into five-objectives:

(i) To devise easy and effective method for knowledge modeling and representation for CPGs computerization into concise yet focused segments, which incorporates context, semantics, and other meta-information related to CPGs content. Such method should help healthcare professional in participating the CPGs computerization process without having technical knowledge.

(ii) To develop a technique that would link relevant evidence-based medical literature with computerized CPGs. The technique would generate, automatically, focused search queries by exploiting the syntactic, contextual, semantic, and meta-
information of computerized CPGs segments. These queries would be used to retrieve medical literature through online evidence-based knowledge source(s), which will be linked to corresponding CPGs segments.

(iii) To develop a method that processes and analyses medical information from referral and response letters to help healthcare practitioners in formulating customized and focused information specifications and a technique that provides alerts for critical medical conditions in the letters.

(iv) To devise a technique that retrieves concise and relevant CPGs segments and corresponding linked medical literature to healthcare professionals’ information specification. It will exploit context, semantics, and other meta-information related to CPGs content for contextually relevant access to CPGs segments and related literature at point of care.

(v) To develop a computer system for Clinical Knowledge Assistance (CKA) that provides evidence-based clinical knowledge and information to assist healthcare practitioners for better interpretation and understanding of clinical referral letters that would improve decision making at point of care.

In the next section we frame the research problems and issues to be addressed in this thesis.

1.4 Research Focus

1.4.1 Problem Statement and Solution Approach

In healthcare setting, patient care depends, to great extent, on adequate and timely information and knowledge exchange between treating doctors in general and chronic
Medical referral and response letters are standard and typically the common means of communicating information about patient care in particular (Tattersall, et al., 2002). The content of such letters varies from straight forward technical problem to complex cases in which extensive details need to be communicated in both directions (John, et al., 1992; Westerman, et al., 1990).

Several studies have shown that there is dissatisfaction at the quality of communication between specialist and general practitioners, because, quality and content of such letters is often poor and unfocused that it does not meet the information needs of letter receipting (Allan 1997; Campbell, et al., 2004; Jenkin, 1993; Jiwa, et al., 2005; Jiwa, et al., 2002; John, et al., 1992; Kentish, et al., 1987; Linne and S, 2000; Syed and Large, 2003; Tattersall, et al., 2002).

Allan (1997) stated in their work that ambiguity present in medical referral letters makes it difficult for hospital practitioners to make accurate decisions on patient management. In view of that, the role of medical referral and response letters is very critical in patient care. Misinterpretation during information exchange due to unfocused content of referral letter may cause serious errors and problems for patient care. Pringle (1991) stated that the referral letter is “the most underexploited method to influence consultant attitudes” and the reply letter is “the most neglected route of GP education”.

In consultation and discussion with ‘Specialists’ (from cardiology) and ‘General Practitioners’ (GPs), they pointed out the problem that persist during the communication and information exchange through referral and response letters. The problem is the interpretation and understanding of patient-centered medical referral and response letters.
at point of care. Often GPs require relevant knowledge and information at point of care to better interpret and understand the letters. Specialists and GPs are in consensus to exploit evidence-based healthcare knowledge from ‘Clinical practice guidelines’ and other form of medical documents like ‘Systematic review’, ‘Randomized controlled trial, ‘Case-control’ etc. Substantial evidence exists that conforming to clinical practice guidelines improves the quality of medical care (Grimshaw and Russel, 1993; Micielo, et al., 2002; Qualigini, et al., 2004). Furthermore, in referral and response letters, some medical conditions and situations seem to be of normal significance but, in reality, these happen to be critical medial conditions. Specialists are interested in providing alerts for such critical medical situations that need attention by GPs.

There are number of research challenges that need to be addressed, in providing relevant and timely medical knowledge and information through evidence-based practices for better interpretation and understanding of referral and response letters. Following, we highlight some of such challenges.

The clinical practice guidelines (CPGs) are one of the important sources of evidence-based medicine. Clinical practice guidelines are “systematically developed statements to assist practitioners and patient decisions about appropriate healthcare for specific circumstances” (Field and Lohr, 1990). In essence, these CPGs are textual documents that consist of many pages. “Referring to such text-based documents is impractical in today’s busy clinical environment that demands fast decision making by practitioners” (Shapoor, 2007). It has been pointed that due to time constraint, healthcare practitioners would like to have access to concise, yet relevant, small segments from clinical practice guidelines for the problem at hand (Hashmi, et al., 2009 b). Retrieving
and providing access to the concise yet relevant content of CPGs at point of care is a challenging research problem.

Another good source of evidence-based knowledge is online external evidence-based medical literature such as systematic review, randomized controlled trial, case-control, cohort, etc. Healthcare practitioners also tend to search online external evidence-based medical literature to acquire insights into diversity of opinions with the best external evidence to support their decisions, to clarify the ambiguities present in CPGs, to supplement their understanding of CPG content, etc (Abidi, et al., 2005a; Jimmy and Dina, 2007). MEDLINE is one of the known sources of evidence-based medical literature that contains millions of medical documents. Finding relevant medical information from such an important and big knowledge source is a challenging task, in particular finding medical literature related to CPGs content, patients’ specific documents (referral and response letters) etc (Abidi, et al., 2005a; Covell, et al., 1985; Gorman and Helfand, 1995; Jimmy and Dina, 2007; Osheroff, et al., 1991; Timpka, et al., 1989; Wanda and Henry, 2000).

In searching and finding relevant medical knowledge and information, healthcare practitioners’ information needs are normally represented in search queries that are based on accurate search terms (Keywords, keyphrases etc). Olena (2005) reported in their research that formulating focused and accurate search terms is one of the key problems in online searching of medical literature. Additionally, identification of keyphrases and key medical concepts from patient medical reports, discharge summary or other forms of patient’s medical documents such as patient referral and response letters adds to the problem of representing clinicians’ information needs. Moreover, patient specific referral
and response letters are free text medical documents. To help healthcare practitioners in formulating better and focused clinicians’ information needs from such letters requires computerized medical information processing method. This should help formulate customized information specification by providing a comprehensive information-view of significant medical information based on their significance and semantics in the letters.

Based on the above problem description, we have identified the research problems and framed them into research questions. Following, we present research problems and corresponding solution approaches to address them.

(RP: 1) **Knowledge modeling and computerization of Clinical Guidelines**: Clinical practice guidelines are long textual documents. In order to make effective use of CPGs at point of care, CPGs need to be transformed into computer-interpretable format (Sonnenberg and Hagerty, 2006). To access, concise and focused segments from CPGs, pertaining to the problem at hand, demands new knowledge modeling and computerization techniques. CPGs computerization approaches normally use decision-logics methods to execute CPGs based on run-time execution engine. These methods are divided into two types (a) event-based approach, and (b) rule-based approach (David and Antonio, 2008). The basic requirement for such approaches is the integration of patient data via EMR or mechanism of accessing of patient data for triggering rules or events. The problem of finding the best possible representation of an EMR for patient data has not been solved yet (David and Antonio, 2008). Realistically, in most countries (including technologically advanced) healthcare is not yet fully computerized, So, it is unrealistic (or hard and expensive) to include automated guidelines enactment systems in real clinical settings (David and Antonio, 2008). Additionaly, in those clinical settings
where input would not be from EMR or EHR, there is a need to find new methods to computerize CPGs.

The research question for this research problem is defined as:

- How to model and represent clinical practice guidelines knowledge in order to computerize them into concise yet focused segments and enrich them with meta-information?

(Solution Approach 1) We address this research question by developing a CPGs computerization framework. In this framework, we develop a knowledge modeling technique that is based on a document-centric model of clinical practice guidelines known as Guideline Element Model (GEM). The knowledge modeling technique depends on our Encoding strategy. This Encoding strategy helps involve healthcare practitioners in the modeling and computerization of CPGs without the dependency on knowledge engineer or software engineer. We develop an “Extended-Knowledge Component Ontology” that provides a mechanism to structure and represent, CPGs content, standardized medical terms, and additional meta-information related to CPGs content (such as contextual impact factor, semantics, UMLS score, etc). The CPGs computerization framework deploys our techniques to process CPGs content, add related meta-information, and transform CPGs into small computerized segments called “Extended-Knowledge Components” (Ex-KCs). These Extended-Knowledge Components are created automatically and are computer-interpretable.

(RP: 2) Linking evidence-based medical knowledge with CPGs: The second research problem of this thesis is related to the automatic linking of computerized clinical practice guidelines with relevant online evidence-based medical literature. Clinicians
normally refer to external medical evidences to have insights into past clinical trials, diversity of opinions, systematic reviews, randomized control trials, current clinical evidences pertaining to some recommendations, treatment procedures etc specified in clinical practice guidelines (Chambliss and Conley, 1996; Covell, et al., 1985; Ely, et al., 2005; Gorman, et al., 1994; Jimmy and Dina, 2007). However, finding such medical evidence is non-trivial task. Some of the main attributes of this task are related to its time-consuming and expensive nature (Chambliss and Conley, 1996). Furthermore, as described earlier, there are many factors that add complexities and problems in seeking evidence at decision points (Covell, et al., 1985; Gorman and Helfand, 1995; Wanda and Henry, 2000).

Healthcare practitioners need efficient ways to fulfill their information needs and clinical inquires at point of care (Chambliss and Conley, 1996; Ely, et al., 2005; Jimmy and Dina, 2007). Linking computerized clinical practice guidelines (C-CPGs) with relevant and current evidence-based medical literature is relatively unexplored area (Abidi, et al., 2005 a). The premise for this is that healthcare practitioners using clinical guidelines or computerized clinical guidelines supplement their expertise with the best external evidence to validate their understating and to support their clinical decisions (Abidi, et al., 2005 a; Chambliss and Conley, 1996; Chueh and Barnett, 1997; Sackett, et al., 1996).

Automatically linking current best evidences to concise segments of computerized clinical practice guidelines presents one of the interesting approaches for providing healthcare practitioners relevant current medical evidences. This imposes certain research
issues that need to be answered. The research question for this research problem is defined as:

- How to link relevant current medical evidences with corresponding segments of computerized clinical practice guidelines?

(Solution Approach 2) The research question related to second problem is addressed by developing a method that generates focused and context-sensitive queries from computerized CPGs segments and determines their query types. These queries are submitted to online evidence-based clinical literature repository “MEDLINE” to retrieve relevant documents. These documents are linked with the corresponding segments of computerized CPGs. In generating clinical query automatically, we exploit the information from computer-interpretable segments of CPGs.

A series of filtration process are devised based on semantics, semantic types and redundancy of medical concepts. Contextual importance and relation among medical concepts are quantified to assign weights. Diseases ontology is developed to remove contextual noise, to specify the meaning and relation of the medical concepts in computerized CPGs. Finally, strategy to select final medical terms and their query type is crafted by using the information and weights assigned to those terms. To perform these tasks our method and techniques constitute a framework i.e. called as “Context-Specific Query Generation Framework” (CQGF).

(RP: 3) Medical information processing for information analysis, alerts, and customized query: Formulating correct and accurate key-terms, is the main problem in electronic search (Olena, 2005). Furthermore, poorly formulated and unfocused
queries have been identified as one of the hurdles in accessing relevant medical information and knowledge (Ely, et al., 2005).

Medical referral and response letters in narrative textual form contain important and critical medical information related to patient’s medical problems. Computerized processing of medical information from referral and response letters has important implications in generating potential query, in formulating customized query to represent information needs of healthcare practitioners, and in delivering automatic alerts for critical medical situations.

In computerized processing of referral and response letters, automatically analyzing and extracting significant information is a challenging task. This includes research issues such as, finding significant terms, finding semantic relation among the terms, classifying them according to their semantics, standardizing medical terms, specifying the meaning of such significant medical terms to be used as information, etc.

The research question for this research problem is defined as:

- How to enable computerized processing and analysis of medical referral and response letters to provide comprehensive information-view for customized query formulation?

(Solution Approach 3) Our solution approach for third problem is as follows: We design a method that incorporate existing natural language processing algorithms to extract medical information (terms, phrases). We use UMLS Metathesaurus to standardize medical information and to retrieve their meta-information e.g. semantic types etc. We devise a filtration strategy to filter out insignificant terms. We also develop
a technique for critical situation detection. Negation detection algorithm is deployed and terms classification technique is developed to generate potential query terms and information-view of the information in the letter that helps formulate customized query. Techniques and methods for our approach are defined and combined in a framework i.e. called as “Automatic Medical Information Processing Framework” (AMIPF).

**RP: 4)** Content and knowledge retrieval from computerized clinical guidelines: Clinical guidelines are normally transformed into computer interpretable format based on a model representation or ontology (e.g. asbru, GEM etc). This model or ontology is referred as contextual model for CPGs (Moskovitch, et al., 2007). Based on a contextual model CPGs are structured into labeled semantic tags, each representing a context. Additionally, computerized clinical guidelines contain semantically meaningful segments and internal structure. Such segments and structure expose a sense of context (Hashmi, et al., 2009b; Robert and Yuval, 2009). Such segments are highly useful for contextually-relevant accessibility of CPGs content and knowledge at point of care. To provide access to the concise, contextually-relevant content of CPGs, it is required to develop a technique for computerized clinical practices guidelines (C-CPGs) content and knowledge retrieval. This technique would exploit above defined information and would perform statistical analysis of medical content within C-CPGs.

The research questions for this research problem are defined as:

- What approaches from knowledge management and information retrieval could be used to develop a ‘concise content retrieval’ technique that exploits the information provided by the computerized clinical guidelines?
• How to develop information structure and indexing strategy to support effective retrieval of the semi-structured medical documents (in our case C-CPGs)?

**Solution Approach 4** The research questions framed for fourth research problem are addressed as follows. We develop a CPGs content retrieval technique that exploits contextual information, information obtained by the semi-structured representation of C-CPGs (Extended-Knowledge Components), and meta-information of CPGs content. We devise indexing strategy to obtain high Recall and high Precision, which is based on a principal defined by (Salton and McGill, 1983). This states that in order to obtain high Recall, an exhaustive indexing is required and to ensure high Precision, highly specific terms should be used. Additionally, terms should carry additional content indication. We develop weighting and ranking algorithms by utilizing contextual information and other meta-information of C-CPGs content. These techniques, method along with some other algorithms like “Ex-KCs” clustering, context analyzing, etc are developed and incorporated in a framework to retrieve concise, yet relevant CPGs’ segments. This framework is called as “Content and Knowledge Retrieval Framework” (CKR).

**RP: 5) Designing and development of a clinical knowledge assistance infostructure**: The fifth research problem of this thesis is related with the prototyping of a clinical knowledge assistance info-structure. The clinical knowledge assistance info-structure would be a computer system that will assist in better understanding and interpretation of referral and response letters by providing access to relevant evidence-based knowledge. It will deploy the techniques and methods developed in this thesis for
the delivery of evidence-based knowledge and its management. The associated research problem has been framed in the following research question:

- How to design and develop a clinical knowledge assistance infostructure that integrates techniques and methods for evidence-based knowledge delivery and defines the information workflow, coordination and communication between related modules?

(Solution Approach 5) The research question for the fifth problem is addressed as follows. We design a “Healthcare-Knowledge” base mediated system architecture for the development of clinical knowledge assistance system. The “Healthcare-Knowledge” base (HKB) consists of Ex-KCs knowledge base, evidence-based medical literature source(s), and different information indices. The HKB provides access paths to evidence-based knowledge sources. This architecture integrates the modules of the frameworks and defines information-structure of the HKB while keeping the independent behavior of each framework.

It provides a mechanism to populate HKB only from those sources that need to be accessed for relevant content. It ensures inter-frameworks communication for collaboration to traverse HKB for information and knowledge access. The clinical knowledge assistance based on our system architecture facilitates a mechanism to help formulate healthcare practitioners’ information needs related to medical referral letters as information specification—akin to query. It interacts with HKB to provide access of evidence-based information and knowledge at point of care.
1.5 Contributions

This thesis makes contributions in the multidisciplinary field of health informatics. The work is focused on medical knowledge modeling approaches, context-sensitive knowledge retrieval techniques and knowledge-based search strategies in healthcare. The thesis makes the following contributions:

(a) A novel knowledge modeling method for CPGs computerization that transforms CPGs into concise yet focused segments called Extended-Knowledge Components. The method consists of an innovative Encoding strategy, a distinctive Extended-Knowledge Component Ontology (Ex-KC-O), and an algorithm that creates Ex-KCs and enriches them with additional meta-information. The Encoding strategy provides effective modeling of CPGs. It is specifically designed to be simple in order to be used without the dependency on a knowledge engineer or software engineer.

(b) An innovative technique for automatic generation of context-sensitive clinical queries from the segments of computerized CPGs. This technique is used in our Context-specific Query Generation Framework that uses these queries to retrieve online evidence-based medical literature and link them with corresponding segments of computerized CPGs.

(c) A distinctive method to enable computerized processing and analysis of medical referral and response letters to help formulate customized query for accessing evidence-based clinical practices. Our method recognizes semantic categories of medical terms and their relationships, classifies significant medical terms, identifies negated terms, and senses critical medical situations for sending alerts. Using this information, an
organized and comprehensive information-view of patient’s medical problems is created to help formulate customized and focused information specifications.

(d) A method for the retrieval of contextually relevant knowledge from computerized CPGs and corresponding evidence-based literature. This method incorporates our novel technique for contextual and statistical analysis of medical concepts and an innovative indexing strategy. This method exploits the information provided by, statistical analysis, semi-structured representation of medical documents, semantic tags, contextual impact, and standardized representation of medical concepts.

(e) A *healthcare-knowledge* base mediated architecture and a computer system for **Clinical Knowledge Assistance (CKA)**. This brings together the salient features and functions developed in the frameworks for CPGs computerization, linking online evidence-based literature, computerized analysis of medical letters, and contextual retrieval of evidence-based knowledge in a unified platform. The CKA manages evidence-based knowledge, provides knowledge assistance for better interpretation and understating of clinical referral and response letters at point of care. This can improve decision making for patients care.

### 1.6 Clinicians’ Behavioral Interaction with Clinical Knowledge Assistance System

This section presents a high level view of General Practitioner (GP) and Specialist interaction with computerized system for Clinical Knowledge Assistance (CKA). Use cases are used to elaborate the functional interaction with a system in Figure 1.1 and Figure 1.2. First, GP functional interaction use cases are discussed followed by Specialist Functional interaction use cases.
1.6.1 General Practitioner Interaction Use Cases

GP functional interaction uses cases are shown in Figure 1.1. Each GP interaction related use case has number associated with it. We will discuss GP interaction use cases with reference to the number associated with it.
1— Upload the response letter

*Precondition: None*

• GP receives response letter from specialist and uploads it in the system to view it.

2— View automatically processed information from letter

*Precondition: Letter has to be uploaded*

• Letter is processed automatically, once uploaded, and a comprehensive information-view of processed medical information is presented to GP.

• Medical information processing framework is used to process medical information and create information-view

3— Use automatic created query

*Precondition: Letter has been processed*

• GP may use automatically created query terms to find relevant patient-specific information without writing any query.

• Medical information processing framework is used to process medical information and create automatic query.

4— Create query using processed information

*Precondition: Letter has been processed*

• GP uses processed information (information-view) and response letter to create his/her own query.
• Query is used to find the medical information or knowledge to validate specialist recommendation or to reduce confusion upon some treatments or to understand certain medical procedure and relevant evidences.

5— Retrieve and view relevant medical evidences/recommendations

_Precondition: Query has been submitted_

• System retrieves evidence-based knowledge components from clinical guidelines.

• These knowledge components are segments of clinical guidelines, which are relevant to the submitted query.

• These segments are linked to the online evidence-based medical literature (articles), which are also retrieved.

• GP is presented with relevant information and knowledge in a specific presentation scheme that gives holistic view of retrieved knowledge.

• GP can select any knowledge component based on its summary and see the details.

• GP can also see the relevant articles and other related knowledge components.

• This particular use case interaction uses knowledge retrieval framework.

• Knowledge retrieval framework uses Healthcare Knowledge Base.

• Healthcare Knowledge Base uses medical knowledge modeling framework and medical knowledge linking framework.
1.6.2 Specialist Interaction Use Cases

Specialist functional interaction uses cases are shown in Figure 1.2. We will discuss Specialist interaction use cases with reference to the number associated with them.

![Figure 1.2 Specialist interactions with CKA system](image)

**1— Upload the referral letter**

*Precondition: None*

- Specialist may and may not upload referral letter into system.
2—Create query to find or validate medical evidences

Precondition: None

• Specialist creates query to validate some medical evidences or to find more information about the treatment/test/recommendation to be prescribed.

3—Retrieve and view relevant medical evidences/recommendations

Precondition: Query has been submitted

• System retrieves evidence-based knowledge components from clinical guidelines.

• These knowledge components are segments of clinical guidelines, which are relevant to the submitted query.

• These segments are linked to the online evidence-based medical literature (articles), which are also retrieved.

• Specialist is presented with relevant information and knowledge in a specific presentation scheme that gives holistic view of retrieved knowledge.

• Specialist can select any knowledge component based on its summary and see the details.

• Specialist can also see the relevant articles and other related knowledge components.

• This particular use case interaction uses knowledge retrieval framework.

• Knowledge retrieval framework uses Healthcare Knowledge Base.

• Healthcare Knowledge Base uses medical knowledge modeling framework and medical knowledge linking framework.
1.7 Thesis Outline

Chapter 2: This chapter presents a literature review on formalisms for representing clinical practice guidelines, methods for accessing computerized guidelines content, search strategies for semi-structured documents, clinical query models and methods for the delivery of medical literature, techniques for automatic processing of medical information, and describes the knowledge resources and tools used in this thesis.

Chapter 3: This chapter outlines our research methodology. The research methodology is divided into five phases to achieve the objectives.

Chapter 4: This chapter describes our Knowledge modeling methodology. Encoding strategy and CPGs computerization framework. It also presents a working example of the CPGs computerization process and provides the engineering techniques for the implementation of the framework.

Chapter 5: This chapter presents our technique for automatic query generation from computerized CPGs segments. It provides in details the functionalities of Context Specific Query Generation framework (CQGF). It includes a working example for automatically generating query and linking online medical literature to corresponding CPG segment.

Chapter 6: This chapter explains our method and techniques developed for the Automatic Medical Information Processing framework (AMIPF). It presents a working example to illustrate the functional workflow of AMIPF.

Chapter 7: This chapter describes the indexing strategies for Content and Knowledge Retrieval (CKR) framework. It explains the technique for contextual and
statistical analysis. It provides in details the functionalities and functional flow of CKR framework. It presents the architecture of a system for Clinical Knowledge Assistance.

**Chapter 8:** This chapter is about the evaluation of the techniques developed in this thesis and the discussion of the results.

**Chapter 9:** This chapter concludes the thesis by revisiting the contributions, the objectives, and by presenting the future directions.
CHAPTER 2

BACKGROUND: A LITERATURE SURVEY

2.1 Introduction

In this chapter, we present a survey of the literature related to the research presented in this thesis. The concepts reviewed provide the theoretical and conceptual background for this thesis. We present the survey of different existing formalisms (models) for the computerization of clinical practice guidelines. In next section, we review the techniques and methods for accessing CPGs content at point of care. Proceedings to next section, review of search techniques for semi-structured documents for effective retrieval is presented. Concepts related to clinical query models and strategies for the delivery of medical literature are surveyed and analyzed. Building blocks and techniques for computerized processing of medical text and information are described. Towards the end of this chapter, knowledge sources and tools used in this research are described, and summary of the chapter is presented.

2.2 Formalisms for Representing Clinical Practice Guidelines

Clinical practice guidelines (CPGs) formalism defines the mechanism to transform CPGs into computer-interpretable format so that their content could be processed by computer's applications. There are different formalisms for CPGs, which have different methods and properties. These formalisms can be divided into two types of approaches (i)
knowledge-base-centric or model-centric representation, and (ii) document centric representation (Shankar, et al., 2001; Stephen, 2005).

Knowledge-base-centric: formalism under this category uses complex knowledge models to support guideline conversion for patient-specific decision-support (Shankar, et al., 2001). These formalisms have been popular for guideline-based decision support system researcher and developers (Stephen, 2005).

Document-Centric: this type of formalism uses guideline document as the authentic source of domain knowledge to be the reference repository of guideline specification (Shiffman, et al., 2004 a; Sonnenberg and Hagerty, 2006). Formalisms like GLIF, PROforma, Arden Syntax, EON, Asbru, SAGE etc fall under knowledge-based centric formalism and GEM, HGML, ActiveGuidelines etc fall under document-centric representation.

In this section, we present the significant models to elaborate their methods to support the modeling process.

2.2.1 Overview of Knowledge-base-centric Formalisms

We review first those guideline formalisms that fall under Knowledge-base-centric category.

2.2.1.1 PROforma

PROforma is a knowledge representation formalism designed for executable process modeling of the clinical guidelines (Fox J, et al., 1998). It was developed at Cancer Research UK. PROforma represents guidelines as a directed graph whereby nodes represent tasks and arcs represent scheduling constraints (David, et al., 2003). In
PROforma guidelines are modeled as a set of tasks and data items. Its task ontology divides task into four classes: plans, decisions, actions, and enquiries.

Keystone: It is also termed as root task. It contains number of attributes that are common to all four tasks.

- Plans: These are collection of tasks, which are grouped together based on some goals and are the building blocks of guidelines.
- Decisions: These are the points where choices are made from the presented options.
- Actions: These are certain procedures that are executed in external environment based on decisions.
- Enquiries: These are the requests for information needs to be acquired from some person or external system.

Guidelines in PROforma are stored using Red Representation language (R²L) i.e. a time-time oriented representation language (Clercq, et al., 2004; Peleg, et al., 2002).
PROforma is supported by a graphical editor for a guideline authoring process, and an
engine for the execution of guidelines. The modeling process consist of two steps, (i)
constructing guidelines into four tasks types using graphical editor, and (ii) automatic
conversion of populated graphical structure into database to be used for execution
purpose.

2.2.1.2 GLIF

The Guideline Interchange Format (GLIF) was developed by InterMed Collaboratory laboratory including researcher at Stanford Medical Informatics, Harvard University, McGill University and Columbia University. It was first published in 1998 (Ohno-Machado, et al., 1998). It was developed to share guidelines among different institutions and software systems. In GLIF modeling of a guideline is performed as a flowchart with a specific syntax (Georg, 2005 b). GLIF formal representation consists of an ontology for representing guidelines as well as a medical ontology for representing medical data and concepts.

Each guideline in the GLIF is represented as a set of nodes, which are linked together in a temporally sequenced graph. Each node represents one of the five classes defined in the ontology (Clercq, et al., 2004; Katharina, 2005; Ohno-Machado, et al., 1998). Due to extensive coding efforts and some other problems (Georg, 2005 b), a new version of GLIF was introduced i.e. known as GLIF3. In GLIF3, XML based syntax is used to represent objects and instances of objects in text format (Clercq, et al., 2004). However, a formal expression language ‘GELLO’ (Sordo, et al., 2003) is included for specifying decisions criteria and patient states in GLIF3 (Clercq, et al., 2004).
Encoding of guidelines in GLIF3 is performed at three levels of abstraction: (i) Conceptual level, (ii) Computable specification, (iii) Implementable specification (Katharina, 2005). The guideline content is represented in its algorithm that is composed of guideline steps. The guideline steps are (i) action, (ii) decision, (iii) branch, (iv) synchronization, and patient state (Clercq, et al., 2004; Ohno-Machado, et al., 1998). Normally, two tools are used to perform GLIF-based knowledge modeling, Protégé’ (Grosso, et al., 1999) and GEODE (Greenes, et al., 1999).

2.2.1.3 GLARE

GLARE, “Guideline Acquisition, Representation, and Execution” is a domain-independent formalism for clinical guidelines. It consists of three components: (i) a representation formalism, (ii) a knowledge authoring tool, and (iii) a guideline execution tool. GLARE is based on the modular architecture that separates out the representation of clinical guidelines and their execution (Terenziani, et al., 2001). Its representation language is designed to achieve a balance between expressiveness and complexity (Terenziani, et al., 2001). GLARE defines an ontology that has limited but focused primitives. The ontology consists of different classes of actions that can be classified into two types: (i) Composite Actions, and (ii) Atomic Actions.

Composite Actions are kind of plans that decompose into hierarchical sub actions. Atomic actions are further divided into four actions (i) Decision Action, (ii) Query Action, (iii) Conclusion, and (iv) Work Actions (see Figure 2.2).
• Decision Action: represents the choices to be made,
• Query Action: represents the request for information needs,
• Conclusion: represents the different outcomes of decision process.
• Work Action: represents the activities of care personnel.

GLARE has an intelligent guideline acquisition interface for expert clinicians to model the guidelines and Artificial Intelligence (AI) temporal reasoning technique to check the consistency of temporal constraints imposed between actions (Terenziani, et al., 2002; Terenziani, et al., 2001).

2.2.1.4 SAGE

SAGE, “Standards-based sharable Active Guideline Execution” was introduced in 2002 to create a methodology and an infrastructure required for the integration of decision-support technology for guideline-based care in commercial clinical information systems (Samson, et al., 2007; Samson, et al., 2004). It uses a deployment-driven
SAGE model transforms guideline as recommendation sets. These sets consist of either Activity Graph or Decision Maps. Activity Graph represents guideline-directed processes and Decision Maps represent recommendation involving decisions at a time point. In SAGE, a recommendation either describes the preferred choice in a management decision or it recommends series of actions to be performed (Katharina, 2005; Samson, et al., 2007). For encoding guidelines recommendations, SAGE recommendation set has four nodes: (i) Context, (ii) Decision, (iii) Action, and (iv) Route. Guideline representation, in SAGE, is performed using Protégé (Grosso, et al., 1999; Samson, et al., 2007).

2.2.1.5 EON

EON model was developed at Stanford University in 1996 (Mark, et al., 1996). It is a component-based architecture suite of models and software components, which are used for guideline-based applications. The EON guideline modeling and execution systems are the parts of EON architecture. EON uses Dharma model as a guideline model to define the knowledge structure of guidelines. Dharma model is a non-monolithic that does not consist of fixed number of primitives (e.g. decisions, actions etc) rather it has a standard set of primitives that can be extended with task specific sub-models. The standard primitives are: scenarios, decisions, actions and goals.

EON contains patient data model in a form of ontology that defines the classes and attributes to represent the patient data. It also contains Medical-specific model that
consists of a medical ontology to model the structure of domain concepts (e.g. drugs, treatments etc). EON model uses the internal frame-based Resource Description Format (RDF) and makes use of Protégé (Grosso, et al., 1999) to describe the models and guidelines. EON defines the execution architecture for the development of guidelines’ execution engine (Clercq, et al., 2004).

2.2.1.6 Asbru

Asbru is a task-specific plan representation language. It represents clinical guidelines as time-oriented skeletal plans, which are plan schemata at various levels of details (Clercq, et al., 2004; Yuval, et al., 1998). The skeletal plans are supported by (i) representation of high-level goals (intentions), (ii) representation of temporal patterns and time annotations, (iii) development of user interface to visualize plans. In Asbru intention-based model is used to represent the clinical guidelines as skeletal plans whereby the functionality of each plan is modeled by number of knowledge roles. The knowledge roles for Asbru intention-based model to represent high-level goals are (i) Preferences, (ii) Intentions, (iii) Conditions, (iv) Effects, and (v) plan body (Yuval, et al., 1998). To specify the temporal aspect of a plan, a time annotation specifies four points (Clercq, et al., 2004). The four points are: the earliest starting shift (ESS), the latest starting shift (LSS), the earliest finishing shift (EFS) and the latest finishing shift (LFS). In addition, two durations can also be defined: maximum duration and minimum duration.

Guidelines are represented in terms of plans in Asbru. To represent different relation among plans and to describe the behavior related to execution of the plans and synchronization, Asbru defines different types of plans, which are: Sequential plans,
Parallel plans, Any-order plans, Unordered plans, Subplans, Cyclical plans (Andreas, et al., 2002). The Backus-Naur Form (BNF) is used as a formal syntax of Asbru language beside this, XML-based version of the Asbru was defined and published (Clercq, et al., 2004; Miksch S, et al., 1997). Since a plan can be represented in XML, it is human readable but understating is very difficult for clinicians without the proper training of semantic and syntactic knowledge about the representation language (Katharina, 2005).

2.2.1.7 Arden Syntax

Arden Syntax developed in 1989 is intended as an open standard for the representation and sharing of medical knowledge in a procedural manner (Clayton, et al., 1989). It consists of “Medical Logic Modules” (MLM), which are rules to trigger certain actions. Each MLM uses four main slots: ‘evoke slot’, ‘logic slot’, ‘data slot’, and ‘action slot’ (Clercq, et al., 2004). Arden Syntax facilitate sharing of simple guidelines, it was not designed for complex guidelines.

2.2.1.8 GUIDE

GUIDE, developed at University of Pavia in Italy (Quaglini, et al., 2001), is a part of “Guideline modeling and execution framework”. It provides an integrated infrastructure of the medical knowledge management thorough a workflow called careflow (Georg, 2005 b). Its formal model is based on ‘Petri nets’ to model guideline workflow. It also provides formal decision analytic models for decision making (Sonnenberg and Hagerty, 2006).
2.2.2 Overview of Document-Centric Formalisms

In this section, we present a review of significant formalisms that fall under Document-Centric category.

2.2.2.1 HGML

Hypertext Guideline Markup Language (HGML) is a guideline representation method based on the markup concept of converting text-based clinical guidelines into computer-interpretable format. HGML consists of basic set of tags for annotation of guidelines and is XML compatible. Guideline developer simply tags the guideline within the web browser using HGML structure (that consists of its basic tags) (Hagerty, et al., 2000).

HGML conforms to the standard mark up language format and allows to add additional attributes with the begin/end tag structure. It also provides advanced HGML tags to support the inferences about the guideline content. HGML mark up editor insert tags directly into the original document. This produces XHTML document with original text as well as tags representing its basic structure like statements, variables, conditions, recommendations (Sonnenberg and Hagerty, 2006). In HGML, ability to define the relationships between tags is an important difference from rule-based languages (Georg, 2005 b).

2.2.2.2 GEM

Guideline Element Model (GEM) is an XML-based guideline document model that is intended to support the transformation of natural language guidelines into a standard computer-interpretable format. It can store and organize heterogeneous information in the
clinical practice guidelines (Shiffman, et al., 2000). It describes concepts pertinent to guideline representation, attributes of those concepts, relationship among the concepts. In GEM elements are basic units of information that store data and define structure by virtue of their position in the hierarchy of the document (Georg, et al., 2005 a; Shiffman, et al., 2000). GEM is constructed as a hierarchy with more than 100 discrete elements (see Figure 2.3). Main concepts in GEM are associated with guideline’s identity, purpose, developer, intended audience, target population, method of development, testing, review plan, and knowledge components (Shiffman, et al., 2000).

Knowledge components are significant elements in terms of capturing guideline’s clinical information and knowledge. A knowledge component includes recommendations that consist of conditionals and imperatives, definitions, and algorithms (see Figure 2.4 for knowledge component hierarchy). GEM is considered more comprehensive and expressive to represent the information and knowledge in guidelines. It helps in the flexible, comprehensible, sharable, and reusable knowledge representation that is

![Figure 2.3 GEM elements hierarchy (Yale School of Med, 2006)]
As GEM is an abstraction of a guideline document, it has several limitations. GEM has limited ability to resolve the ambiguities present in the guidelines, however, it can present them to user for resolution. GEM is not full comprehensive, therefore, additional elements, attributes and relationships may be required to adequately encode the guidelines depending on the needs of stakeholders (Shiffman, et al., 2000).

Guidelines are transformed into computer-interpretable format by marking up process using GEM Cutter tool (Yale School of Med, 2006). Marking up of guidelines depends on the encoding strategy as the GEM model is simply an abstraction of the guidelines.

![Figure 2.4 Knowledge component element's hierarchy in GEM II(Yale School of Med, 2006)
guideline document. It relies on extrinsic systems to apply in ways that are useful. Based on some encoding strategy semantic refinement (precise specification of guideline knowledge) may be performed during marking up process or after marking up. The marked-up guideline is used in workflow integration depending on the local requirements (Shiffman, et al., 2004). Geroje et al. (2003) presented their interpretative framework for GEM-encoding that is based on different levels of normalization and atomization of concepts of the guideline content and help represent the guideline in “If–then–with’ rules.

Representation of GEM in XML format provides number of advantages (McGrath, 1998). As XML is self descriptive that may improve the searching, indexing, and locating information (Shiffman, et al., 2000). Furthermore, it is an open standard that support the development of tools for document processing, etc (Shiffman, et al., 2000; Shiffman, et al., 2004).

2.2.3 Guideline Formalisms Representation Languages

GLIF, GUIDE, PROforma, Asbru, EON have almost similar approach in terms of primitives (steps, tasks or Plans) that describe the control structure of a guideline. GALRE represents the guideline structure in terms of composite actions and atomic actions. SAGE defines the guideline as recommendation sets. These sets consist of either activity graph or decision maps. Arden Syntax models each guideline as an independent modular rules. Arden Syntax, PROforma, and Asbru use BNF language for their representation. GLIF uses UML, GUIDE uses workflow process definition language, EON and SAGE rely on the internal syntax of Protégé. GLARE uses its own representation language designed to achieve a balance between expressiveness and complexity.
GEM and HGML use markup methodologies to serve guideline text segments relevant to patient context. GEM uses XML for its representation while HGML uses HTML and XML (XHTML).

2.3 Computerized Clinical Practice Guidelines Content Access at Point of Care

Clinical practice guidelines are transformed into computer-interpretable format using guideline formalism. These computer-interpretable guidelines are used by applications such as: decision-support systems, clinical information systems, clinical assistance systems, etc to provide clinicians access of guidelines content (data, information, knowledge) at point of care. Most of the formalisms represent computerized clinical practice guidelines (C-CPGs) by decision logics (in plans, actions, tasks,) so that C-CPG content can be accessed by an execution engine. Execution engines can be divided into two types:

(i) event-based approach,

(ii) rule-based approach.

The event-based execution engine can be used in asynchronous way in a continuous system where events-are handled asynchronously. The rule based execution engines are monitored by another operator that supervises and controls the rules that can be triggered in any time in a synchronous manner (David and Antonio, 2008). There are few other approaches that make C-CPG content available at point of care. One such approach uses predefined clinical questions which are grouped for certain “infobuttons” that represent the problem from EMR. These questions are linked with the tagged answers from the guidelines. Retrieval of relevant answers is performed by clicking the “infobutton” for
specific question(s). All approaches for accessing CPGs content at point of care depend on the electronic patient record (EMR) or patient data and need to be integrated into careflow framework.

One approach that developed guideline search engine called ‘Vaidurya’, uses the clinician’s information needs in terms of a query and performs context-sensitive retrieval. It retrieves full guidelines (instead of content of guideline) that match the clinician’s query. The list of retrieved guidelines is ranked based on the relevancy to the clinician’s query.

In this section we present some of the computer applications developed for accessing the guidelines content in clinical settings.

2.3.1 GLEE

GLEE is a guideline execution tool, which was developed to provide access of those guidelines’ content which are encoded in GLIF3 (guideline formalism) (Wang, et al., 2004; Wang and Shortliffe, 2002). The execution model of GLEE is based on “system suggests, user controls” approach. It deploys a tracing system to record an individual patient’s state when the guideline is being applied to that patient (to record the state of guideline steps and their transitions). GLEE supports an event-driven execution model once it is attached to a clinical event monitor in a local environment (Wang and Shortliffe, 2002). Its architecture has three conceptual layers. These conceptual layers are: data, business logic, and user interface (David and Antonio, 2008).
Data layer: this layer represents the EMR, guideline repository (guidelines in RDF files) and the clinical event monitor. The clinical event monitor initiates the execution of clinical guidelines using an event-driven model (David and Antonio, 2008).

Business logic layer: consists of GLEE run time execution engine. It is based on client/server model. The server side of execution engine interacts with the data layer and clients interact with the users.

User interface layer: This layer contains clinical applications that exchange data with the upper layers (David and Antonio, 2008).

2.3.2 GLARE

Guideline Acquisition, Representation and Execution (GLARE) is a framework to acquire and execute clinical guidelines (Terenziani, et al., 2001). It executes the clinical guidelines encoded in its own guideline representation formalism (see 2.2.3). GALRE represents the guideline structure in terms of composite actions and atomic actions. For each action, there is a set of preconditions and set of conclusions. An action is triggered once its preconditions are satisfied and conclusions’ set follows after the activation of the action.

GLARE architecture is modular in nature that separates out acquisition phase (guideline representation phase) when a guideline is introduced in a system and execution phase and when a guideline is applied by physicians to a specific situation (Anselma, et al., 2007; David and Antonio, 2008). Its architecture has three layers: upper, intermediate and lower layer. The upper layer is called system layer, the intermediate layer is called XML layer, and the lower layer is called DBMS layer. Acquisition and execution module
are at system layer. DBMS layer has access to all databases where all data for creating and executing guidelines are stored. Furthermore, this layer physically connects the upper layer. The XML layer allows the data exchange between DBMS and System layer. Each patient has its own medical record contained in a patient database (DBMS layer). Execution module executes clinical guidelines using the appropriate retrieved data from database (David and Antonio, 2008).

2.3.3 Arezzo

Arezzo is a commercial guideline execution engine that uses clinical guidelines encoded in PROforma (Fox, et al., 2006). PROforma is an executable process modeling formalism (see section 2.2.1). Its architecture consists of three components, which are: (i) composer, (ii) Tester, and (iii) Performer. The Composer is used to transform the guidelines using PROforma, the Tester is used to test and verify the guideline logics before deployment, (iii) The Performer is an inference engine that executes the guideline. The execution of the guideline is carried out by using patient related data that is stored in EMR (existing healthcare systems). During execution of a guideline in a Performer engine, a task can take few states: it may change its internal state that depends on whether the task is awaiting for data, suspended, finished or it can not be accomplished in the current state of the patient (David and Antonio, 2008; InferMed, 2008).

2.3.4 Digital Electronic Guideline Library (DeGeL)
DeGeL is a web-based modular and distributed architecture that uses Asbru for the transformation of text-based guidelines to a formal representation (Shahar, et al., 2004; Young, et al., 2007). It maintains the repository of the guidelines and facilitates user to search, browse, and retrieve all available guidelines.

The textual guidelines are gone through a process where experts add semantic information to the guidelines and then the guidelines are processed to take the final representation form. Asbru organizes a clinical guideline as a library of Asbru plans (see section 2.2.6).

DeGeL contains an execution engine module that is called “Spock”. Spock incorporates an inference engine that can retrieve data stored in a patient’s medical record for execution. It consists of different modules (David and Antonio, 2008):

- a module to store guidelines,
- a parser to interpret the content of the guidelines,
- a controller that synchronizes the communication between system and external services.

In Spock, actions are monitored in asynchronous manner. It also controls starting or resuming guideline execution (David and Antonio, 2008; Young and Shahar, 2005).

2.3.5 NewGuide

NewGuide is a framework for modeling and executing clinical guidelines (Ciccarese, et al., 2004; Ciccarese, et al., 2005). NewGuide uses GUIDE (see section 2.2.10) formalism for the representation of clinical practice guideline. It facilitates modeling of complex concurrent processes, temporal data, and hierarchical issues.
NewGuide uses UMLS codification (Lau and Shakib, 2005) to describe medical terms and procedures. It deploys a “Medical Text Mark-up language” termed as “Guideline text Mark-up” to describe tasks within a guideline (David and Antonio, 2008; Kumar, et al., 2002).

In NewGuide, the guidelines are executed by an inference engine. The inference engine consists of three main components: a general manager, a message manager, and an instance manager. A clinician invokes the inference engine that creates an instance of a clinical guideline for the management of an individual patient. Instance manager supervises all the steps taken in the execution of the guideline and at the same time all instances are controlled by the general manager. Once guideline is loaded, instance manager has to collect patient’s data stored in his patient record. The execution engine goes step by step to recommend actions. Message manager is used to facilitate the communication between NewGuide and the external world (environment) (Ciccarese, et al., 2005; David and Antonio, 2008).

2.3.6 Guideline-Centered Clinical Decision Support

Shiffman et al. (2004 b) presented their approach for guideline-centered clinical decision support design. They used GEM formalism for document-centric approach to develop a standalone CDSS for tobacco cessation counseling in primary care office. They defined guideline-based CDSS as a 2 part process. In first part guideline knowledge is translated into computer-interpretable form, which is performed using GEM markup. In second part information not provided by guideline are added. They selected the important information relevant to in-office smoking cessation practice from the guideline. This important information was less than 10% of the original guideline text. They used GEM
cutter to mark up the important but focused information from the guideline. Knowledge components are used to classify definitional material, algorithmic information and guideline recommendation. After marking up, GEM-Based guideline is submitted to extractor to extract the decision variables and the actions. These actions and the decision variables are formatted in separate lists to perform atomization of concept, disambiguation, and adjustment of level of abstraction and other tasks. Decision variable are represented like DV1, DV2, etc and actions are represented like A1, A3, A5, etc. They categorized the actions into the types described by (Essaihi, et al., 2003) e.g., monitor, prescribe, perform, etc. Decision logic is defined for one or more decision variables and corresponding action or actions. These actions are triggered based on the inference engine that is integrated with patient data.

2.3.7 Rapid Answer Retrieval from CPG

Poon et al. (2006) developed a method to the retrieval of answer from a clinical practice guideline at point of care. Their method is based on the “infobuttons” approach (Cimino, et al., 2002). In their method predetermined clinical questions are linked with the answers from clinical practice guidelines. Clinical question or set of clinical questions were classified using (Ely, et al., 1999) taxonomy. Each question or set of question were grouped under “infobuttons” that are linked to the problem in an electronic medical record. In order to provide quick answer to the question they tagged the guideline using “named destination” tagged. They developed an XML schema to define the structure of question and “named destination” tag. So, each question(s) was/were linked to the answer(s) from the clinical guidelines. Each “infobutton” display a list of questions, which can be invoked by clicking at “infobuttons”.
2.3.8 Vaidurya Clinical Guidelines Search Engine

Robert and Yual (2009) developed a clinical guidelines search engine that is called Vaidurya. It is a part of DeGeL framework (Shahar, et al., 2004). In Vaidurya, full text search and two additional search methods are implemented:

(i) concept based search: it relies on pre-indexing guidelines in a clinically useful fashion,

(ii) context-sensitive search: it relies on semi-structuring the guidelines using the given ontology and then searching for terms within specified labeled term.

They used similarity function and modified vector space model for the retrieval of guidelines. Vaidurya retrieves full clinical guidelines in response to a query from the digital guidelines repository.

2.4 Semi-structured Document Searching

Many guideline formalisms represent the guideline content in semi-structured format. Incorporating, the best practices and approaches from semi-structured document searching methods, would help enhance the efficacy of the search techniques for CPGs content access at point of care. In this section, we present significant approaches for efficient and effective search in semi-structured documents.

In searching of semi-structured documents, advanced approaches exploit the content and document structure. Different indexing strategies have been proposed to capture the content of semi-structured documents and their structure. Some methods exploit semantic information, and meta-information pertaining to content by using
ontology and structure. Methods define the ranking functions and the searching algorithm based on their index structure and approaches for semi-structured documents searching. Existing IR algorithms are also modified to be used for semi-structured documents retrieval. The key idea for searching and retrieval of semi-structured documents is to exploit the content, structure and additional meta-information (if available) to enhance the efficiency. It requires development of techniques, algorithms and indexing strategy based on the requirements, approaches and nature of semi-structured documents.

Kotsakis (2002) presented their technique for information retrieval in XML documents where their approach is to view XML documents as a collection of text documents with additional tags. In this approach existing IR (information retrieval) techniques are adapted to achieve sophisticated search on XML documents. An indexing strategy is introduced that extends the inverted file structure to search XML documents. It is accomplished by integrating the XML structure in the inverted file by combining the inverted file with a path index.

They introduce the summary tree that keeps the structure of original documents with only important content (unimportant terms were removed). This summary tree is

![Figure 2.5 Loading an XML summary tree into index structure (Kotsakis, 2002)]
used for indexing. The path index is a hierarchy of tags, which records every single path in the collection. A list of documents is also stored alongside each entry to the inverted file (see Figure 2.5). A ranking function is introduced using inverse document frequency approach for ranking the search documents.

Abbaci et al. (2006) developed an XML retrieval system which takes into account document structure and document content during indexing process. Their system architecture is based on two independent components: (i) XML documents analyzer and (ii) query processor. The XML document analyzer parses the XML document for content and structure analysis. The query processor, receives the user’s query, analyzes it to extract the operands (AND, OR, SIB, Not etc) when the query contains the operators and finally retrieves the corresponding XML fragment (see Figure 2.6).

In their method, content and structure of XML documents are indexed using three indices:

(i) A file manager table: it archives the files indexed,
(ii) A node list table: it keeps the structure of the documents archived,
(iii) A wordlist table: it keeps every word appearing in documents archived.
Botev and Shanmugasundaram (2005) presented their approach for context-sensitive keyword search and ranking for XML. In their view, many XML document collections have a hierarchical structure with semantic tags, which allows users to specify the context of their search more precisely. In their system architecture, context evaluator evaluates the user query for the search context. User query consists of two parts: (1) the keyword search query, and (2) the search context.

![Figure 2.7 System architecture for context-sensitive keyword search (Botev and Shanmugasundaram, 2005)](image)

The Context evaluator takes the specification of the search context from the user query and returns a set of XML element IDs “I” in a way that descendants of I define the search context. This search context consists of all the elements in the sub-trees that root at the elements from the specification value. The query engine uses the search context (set of IDs “I”) and applies the user keyword-search query to rank the elements in the search context with respect to the query (see Figure 2.7). It uses the index structure and ranking formula to produce the ranked query results. In their indexing strategy inverted list index structure (Salton, 1989) is extended to capture the XML hierarchy in the inverted list entries so that nested elements that contain the query keywords can be
returned as the results. Inverted list is also structured so that entries that do not belong to search context can be skipped.

Xin-Ye et al. (2006) developed a search engine for XML documents that defines the semantic search scheme based on their indexing strategy and ontology. In their approach semantics of both XML documents and the query are enriched by domain ontology annotation. Furthermore, additional information provided by the structure of XML document is considered by using indices like ‘node path index’, ‘semantic keyword index’ and ‘element tag index’. A searching algorithm is used to facilitate the semantic search for the XML documents. This algorithm calculates the semantic correlation coefficient (i) between two nodes, (ii) between keyword and a node, and (iii) by inferring from ontology base. Results are ranked based on the ranking function “R-value”.

The index strategy defines three indices to capture the above defined information, these indices are described below.

Node Path Index: it indexes all element nodes in a XML document that is linked to their parent node. Every element node has a unique identifier in a XML document. The “docID, nodeID” scheme expresses an element node where ‘docID’ represents XML document number, ‘PnodeID’ is used to denote its parent node (see Figure 2.8).

Figure 2.8 Node Path Index (NPI)(Xin-Ye, et al., 2006)
Inverted tag node index: This index is used to store all the element tags with links to their element node denoted by “(docID, nodeID)” (see Figure 2.9). In this index element tag is represented by its annotated term that expresses the concept. Every element node has a data type which denoted by ‘D’ in Figure 2.8, which has specific integer value.

Inverted semantic keyword index: It indexes keywords appeared in the element content whereby every keyword is linked to its element node. Every keyword is assigned weight that is denoted by ‘W’ in Figure 2.9.

### 2.5 Clinical Query Models and Delivery of Medical Literature

In this section, we review the clinical query models/generic queries/query types and methods of delivering the medical literature at point of care. In these methods and techniques concepts like scenario specific query, query types, tasks oriented query, query templates, and query categories have almost similar meaning. Recent studies reveal that almost 60% of clinicians query center on specific query types or scenarios (Cimino, et
Several studies have shown the need of healthcare professionals to access evidence-based literature, information and knowledge in clinical practice, and other knowledge sources, however, these studies have also identified the problems healthcare practitioners face in seeking such information (Chambliss and Conley, 1996; Covell, et al., 1985; Ely, et al., 1999; Gorman and Helfand, 1995; Gorman, et al., 1994; Hersh, et al., 1996; Shelstad and Clevenger, 1996; Timpka, et al., 1989).

In order to minimize hurdles in seeking medical information and knowledge, multiple studies have investigated healthcare practitioners’ information needs and information seeking behavior. Based on the healthcare practitioners’ information needs and information seeking behavior, different query models, search strategies and techniques to categorize queries into query types have been proposed (Cimino, 1996; Cimino, et al., 1993 a; Cimino and Barnett, 1993 c; Cimino, et al., 1993 b; Ely, et al., 2000; Ely, et al., 2005; Haynes, et al., 1994; Haynes, et al., 2005 ; Mendonça, et al., 2001; Price, et al., 2002; Wanda and Henry, 2000; Wilczynski, et al., 2001).

For example, Cimino et al. (2002) showed that clinicians’ questions are predictable that center on specific tasks. A focused and context-sensitive clinical query either from healthcare practitioner or system-generated should appropriately frame context, problem and clinical intention. It may be done by using a list of specialized problem-specific medical-terms/concepts as the search query (Abidi, et al., 2005 b; Olena, 2005). Following, we present the approaches for clinical query models and delivery of medical literature.
Cimino et al. (1993 a) developed a set of general-purpose questions for clinicians called generic queries with a noun—relation—noun structure. To construct, generic MEDLINE search strategies, relational information encoded in the query structure were used. MeSH subheading were used to express the relationship between search terms (Cucina, et al., 2001). User could select generic queries in one of two ways: (i) user may type in question that is analyzed to identify most relevant generic query or (ii) user may indicate patient data of interest, from a admission profile of patient record system, and then pick one of several potentially relevant pre-defined questions. The rationale of this approach is that clinician’s needs are matched to one of a set of pre-defined general query types for which retrieval strategies have been developed in advance (Cimino, et al., 1993 a). Using these generic clinical questions, an application Medline Button was developed that was used to perform bibliographic searches directly from patient data (Cimino, et al., 1993 b). It was based on five steps to perform bibliographic searches: (i) identifying the patient data from patient record that raises question, (ii) selection of pre-defined generic question that conforms to selected patient data, (iii) translation of patient data into terms used for bibliographic indexing, (iv) conversion of user selected question into search strategy, and (v) transfer of search strategy to search engine for bibliographic searches (Cimino, et al., 1993 b).

The intention of this approach was to allow healthcare practitioners to select relevant data items from clinical information system and match up a relatively small list of generic queries based on the terms selected. The specific terms selected from the clinical system were then incorporated into generic question, which are made specific to user’s question. Once the question was selected, the terms were filled into appropriate
blanks of a corresponding MEDLINE search strategy, which were sent to MEDLINE system (Cimino, 1996). Modified form of Medline Button known as Infobuttons. It was developed using these generic questions and terminological knowledge for the retrieval and linking of online medical sources to user selected clinical questions based on patient data (Cimino, et al., 1997). The idea of infobuttons is almost similar to Medline Buttons with some additions like terminological knowledge sources e.g. MED (Cimino, et al., 1997).

For example, in infobutton application a generic question might be e.g. “Tell me about <term>”. The <term> is filled in at run-time with a term (data) selected by user from patient record/clinical information system. This term is used to exploit the hierarchical and semantic links in the terminological knowledge source (MED) to identify the additional terms that can be used in the generic questions (Cimino, et al., 1997). Furthermore this additional information about the term is used to allow the infobutton to select the generic questions that are relevant. Different prototype infobuttons applications have been developed based on this approach (Cimino, et al., 1995; Cimino, et al., 1995; Elhanan, et al., 1996; Zeng and Cimino, 1997).

Haynes et al. (1994) proposed a method to develop optimal MEDLINE search strategies for the retrieval of sound clinical studies of the four query types: etiology, prognosis, diagnosis, treatment. In this study, retrieval performance of methodologic search terms and phrases in MEDLINE was compared. In this comparison manual review of each article for each issue of ten internal medicine and general medicine journals for two years 1986 and 1991 was conducted. The search strategies were based on the MeSH terms and text words related to research design features. These search strategies were
used to evaluate MEDLINE strategies designed to retrieve literature that met the basic methodologic criteria for clinical practice (Haynes, et al., 1994).

“These search strategies were treated as diagnostic tests for sound studies and the manual review of the literature was treated as the gold standard” (Haynes, et al., 1994). The sensitivity, specificities, accuracy, and precision of MEDLINE searches were determined based on the concepts of diagnostic test evaluation and library science (Haynes, et al., 1994).

This method has been incorporated at PubMed as a built-in clinical query search filters (Haynes, et al., 1994; Wilczynski, et al., 2001). Abidi et al. (2005 a) used query types defined by (Haynes, et al., 1994) and proposed a method to generate automatic query from selected text of ‘Computerized-CPG by medical practitioner, to retrieve medical literature from PubMed.

Mendonca and Cimino (2001) studied MEDLINE MeSH terms associated with the four basic clinical tasks: diagnosis, etiology, prognosis, therapy. Their goal was to automatically categorize citations for the task-specific retrieval.

Ely et al. (1999) conducted an observational study for the analysis of the information needs. In their observation, 10 most common forms of generic queries were reported containing one or two concepts and a semantic relationship (Cucina, et al., 2001; Ely, et al., 1999). Ely et al. (2000) developed the taxonomy of doctors’ questions about patient care that could be used to help answer clinicians’ questions. This study resulted in the taxonomy of 64 generic questions types. The rationale behind is that clinical questions in primary care can be categorized into a limited number of generic types (Ely, et al., 2000).
Poon et al. (2006) developed a method that is based on the relevant questions for the retrieval of answer from clinical practice guidelines by using the taxonomy proposed by (Ely, et al., 2000). Mendonca et al. (2001) proposed a model of information needs for query definition and refinement based on the taxonomy of generic clinical questions developed by (Ely, et al., 2000). This model is to be used for bibliographic search on medical databases.

Cucina et al. (2001) developed a comprehensive set of generic queries for information retrieval from electronic medical information resources. The formulation of these queries was performed by using unions of UMLS semantic types to define the allowed value for each query concept. The difference in their work from others (Cimino, et al., 1993 a; Ely, et al., 2000) is that they tried to develop a query model suitable for both full-text indexing and retrieval by defining more compact query set to keep the indexing task manageable (Cucina, et al., 2001). These generic queries were used in the system (Berrios, 2000; Berrios, et al., 1999; Dugan, et al., 1999) for semi-automated indexing and information retrieval for full-text information resources (Cucina, et al., 2001).

Price et al. (2002) developed a prototype application ‘SmartQuery’ that provides context-sensitive links from an electronic patient record to medical knowledge sources. This system aid in forming complex queries from information presented in various parts of EMR. Once patient’s EMR is accessed, a base set of MeSH terms is collected from patient’s diagnosis list. ICD 9 code is used as a proxy for a real patient problem that is translated into to MeSH terms. In SmartQuery user checks the boxes next to items relevant to question and click add buttons. This causes MeSH term (created form ICD-9
codes) to be added corresponding to data underlying the displayed information. User may also enter additional terms. To send query, user needs to check the collected terms of interest and the information sources to be queried (Price, et al., 2002). These user selected terms are used to formulate appropriate query for the corresponding selected resource. Each resource requires different query formulation process. This system resembles the Infobuttons (Cimino, et al., 1997). Unlike Infobuttons, it does not use generic queries. SmartQuery tries to elicit terms that describe the clinical situations. The hypothesis of this approach is that allowing user to add information from different areas of EMR and to aid their own search terms can be used to facilitate a richer collection of terms that can be used to formulate queries involving multiple concepts (Price, et al., 2002).

Wanda and Henry (2000) proposed a system called “QueryCat” for automatically categorizing user’s initial clinical queries into query types. Their query model consists of ten query types or categories that are the abstraction of user’s specific query. In their method, clinical queries are categorized into their defined query types by lexical analysis and semantic analysis techniques (Wanda and Henry, 2000).

Zhenyu and Wesley (2007) developed a knowledge-based query expansion method. In this method original query is enhanced by adding additional relevant terms, which are query’s scenario specific. In essence, query’s scenarios are similar to query types or categories. In this method UMLS knowledge source is used to extend the original query with additional terms to enhance the retrieval performance.

Demner-Fushman et al. (2008) proposed “infobot” system prototype that is designed for automatic delivery of patient-specific information from evidence based
resources. In this system, search and delivery process is initiated by a new entry in one of the monitored EMR fields. The new value entered into the field is extracted from the EMR and sent to the NLM server to identify the clinical terminology and to deliver information from evidence-based resources, which are pre-selected by the healthcare organization (Demner-Fushman, et al., 2008).

### 2.6 Automatic Processing of Medical Information

Computer-based medical information and text processing in biomedical literature and different medical documents has broad scope in many applications such as information retrieval (Jones and Staveley, 1999; Li, et al., 2004), browsing interfaces in digital libraries (Gutwin, et al., 1999), vocabulary construction (Kosovac, et al., 2000), summarizing medical article (Natioanl library of medicine, 2006), indexing medical literature (Clifford, et al., 2005), document classification and clustering (Jonse and Mahoui, 2000), Knowledge mapping (Shepherd, et al., 2006), knowledge discovery, text mining (Ng and Wong, 1999; Ono, et al., 2001; Tanabe, et al., 1999) and many more. In medical text and information processing applications, concept identification is very crucial.

Concepts in medical text are normally described by noun phrases that also contain the primary information of documents as well (Quanzhi and Yi-Fang, 2006). Majority of concept terms represented by noun phrases make identification of noun phrases one of the fundamental problems for applications in mining medical documents (Quanzhi and Yi-Fang, 2006). Among the noun phrases, identification of keyphrases has been challenging research problem, since keyphrases are more domain-oriented and selective as opposed to noun phrases (Quanzhi and Yi-Fang, 2006). In medical documents noun
phrases and keyphrases contain important information, therefore their classification, mapping, semantic processing using computerized techniques produce more useful information and meanings to medical information processing applications. Concepts, noun phrases and keyphrases identification may require key processes like recognition, extraction, semantics-processing, mapping and classification, which are challenging tasks in medical information processing (Krauthammer and Nenadic, 2004).

Different methods and techniques have been presented for the medical information processing in hospital discharge summaries, radiology reports, abstracts of MEDLINE database articles and patient’s semi-structured data for different applications.

Friedman et al. (1994) developed a natural-language text processor for clinical radiology (known as MedLEE). This system consists of three phases of processing:

(i) Parser,

(ii) Phrase-regularization phase,

(iii) Encoder.

The parser uses a grammar and a lexicon to find the structure of the text and generates preliminary structured output form for clinical information in the text (Friedman, et al., 1994).

The Phrase-regularization phase is used to further reduce the stylistics vacations occurred during natural language processing by incorporating mapping knowledge base (Friedman, et al., 1994).

In the encoding phase standard forms are mapped into unique concepts associated with the controlled vocabulary. The mapping in this phase is performed using synonym
knowledge base. The synonym knowledge base consists of standard forms and their corresponding concepts in controlled vocabulary (Friedman, et al., 1994).

Croft et al. (1991) developed a method where phrases in natural language queries were identified to build structured queries for probabilistic retrieval model. In their method they showed that using phrases in this way improved performance of retrieval (Croft, et al., 1991). Soderland et al. (1995) proposed technique to recognize information about diagnosis and symptoms. In their technique, they used semantic lexicon and dictionary of syntactic and contextual rules.

Nadkarni et al. (2001) used the UMLS Metathesaurus as domain knowledge to identify concepts in medical text for indexing discharge summaries and surgical notes. They used commercial phrase identification program to identify medical concepts. The identified concepts were matched against the entries in the UMLS Metathesaurus whereby matched concepts were used for indexing (Nadkarni, et al., 2001).

Blake and Pratt (2002) conducted study to find potentially useful connections among the medical literature automatically. In their method noun phrases were used to identify the concepts along with additional semantic constraints to determine the connections among the documents. In their study, they used MEDLINE as a source of medical literature and their results showed that noun phrases are more effective than single words in connecting and finding medical literature (Blake and Pratt, 2002).

Majoros et al (2003) developed a method of improving the quality of automatically extracted noun phrases by introducing prior knowledge during the Hidden Markov Model (HMM) training procedure for the tagger. They used a curated lexicon (controlled vocabulary) of phrases extracted from UMLS to modify the HMM probabilities. In their
results they found that this enhancement when combined with appropriate training data improved the quality and relevance of the extracted phrases that in turn enables better accuracy in literature mining tasks (Majoros, et al., 2003).

Kumar et al. (2004) presented a method for developing a knowledge base called BioMap. They used noun phrases extracted from the MEDLINE collection and deployed the UMLS to resolve noun phrases into number of categories. A MetaMap transfer (MMTx) program was used to map the noun phrases to UMLS concepts. BioMaP was regarded as a secondary knowledge resource derived from primary resources such as MEDLINE for medical literature (Kumar, et al., 2004).

Long (2005) developed a technique for extracting diagnosis and procedures from semi-structured discharge summaries by using the structure of a summary to locate the appropriate text. The past medical history and the discharge diagnosis sections in the discharge summaries were used for extracting and coding. In this method limited amount of natural language processing was used to find phrases that might contain the desired data while the UMLS was taken as a knowledge source to find concepts.

Sibanda et al. (2006) developed a technique for statistical semantic category recognizer (SCR) to identify eight semantic categories in discharge summaries. In this technique SVMs are used to predict the semantic category of each word in discharge summaries. Syntactic, orthographic and lexical contextual features as well as ontological information from the UMLS are taken into account during the training of SCR.

Sotelsek-Margalef and Villena-Román (2008) described an approach of information extraction for medical text classification. This approach was deployed in a prototype of expert system called Medical Diagnosis Assistant (MIDAS). MIDAS is
used to suggest medical diagnosis from radiological/clinical patient records. It is based on information extraction and machine learning from clinical histories of previously diagnosed patients. In this approach free text is converted into actionable knowledge using NLP techniques. This actionable knowledge is used to train machine learning system to perform a clinical free text classification (Sotelsek-Margalef and Villena-Román, 2008).

Methods and techniques for medical text and information processing can be classified into four approaches (Krauthammer and Nenadic, 2004):

(i) Dictionary-based approaches,

(ii) Rule-based approaches,

(iii) Machine-learning and statistical approaches,

(iv) Hybrid approaches.

Dictionary-based methods normally use existing vocabularies to locate terms in text. However, Krauthammer and Nenadic (2004) reported that one of the problems in the dictionary-based methods is associated with using straightforward dictionary, which causes many terms in text not to be recognized. It has also been reported that simple pattern matching resulted in low precision and recall rate due to homonymy and spelling variations (Gaizauskas, et al., 2000; Hirschman, et al., 2002; Krauthammer and Nenadic, 2004; Tuason, et al., 2004). Some techniques combine the dictionaries-based approaches with additional processing to supplement medical text and information processing tasks (Krauthammer and Nenadic, 2004; Krauthammer, et al., 2000; Tsuruoka and Tsujii, 2003).
Rule-based approaches normally try to identify terms by re-establishing associated term formation patterns. In these approaches rules are developed to describe the common naming structures for certain term classes using either orthographic or lexical info. These techniques are known to be extremely time-consuming for development and difficult in adjustment to other entities as rules are normally specific in essence (Collier, et al., 1999; Krauthammer and Nenadic, 2004).

Statistical approaches are deployed for general terms recognition i.e. keywords (Andrade and Valencia, 1998). Machine learning approaches are normally developed for specific entities and are integrated with term recognition and term classification (Sibanda, et al., 2006). Machine learning approaches are based on training data to learn the prominent features for term recognition and classification. One of the main problems in such approaches is the reliable training resources, because such resources are not widely available (Benson, et al., 2000; Krauthammer and Nenadic, 2004). Other challenges in such approaches are selection of discriminating features and detection of term boundaries (known to be the most difficult to learn)(Krauthammer and Nenadic, 2004).

Hybrid approaches usually combine different methods, for example, ‘rule and statistically –based approaches’. Such approaches also use various resources such as pre-compiled lists of specific terms, words etc for term recognition task (Krauthammer and Nenadic, 2004; Tanabe and Wilbur, 2002).
2.7 Knowledge Resources and Tools

In this section, we briefly describe the tools and knowledge resources used in this thesis: (i) Unified Medical Language system (UMLS), (ii) MEDLINE, (iii) PubMed, (iv) MetaMap.

2.7.1 Unified Medical Language System (UMLS)

The UMLS project was commenced in 1986 by the National Library of Medicine (NLM) to reduce the fundamental barriers to the application of computers to medicine by providing a complete (to great extent—almost) spectrum of current exiting medical knowledge (Humphreys, et al., 1998). The UMLS is a large medical domain knowledge base to support the computer systems and applications in understanding the meaning of biomedicine and health language and to find the different information about the biomedical concepts (USNLM, 2008 a). It integrates and maps information procured from different medical information vocabularies, resources, dictionaries, and ontologies (Humphreys, et al., 1998).

The UMLS consists of three main knowledge sources that provide different information pertaining to medical concepts (USNLM, 2008 a):

- Metathesaurus—providing access to more than one million biomedical concepts,
- Semantic network—defining categories and relationships,
- SPECIALIST Lexicon & Lexical Tools—providing lexical information and programs for information processing.
2.7.1.1 Metathesaurus

The Metathesaurus is a large repository that is multi-purpose and multi-lingual in nature. It is constructed from more than 100 different thesauri, code sets, lists of controlled terms used in patient care, health service billing, public health statistics, indexing biomedical literature, and basic clinical and health services research. It provides the information about the concepts pertaining to biomedical and health domain, their synonyms and the relationships among them. Its organization is based on concepts or meaning. It links alternative names and views of the same concepts and determines the useful relationship between different concepts (USNLM, 2008 a). As the UMLS Metathesaurus is a quite large repository, it must be customized to be used effectively (USNLM, 2008 a). In our work we are using two of its source vocabularies: (i) MeSH and SNOMEDCT.

2.7.1.2 Semantic Network

The Semantic Network assigns categories (semantic types) to all concepts defined in Metathesaurus and provides relationships among these concepts through semantic types. It contains 135 semantic types e.g. “Disease or Syndrome” or “Pharmacologic Substance” and 54 relationships. Semantic network is hierarchical in structure where Semantic Types are the nodes in the network and the Semantic Relations between them are the links. The current scope of UMLS Semantic Types is quite broad (USNLM, 2008 a).

2.7.1.3 SPECIALIST Lexicon and Lexical Tools

The SPECIALIST Lexicon is a general English lexicon that contains many biomedical terms. It covers both common English words and biomedical vocabulary. It
contains the following lexical information for each word or term: (i) syntactic, (ii) morphological, and (iii) orthographic. Such information is required for SPECIALIST Natural Language Processing System. The Lexical Tool are developed to address the degree of variability in natural language terms (USNLM, 2008 a).

2.7.1.4 MetaMap Transfer (MMTx)

MMTx is a software library (APIs) to make MetaMap program available to biomedical researcher in a generic configurable environment. MetaMap maps text to concepts in the UMLS Metathesaurus. With this software text is processed through series of modules that is used in different tasks for a developed heuristic of computer application or algorithm (USNLM, 2009).

2.7.2 MEDLINE

Medical Literature Analysis and Retrieval System Online (MEDLINE) is a large bibliographic database that contains over more than 16 million references to journal articles in life sciences and biomedicine. It is mainlined by U.S. National Library of Medicine. The MEDLINE records are indexed with NLM’s Medical Subject Headings (USNLM, 2008 b).

2.7.3 PubMed

PubMed was developed by the National Center for Biotechnology Information (NCBI) at the NLM. It is available on Entrez retrieval System. PubMed provides free access to MEDLINE and links to free sites providing full text articles and related resources. It provides E-utilities (WebServices, APIs) to access MEDLINE and other associated databases through software programs (USNLM, 2008 c).
2.8 Summary

In this chapter, we have provided the literature survey for the theoretical background pertinent to this thesis. Existing formalisms for the computerization process of clinical practice guidelines have been reviewed, which provide the conceptual ground for the basic steps required for computerization process.

Techniques to access CPGs knowledge at point of care for different scenarios and applications have been discussed and problem issues and tasks have been presented. Most of the formalisms of CPGs computerization transform CPGs into semi-structured format. To make use of the semi-structured format in effective delivery of CPGs knowledge at point of care, a literature survey has been undertaken for the semi-structured documents search techniques.

A review of clinical query models and the methods for the delivery of medical literature from evidence-based knowledge sources has been conducted for the understating of clinical query formulation related to cognitive and technical tasks. Different techniques and methods have been discussed for the computerized processing of medical text and information. We have descried the knowledge sources and the tools used in this research.
CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter gives a brief description of research methods from the field of computing and information systems that implies to the field of health informatics and describes the tenets of our research methodology for this thesis.

3.2 Overview of Research Methods

Research methods can be classified in various ways but one of the most common classifications is into quantitative and qualitative research methods.

**Quantitative research methods:** These were originally developed in the natural sciences to study natural phenomena. Quantitative methods are research techniques that are used to gather quantitative data - information dealing with numbers and anything that is measurable. Statistics, tables and graphs, are often used to present the results of these methods. Quantitative methods are now well accepted in information systems that include surveys, laboratory-based experiments and simulations, formal methods (for example, econometrics), and numerical methods (David and Jan, 2005).

**Qualitative research methods:** These were developed in social sciences to enable researchers to study social and cultural phenomena. Examples of qualitative methods are action research, case study research and ethnography (David and Jan, 2005). Qualitative researchers typically rely on the following methods for gathering information: participant
observation, field notes, reflexive journals, structured interview, analysis of documents and materials, etc (Marshall and Rossman, 1998). "Qualitative research seeks out the ‘why’, not the ‘how’ of its topic through the analysis of unstructured information…..It doesn’t just rely on statistics or numbers, which are the domain of quantitative researchers” (QSR, 2007). These methods “are useful for the exploration, requirements analysis, usability engineering and the formative and summative evaluation of information technology in health care” (Webel, et al., 2006).

Jeffery and Votta (1999) described their opinion on quantitative approach to the field of computer science/software engineering research that “At this point in time, there is no widely held collective agreed model of the definition and role of empirical (quantitative) software engineering”. Seaman (1999) supports qualitative methods in software engineering by pointing out that “The principal advantage of using qualitative methods in software engineering is that they force the researcher to delve into the complexity of the problem rather than abstract away from it. Thus, the results are richer and more informative…however, qualitative results are often considered ‘softer’ or ‘fuzzier’ in technical communities, but then so are the problems we study in software engineering”.

Indeed, there is no single or commonly agreed approach to conducting research in the field of computing, information systems, and health informatics/medical informatics (Rory, 2000).
3.3 Our Research Methodology

The research methodology presented in this thesis derives from a synthesis of qualitative and quantitative methods. Techniques, approaches, and strategies developed in this thesis are based on healthcare knowledge management, contextual information retrieval, and knowledge-based search strategies. The qualitative research efforts for this thesis were centered on analysis of clinical practice guidelines’ (CPGs) models and encoding strategies for CPGs knowledge representation, creation of heart disease ontology, generation of information-view for query customization, medical literature and CPGs knowledge linking strategy. The quantitative research efforts were centered on CPGs computerization into segments, computerized medical text analysis for extraction of medical candidates, automatic generation of query and its query type, CPGs content retrieval algorithm and indexing strategy, and the development of clinical knowledge assistance system.

This thesis addresses the designing and development of methods and techniques for a computer system that provides evidence-based clinical knowledge and information to assist healthcare practitioners for the better interpretation and understanding of clinical referral letters and to provide alert messages for critical situations at point of care.

The research problems of this thesis have been grouped into five categories that have been framed into corresponding research questions (see section 1.4). To answer the research questions of each research problem, the research methodology is grounded into five phases.
Each phase consists of steps that are performed to achieve the objective of a corresponding phase and in turn proceeds to accomplish the goal of this thesis. Figure 3.1 provides the blueprint of our research methodology. The left hand side of Figure 3.1 illustrates how the phases and their corresponding steps are linked while the right hand side of Figure 3.1 shows main methods and techniques associated with each step.

- **Phase 1:** *knowledge modeling and computerization techniques for clinical practice guidelines.*

- **Phase 2:** *search strategies to link online Evidence-based-medical (EBM) literature with computerized clinical guidelines.*

- **Phase 3:** *computerized processing of medical information (of referral and response letters) for customized focused query formulation and generating alert messages for critical situations.*

- **Phase 4:** *contextual retrieval techniques for CPGs’ content and knowledge (that are linked with EBM-literature).*

- **Phase 5:** *developing a computer system that incorporates the techniques and frameworks developed to provide clinical knowledge assistance.*

We describe each phase, its objective and corresponding steps in the following sections.
3.3.1 Knowledge Modeling and Computerization

The objective of this phase is to design a framework for knowledge modeling of the clinical practice guidelines to transform them into concise yet, focused segments, which are computer interpretable and enriched with additional meta-information. This phase consists of two steps:

(i) CPGs Knowledge Representation

(ii) CPGs Computerization into Segments

Details of the techniques, strategies and algorithms associated with corresponding steps of this phase are given in chapter 4. This phase resulted in a Knowledge Modeling Framework for Clinical Practice Guidelines Computerization.

3.3.1.1 CPGs Knowledge Representation

We have devised an ‘Encoding strategy’ to encode CPGs knowledge using GEM model. In this step, we used our Encoding strategy to transform CPGs content into GEM-encoded CPGs.

3.3.1.2 CPGs Computerization into Segments

To transform, CPGs content into concise yet focused computerized segments, we have extended guideline element model “GEM”. The extension of GEM model is performed by creating Extended-Component Ontology (Ex-KC-O). The Ex-KC-O extends the GEM model at Knowledge Component level. We transform CPGs into computer-interpretable format by creating Extended-Knowledge Components (Ex-KC). In this step, we have developed a method to transform GEM-encoded CPGs into Extended-Knowledge Components (Ex-KCs) using Ex-KC-O and some additional
modules (see section 4.3 for details). We have also developed heuristics to identify and standardize medical terms, concepts and phrases from CPGs content using UMLS thesauri (i) SNOMEDCT and (ii) MeSH. The transformed Ex-KCs are concise computerized segments of CPGs, which are enriched with meta-information like, contextual impact factor, semantic type, UMLS score, vocabulary source etc.

3.3.2 EBM-Literature Linking Strategy

The objective of this phase is to develop a framework that links the relevant evidence-based medical literature from online EBM knowledge sources to computerized CPGs segments, which are represented with Ex-KCs. This phase consists of two steps:

(i) Key-medical terms and Query type selection

(ii) Linking CPG with EBM literature

Details of the techniques, strategies and algorithms associated with corresponding steps of this phase are given in chapter 5. This phase resulted in a Context-specific Query Generation Framework.

3.3.2.1 Key-medical Terms and Query Type Selection

To perform this step, we developed a method that identifies significant medical terms based on their contextual importance and semantic types and determines key-medical terms to generate a query.
Research Methodology (Phases and Steps)

Phase 1: KNOWLEDGE MODELING & COMPUTERIZATION
- CPGs Knowledge Representation
- CPGs Computerization into Segments

Phase 2: EBM-LITERATURE LINKING STRATEGY
- Key-medical terms and Query type Selection
- Linking CPG to EBM literature

Phase 3: MEDICAL INFORMATION PROCESSING
- Extraction of Medical Candidates from Referral Letter
- Classification of Medical Candidates
- Information-View for Query Customization

Phase 4: CPGs CONTENT RETRIEVAL
- Computerized CPGs Indexing Strategy
- Content & Knowledge Retrieval

Phase 5: CLINICAL KNOWLEDGE ASSISTANCE
- Clinical Knowledge Assistance Info-structure

Methods and Techniques
- Analysis of CPGs models
- Encoding Strategy
- Ex-KC Ontology-based Computerization
- Filtration Techniques
- Heart Diseases Ontology
- Search Strategy
- Natural Language Processing Method
- Semantic Types Categorization
- Semantic Relations Analysis
- Multi-level Classification
- Automatic Query Generation
- Ex-KCs Content and Structure Analysis
- Contextual Retrieval Algorithm
- System Architecture
- Internet/Client-Server
- Extensible Markup Language (XML)
- Programming Languages (Java EE)

Figure 3.1 Research Methodology. UMLS = Unified Medical Language System, CPG = Clinical Practice Guidelines, Ex-KC = Extended-Knowledge Component, EBM literature = Evidence-based medical literature
This method determines key-medical terms by using, a heart disease ontology, a technique for quantification of semantic relations of medical terms, a weighting scheme, a semantic type filtration strategy, and a strategy for common term analysis. We have developed a technique to find the query type of generated query based on the semantic types of key-medical terms in a query. A query type specifies the scenario or intention of a query e.g. prognosis, therapy etc. The query and query type are generated for each Ex-KC.

3.3.2.3 Linking CPG with EBM-Literature

We are using MEDLINE as a source of EBM literature. In this step, we developed a heuristic that uses the E-utilities at PubMed to query MEDLINE using generated query for Ex-KC. The retrieved relevant EBM literature is linked to corresponding EX-KC.

3.3.3 Medical Information Processing

The objective of this phase is to develop a framework that processes information in medical referral/response letter to create an information-view. The purpose of the information-view for a referral/response letter is to assist healthcare practitioners in formulating an optimal customized clinical query that reflects their information needs. This phase consists of three steps:

(i) Extraction of Medical Candidates from Referral Letter

(ii) Classification of Medical Candidates

(iii) Information-View for Query Customization.
Details of the techniques, strategies and algorithms associated with corresponding steps of this phase are given in chapter 6. The output of this phase is an *Automatic Medical Information Processing Framework*.

**3.3.3.1 Extraction of Medical Candidates from Referral Letters**

Referral letter could be in any form of free text document, in Microsoft Word, PDF (Portable Document Format) or other text format. In our case, referral letters are in Microsoft Word format. We have developed a heuristic to extract text from a letter. To identify and extract medical candidates from the text, a method has been developed using the existing natural language tools and UMLS thesauri.

**3.3.3.2 Classification of Medical Candidates**

To perform this step, we developed a strategy for the classification of medical candidates. This strategy is based on our technique of a redundancy filter, a semantic type filter, a semantic categories scheme, and a heart diseases ontology. In performing this step we also developed a heuristic for negation detection using the existing algorithm NegEx (Chapman, 2003).

**3.3.3.3 Information-view for Customized Query**

For this step, we devised a presentation scheme of analyzed medical information extracted from referral letter that is called “Information-view”. This presentation scheme includes:

(a) Classification of medical information into five general categories:

- symptoms,
- diseases,
• diagnostic procedures,
• therapies,
• medications,

(b) Negated medical terms along with their sentences,

(c) Alert messages for critical situations,

(d) Classification of medical information into five cardiology categories:
• heart disease symptoms,
• heart diseases,
• diagnostic procedure for heart diseases,
• heart diseases therapies,
• heart diseases medications,

(e) Proposed automatic query terms.

All these information are presented to help formulate a focused customized query so that relevant information from evidenced-based knowledge source(s) could be accessed.

3.3.4 CPGs Content Retrieval

The purpose of this phase is to retrieve relevant C-CPG segments (Ex-KCs, representing CPG content) along with the linked EBM-literature, given the information needs as a query. This phase consists of two steps:

(i) Computerized-CPGs Indexing Strategy
(ii) Content and knowledge Retrieval

Details of the techniques and algorithms associated with corresponding steps of this phase are given in chapter 7. This phase resulted in Content and Knowledge Retrieval Framework.

3.3.4.1 Computerized CPGs Indexing Strategy

In creating an Ex-KC knowledge base, we have developed an effective computerized CPGs indexing strategy to index Ex-KCs structure and content. The purpose of the indexing strategy is to assist an efficient retrieval with high recall and high precision.

3.3.4.2 Content and Knowledge Retrieval

For this step, we have developed a method to retrieve semi-structured segments of CPGs content (Ex-KCs) for the information specifications, in form of clinical query. The aim of this method is (i) to retrieve CPGs content, which are contextually relevant to the clinical query and (ii) to achieve high recall and high precision of Ex-KC retrieval. In pursuing this aim, we have developed a “retrieval algorithm” that uses contextual, semantics, structural and meta-information contained in Ex-KCs to retrieve relevant CPGs segments and the corresponding linked medical literature.

3.3.5 Clinical Knowledge Assistance

The objective of this phase is to develop a computer system that uses the methods and techniques developed in above four phases to provide evidence-based clinical knowledge and information assistance. The purpose of the computer system is to assist healthcare practitioners for better interpretation and understanding of clinical referral
letters and to provide alert messages for critical situations at point of care. This phase has only one step that is performed to achieve the objective of this phase. Details of this phase are presented in chapter 7.

3.3.5.1 Clinical Knowledge Assistance (CKA) Info-Structure

We have designed a system architecture that specifies the functional flow of the frameworks and modules and defines inter-frameworks communication and coordination, to develop a system for Clinical Knowledge Assistance (CKA). We have developed a healthcare-knowledge base that contains Ex-KCs and linked medical literature from MEDLINE and related indices. We used web-based client/server platform to implement a system prototype.

3.4 Summary

We have presented our research methodology that is based on a mix of qualitative and quantitative research approaches. In our research methodology, we divide the research into phases based on the research problems, objectives and tasks. Each phase has certain steps, which are performed by using existing and developing new techniques, methods, and heuristics. The techniques, approaches and methods are based on healthcare knowledge management approaches, contextual information retrieval techniques, and knowledge-based search strategies.
4.1 Introduction

Knowledge is regarded an intensive asset of enterprises, in particular for those which are knowledge centric such as healthcare. The quality of knowledge and best practices based on evidences are the measures of the effective operationalization of knowledge assets. On the other hand contextually sensitive and semantically related evidence-based knowledge play an important role in decision-making at point of care. Clinical practice guidelines (CPGs) are being developed to bring evidence-based practices at point of care. It has been argued that efficient computerization of the CPGs is needed to make effective use of such knowledge source. This chapter deals with the computerization of clinical practice guidelines to transform them into computer-interpretable format. We first present an overview of CPGs computerization approaches and revisit associated problem with corresponding research question. Secondly, we describe and explain our CPGs computerization framework and its objectives. In the next section functional flow of the CPGs computerization framework is elaborated, its different modules and techniques are explained in following sub-sections. We proceed further to discuss engineering techniques and algorithms for the implementation of CPGs computerization framework.
Clinical practice guidelines are developed to provide rich source of up-to-date knowledge of evidence-based best clinical practices. The main objective of such knowledge source is to assist healthcare practitioners in specific clinical circumstances at their decision points for better healthcare. In essence, “clinical-practice-guidelines (CPGs) are originally textual documents, usually structured as a set of clinical situations, for which evidence-based therapeutic recommendations are provided” (Georg, et al., 2003). Studies have shown that simply creating and distributing CPGs is not effective in practice of healthcare practitioners or changes in health outcomes of their patients (Kosecoff, et al., 1987; Lomas, et al., 1989).

In order to make effective use of CPGs at point of care, there is a consensus to transform clinical practice guidelines into computer-interpretable format so as to deploy them with clinical information systems (Sonnenberg and Hagerty, 2006). Two types of approaches exist to transform CPGs in computer-interpretable format: (i) ‘Model-centric representation’ or ‘Knowledge-base-centric formalism’ and (ii) Document-centric formalism (Shiffman, et al., 2004 a; Sonnenberg and Hagerty, 2006; Stephen, 2005).

4.2.1 Knowledge-Base-Centric Formalism

‘Knowledge-base-centric formalism’ or ‘Model-Centric representation’ requires a formal representation methodology to encode the guideline text in a knowledge base (Shankar, et al., 2001). “This involves the gradual conversion of a conventional narrative guideline formulated by domain experts into a compact conceptual model that is machine interpretable. However, the "text-to-model relationship is indirect and the mapping is
often inexact. Nevertheless, the conceptual model is semantically close to the operational model” (Stephen, 2005). The dominant modeling approach in knowledge-base-centric formalism is primitive-based modeling. The primitives are ‘steps’, ‘tasks’, ‘plans’, ‘actions’, ‘decisions’ that describe the control structure of the guideline. Most of the approaches that fall into this category have defined a language that entirely describes the representation through a formal syntax. For example, Arden syntax, Proforma and Asbru use BNF( representation language) (Asbru with its latest version is also available in XML-format), GLIF uses UML, and EON depends on the internal syntax of Protégé (Clercq, et al., 2004; Stephen, 2005).

A knowledge engineer or computer scientist is required to encode the guideline text using complex expressions of the representation language. Clinical information systems are designed based on designers’ understanding of the guideline authors’ goals of intentions and their conceptualization of guideline logic. This makes it very difficult for clinicians or healthcare practitioners to directly encode CPGs into computer-interpretable format (Shiffman, et al., 2004a). A ‘run-time engine’ or an execution engine is required to execute the encoded-CPGs using this formalism. Two types of approaches are taken for execution engine (i) ‘event-based approach and (ii) ‘rule-based approach’ (David and Antonio, 2008). The guidelines execution engine executes the logics that are defined by using the different primitives of the formalism-model. In essence, decision-rules are triggered once the condition is met or an event is occurred. There are some complex steps that have to be taken while defining the rules or events, e.g. defining the variables that represent the decision variables, actions and other related information for defining decision logics in terms of AND, OR logical operators etc.
4.2.2 Document-Centric formalism

In ‘document-centric formalism’ guideline document is used as the authentic source of domain knowledge to be the reference repository of the guideline specifications (Shiffman, et al., 2004a; Sonnenberg and Hagerty, 2006). In these approaches, original guideline document is marked-up or annotated to create more structured, with semantics, elements of the model. This process is performed in different stages e.g. marking up, identifying elements, assigning these elements to specific semantic tags etc. This approach can help alleviate some of the difficulties in encoding the guidelines into computer-interpretable format (Sonnenberg and Hagerty, 2006). This formalism provides the possibility of CPGs transformation into computer-interpretable format (at least at the earlier stages) by operators e.g. clinicians, healthcare practitioners who are not skilled in programming (Sonnenberg and Hagerty, 2006). The approaches that correspond to this formalism differ from one another by the ‘defined formal structure’ that they impose on a document (Shankar, et al., 2001). Examples of some of the models of such approaches are GEM (Shiffman, et al., 2001), ActiveGuidelines (Tang and Young, 2000), HGML (Hagerty, et al., 2000).

“The best-known document-centric approach is GEM” (Sonnenberg and Hagerty, 2006) that takes original narrative guideline text and with the help of GEM Cutter II (a special-purpose editor) text segments from CPG are copied into guideline elements. These elements have some semantics and are structured in hierarchal manner. The result of this transformation is an XML document containing guideline excerpts. Shiffman, et al. (2004) proposed steps required for the computerization of CPGs by document-centric
approach (GEM model). The steps are grouped in two phases: Translation of guideline knowledge and Workflow integration. These phases are described in next sections.

### 4.2.2.1 Translation of guideline knowledge

Translation of guideline knowledge refers to a process to “translate the knowledge contained in guideline text into a computable format” (Shiffman, et al., 2004 a). The steps for this phase are given below:

**Guideline Selection**: This is a process that is based on user or organizational imperatives. To select appropriate guidelines, some of the following activities need to be done. For instance, “identified areas in which there is exceptional local practice variation, areas in which new knowledge ought to be put into practice, or areas in which resource use is inappropriate” (Shiffman, et al., 2004 a).

**Markup**: This step is performed after guideline selection and it is the only step specific to an XML modeling approach. In this step, guideline knowledge components and related information needed for computerization are identified and tagged or annotated.

**Atomization**: “Atomization is the process of extracting and refining single concepts from the recommendation’s natural language text. It involves removing unnecessary words, changing verb phrases from passive to active voice, reducing decision variables to prototypic nouns with descriptors occupying the (value) element and stating actions and directives as verbs in active voice with associated direct and indirect objects and modifiers” (Shiffman, et al., 2004 a).
**Deabstraction:** This is the process in which level of abstraction, at which decision variable and action is described, is adjusted to ease the computerization.

**Disambiguation:** “Disambiguation is the process of establishing a single semantic interpretation for a recommendation statement. Ambiguity can be introduced when values of decision variables are not mutually exclusive” (Shiffman, et al., 2004 a).

**Verification of Completeness:** This is a process that assures that all logically possible combinations of condition states are addressed that provide guidance in all situations that a clinician is likely to face. However, it is unnecessary for the conditional recommendations with a small number of decision variables.

**Explanation:** This is a process to describe the reasoning behind the recommendation. Normally, text extracted directly from the guideline can be used to achieve this purpose.

**Build Executable statements:** In this process, the atomized, deabstracted, and disambiguated decision variables and actions are transformed into logical statements. Logical operators can be used to express guideline recommendations into decision logic or statement logic which are best known as decision rules.

In addition to these steps, other steps have been used like *Normalization* where decision variables and actions were normalized with different IDs and representations e.g. {for ‘patient age’ = state_patient.age}, {for ‘under 60 years of age’ = INF_60} {for ‘diabetes’= DIA}, {for ‘hypertension’ = HT} etc (Georg, et al., 2003).
4.2.2.2 Workflow Integration

Workflow integration refers to a process of integrating the information (transformed in last phase) into clinical workflow. Shiffman, et al. (2004a) regarded the integration of guideline’s knowledge with clinical workflow as a critical activity for successfully bridging the guideline implementation gap. Following are the steps that may be taken for workflow integration during CPGs computerization.

**Identify Origins and Insertions:** “The implementer must identify a source or origin in the clinical environment for each decision variable and an insertion point in the care process for each recommended action and directive.” (Shiffman, et al., 2004a).

**Define Action-Type:**

Shiffman, et al. (2004a) divided actions types into four categories: (i) gathering information, (ii) interpreting information, (iii) performing a task and (iv) arranging for or organizing additional care. It has been proposed that guideline-recommended activities should be categorized to these predefined action types that can help implementers for computerization.

**Associated Beneficial Services:** “Action types can be linked to associated beneficial services that offer design patterns for facilitating clinical care. It then remains for the implementer to instantiate specific components appropriate for local circumstances” (Shiffman, et al., 2004a).

**Choose Interface Components:** “Design elements for the user interface must be selected and grouped for optimal usability” (Shiffman, et al., 2004a).
**Requirements Specification:** “The output of the above processes is a requirement specification that can be operationalized by information systems personnel. Such individuals often have high levels of expertise in programming local systems, but have more limited knowledge of clinical domains and informatics skills. The document serves as a starting point for an iterative development process” (Shiffman, et al., 2004 a).

Following, we briefly present a comparison between Knowledge-base-centric and Document-centric formalisms. Knowledge-base-centric formalisms use complex knowledge models to support guideline conversion into computer interpretable format (Shankar, et al., 2001). These formalisms normally have a compact conceptual model that is machine interpretable, however, during conversion process text-to-model relationship is not direct that causes often inexact mapping (Stephen, 2005). These formalisms use complex expression of the representation language that requires a knowledge engineer to encode the text based on their understanding of the conceptualization of guideline logics (Shiffman, et al., 2004 a). On the other hand, Document-centric formalisms, guideline are marked up or annotated to produce semi-structured format. Such approaches represent the closest representation of original guidelines (Sonnenberg and Hagerty, 2006). These approaches offer possibility to transform guidelines at least at early stages by clinicians (Sonnenberg and Hagerty, 2006). Document-centric formalism is considered less complex than knowledge-base-centric formalism in terms of conversion process. However all the approaches require custom-developed execution engine to access the encoded-CPGs knowledge. Guideline execution engine needs to be integrated with a local workflow of clinical setting (integration with patient data) to be used. Moreover, the
problem of best possible representation of an EMR has not been solved yet (David and Antonio, 2008).

A document-centric formalism, GEM (Shiffman, et al., 2000) is an abstract level model that defines the guideline elements. Computerization of CPGs using GEM depends on steps of encoding scheme and nature of execution engine. One of the identified issues with GEM-based computerization is that it requires exhaustive manual process to accomplish this task (Patrick, 2007). Additionally, complex and guidelines-specific encoding schemes and frameworks are difficult to be followed by clinicians who are involved in computerization process (Georg, et al., 2003). Existing computerization approaches based on GEM model (see section 2.3) use execution engine to provide encoded-guideline knowledge at point of care. This requires guideline knowledge to be represented in logical statements defining the decisions-rules by programming expert. Representing guidelines knowledge in decision logics to be accessed using execution engine and patient data integration is a very common approach. In taking this approach some information related to CPGs content are not considered and CPGs content are not enriched with additional information (Hashmi and Zrimec, 2008 a). For example, information related to CPGs content like, semantic correlation, contextual importance, meta-information, information provided by XML-structure (if transformed into XML format e.g. GEM). Furthermore, using standardized medical vocabulary during computerization of CPGs process has its critical implications, in particular for sharable guidelines (Sonnenberg and Hagerty, 2006). For instance, UMLS thesauri can be used during computerization process to standardize medical terms, concepts and phrases, CPG content can be enriched with meta-information such as semantic types, UMLS score.
Vocabulary source etc. Exploiting such information to provide contextually relevant CPGs knowledge at point of care requires new approaches of computerization other than those approaches, which represent guidelines knowledge in decision logics to be used with execution engine. Additionally, in those clinical settings where input would not be from EMR or EHR, there is a need to find new methods to computerize CPGs.

In forthcoming sections, we describe a new approach of CPGs computerization in addressing a research question framed for the first research problem (see section 1.4) of this thesis. The research question is:

- How to model and represent clinical practice guidelines knowledge in order to computerize them into concise yet focused segments and enrich them with meta-information?

4.3 Proposed Clinical Practice Guideline Computerization Framework

Clinical practice guidelines are textual documents, which are developed to support evidence-based medicines by bringing the best clinical practices to healthcare practitioners. However, to make effective use of CPGs in clinical practices, transformation of CPGs into computer-interpretable format is needed (Sonnenberg and Hagerty, 2006). In our research, we are using CPGs as one of the evidence-based knowledge sources to provide better interpretation and understanding of referral letters. In this clinical setting, input to access guideline knowledge will not be from EMR or EHR patient data, which is why a new approach of knowledge modeling and computerization of CPG is developed.

The underlying rationale of our CPGs computerization approach is that:
(i) CPGs consist of recommendations of management of diseases, therapeutic procedures, diagnostic procedures, set of symptoms, etc where each recommendation, therapeutic procedure etc represents a complete activity, so CPGs should be computerized into concise yet focused segments,

(ii) Medical terms, concepts and phrases in CPGs should be standardized,

(iii) CPGs content should be enriched with related meta-information,

(iv) Knowledge modeling of CPGs should be easy and effective so that clinicians could participate in computerization process, and

(v) Automating computerization tasks as many as possible to alleviate manual tasks involved in computerization process.

We took the document-centric approach for our CPGs computerization framework to comply with the underlying rationale of our computerization approach. One of the the best document centric approaches is ‘Guideline Element Model’ (GEM) (Shiffman, et al., 2001), it supports original narrative guideline text and has a special-purpose editor “GEM Cutter” to copy text segment into guideline elements (Sonnenberg and Hagerty, 2006). So, because of its features GEM is selected as a model of our CPGs computerization. GEM is an XML-based guideline document model that can store and organize the heterogeneous information contained in practice guideline documents. It describes concepts pertinent to guideline representation, attributes of those concepts, relationship among the concepts and advocates translation of textual guidelines into a format that can be processed by computers (Georg, et al., 2003; Shiffman, et al., 2004 a). The most
important part of GEM is the “knowledge component” section. It stores and categorizes the recommendations of the CPGs to constitute the essence of practice guidelines. Details of each and every element can be found in (Yale School of Med, 2006).

GEM has several limitations. As described above, model is simply an abstraction of the guideline document that makes it dependant to extrinsic system to apply in ways that are useful. GEM does little to resolve the ambiguities presented in many guidelines (Georg, et al., 2003). Although GEM provides quite good numbers of elements but depending on the requirements, additional elements, attributes and relationships may be necessary to adequately encode guidelines. We have extended GEM at ‘Knowledge Component’ level. We amended the structure of Knowledge Component to make it more comprehensive and compatible to our requirements for CPGs computerization approach. We have added different elements to store context, semantics, automatically generated IDs and meta-information pertaining to CPGs content. GEM extension is performed by using our Extended-Knowledge Component Ontology (Ex-KC-O) (see section 4.3.4).

4.3.1 Functional Flow of Our CPG Computerization Framework

CPGs computerization framework is shown in Figure 4.1. The functionality of our framework can be divided into three phases. In first phase CPGs are marked up using the GEM II cutter tool (Yale School of Med, 2006) that is based on the GEM II schema.
The “marking up” of CPGs is performed based on our “Encoding strategy”. Applying ‘Encoding strategy’ on the CPG produces an instance of “GEM-Encoded CPG” (see Figure 4.1 Phase1). In the second phase GEM-Encoded CPG is passed to “Knowledge-Component-Extractor” (KC-Extractor) to extract each knowledge component separately (see Figure 4.1 Phase2). Extracted Knowledge Components are used by an “Extended-Knowledge-Components (Ex-KCs) Instance Creator” module to create the Ex-KCs knowledge base. This module uses ‘Ex-KC Ontology’ to define the structure of Ex-KC elements. It works with Contextual Weight (CW) Module and Elem-ID Module to add contextual weights and IDs. The contextual weights for medical terms and phrases, under different elements, are assigned by a “CW Module”. The function of an “Elem-ID Module” is to create automatic IDs specific to each element type of the knowledge component of GEM-encoded CPG.
In the third phase (see Figure 4.1), the “Ex-KCs Instance Creator” module works with UMLS-MMTX Manager to add *meta-information* to each Ex-KC and to standardize medical concepts. The “UMLS-MMTX Manager” uses three modules: (1) “sentence noun-phrase (S-P) parser”, (2) “MED-Candidates-Mapper”, and (3) “Redundancy Filter” to perform its tasks. The “UMLS-MMTX Manager” performs the following tasks: (i) parsing text into sentences and sentences into noun medical phrases, (ii) retrieving a “best-final-mapping” of medical concepts from MeSH and SNOMED (iii) adding *meta-information* for these medical concepts and (iv) filtering the redundant medical concepts.

In the following sections, we describe our encoding strategy and detailed functionalities of the main modules of CPGs computerization framework.

### 4.3.2 Encoding Strategy

Our ‘Encoding strategy’ defines the specification and rules that are to be followed during the marking-up of CPG using GEM II cutter. These rules are developed for the knowledge modeling of the CPGs to achieve the objective of our CPGs computerizations. This ‘Encoding strategy’ has been developed specifically for ‘Knowledge Component’ section of GEM model. Following are the tenets of our ‘Encoding strategy’.

We categorize the age groups for target population of guideline into six age groups. These age groups were derived from MeSH, by analyzing 13 different age groups defined in PubMed to classify query, and by consulting healthcare practitioners from cardiology. (i) infant (0 to 23 months), (ii) child (2 to 18 years), (iii) adult (19 to 64 years), (iv) elderly (65 to 79 years), (v) aged (80 years and above), and (vi) general (any age).
To mark up a *Knowledge Component* (KC) element of the GEM model, it is considered as a complete activity or a section in the guideline. This element must be annotated or tagged during a mark-up process to represent that particular activity or section in the guideline. Annotation of this element could simply be some heading or sub-heading (narrative text) from guideline that is selected using GEM cutter for the annotation. Healthcare practitioner may also add suitable medical concept or medical text for annotation of KC.

Every KC element has the *Recommendation* element(s). In our work *Recommendation* element must represent a sub-activity or sub-section within the scope of KC element’s activity. Each *Recommendation* element must be annotated to represent a sub-activity within KC. The annotation text should be taken from the guideline. It may also be annotated by healthcare practitioner by adding suitable medical concepts if required.

The *Conditional elements* are found under the *Recommendation* element. This element has many sub-elements that represent the conditional recommendation or activity in the guideline. Annotation of this element is not mandatory.

There are different sub-elements under *Conditional element*. For our ‘Encoding strategy’, the most important are *Decision Variable*, *Action*, *Reason*, *Recommendation Strength*, and *Linkage*. Other elements like *Logic*, *Cost* etc are not used, as manually defined logic statements are not required for our computerization framework. Elements like *Goal*, *Reference* may be used just to add additional information but these elements would not make any difference on our computerization framework technique.
The *Decision Variable* element contains sub-elements that are used to add more information or data about *Decision Variable*. In our work, we are using *Value* and *Decision Variable Description* sub-elements. Each *Decision Variable* element must be annotated with specific value (in text format) or medical concept within the specific recommendation scope in the guideline, e.g. ‘chest pain’ etc. The *Value* element is used to add some information about the corresponding *Decision Variable*, e.g. ‘severe or lasting soon’ etc. *Decision Variable Description* element may contain some information (in the form of narrative text) to describe corresponding *Decision Variable*. At this level, out of these three elements only *Decision Variable* element’s annotation is mandatory. The content representing the decision variable could be any narrative text that adds significance towards decision.

The *Action* element’s annotation is performed by marking-up full narrative text statement(s) from the guidelines within the recommendation scope and it is mandatory. The *Action* element has different sub-elements but for our computerization technique, only *Action Description* element is used. The *Action Description* element may be annotated to add more description about the action from guideline.

In our technique, one *Conditional* element could have one or more *Decision Variable(s)* along with one or more *Action(s)*. There is no tightly-coupled link for *Decision Variables* elements and *Action* elements. For example, *Decision Variable1* and *Decision Variable2* are for *Action2*, rather in our case, *Decision Variables* and *Actions* provide as a whole complete snap of corresponding *Conditional* element.
In our Encoding strategy every Action element should have Reason and Recommendation Strength elements. These two elements should be under the Action element so that linking of Reason and Recommendation Strength elements to corresponding Action element could easily performed automatically. In GEM model these two elements are under conditional element. To provide the virtual association between Action, Reason, and Recommendation Strength elements, the Reason, and Recommendation Strength elements must be added during marking-up for every action, whether there are any content pertaining to these elements in a guideline for annotation or not.

In GEM model Linkage element is within the scope of Conditional element. According to Encoding Strategy, the Linkage element should be within the scope of the Recommendation element. In our approach, Linkage element may contain association link to any other relevant Knowledge Component or Recommendation element of the same guideline. The link could be provided by simply annotating this element with the name (annotated text) of the associated Recommendation or Knowledge Component element. In this way, related recommendations from clinical guidelines can also be presented if a corresponding recommendation is being considered.

The Imperative element is used if the recommendation in the guideline is not conditional in nature. Complete imperative recommendation is annotated (with narrative text) under the Imperative element. The Linkage element under the Imperative element is used as described above.
In our ‘Encoding strategy’ operator (who is marking up the guideline) does not need to provide manual IDs for the above described elements because these are automatically assigned. Neither ‘Atomization’ nor ‘Normalization’ (Georg, et al., 2003; Shiffman, et al., 2004 a) is performed for any elements described above because we do not transform CPGs knowledge content into decision logics.

It is necessary during marking-up that any abbreviation used in guideline for any medical concept or phrase should not be used rather full concept or phrase should be used for annotation e.g. ‘ACS’, instead of this abbreviation, full phrase ‘Acute coronary syndrome’ must be used. Any special character or symbol like “*” etc should be eliminated from the text.

4.3.3 Working Example of CPG Modeling Using Our Encoding Strategy

In Figure 4.2, excerpt of Australian CPG for Management of acute coronary syndrome (Heart Foundation Australia, 2006) is shown with some tags representing annotated text for corresponding elements in the GEM model based on our ‘Encoding strategy’.

We present an example of marking-up process using our Encoding Strategy with reference to Figure 4.2. Mark-up is first performed for Knowledge Component element. In Figure 4.2, a knowledge component tag is representing a text to be annotated for Knowledge Component element of GEM-encoded guidelines. Knowledge Component element’s annotated text taken from guideline is “Acute management of chest pain”. Figure 4.3 shows GEM-encoded format of annotated text.
Acute management of chest pain

**Recommendation**

**Getting to hospital**

- **Decision Variable**
  - Chest discomfort at rest or for a prolonged period (more than 10 minutes, not relieved by sublingual nitrates; recurrent chest discomfort, or chest discomfort associated with syncope or acute heart failure) are considered medical emergencies. Other presentations of ACS may include back, neck, arm or epigastric pain, chest tightness, dyspnoea, diaphoresis, nausea and vomiting. Very atypical pain, including sharp and pleuritic pain, is more common in women, people with diabetes and older people.

- **Action**
  - People experiencing such symptoms should seek help promptly and activate emergency medical services to enable transport to the nearest appropriate health care facility for urgent assessment (grade D recommendation). Ideally, transport should be by ambulance. However, where ambulance response times are long, alternatives may need to be considered. Patients should be strongly discouraged from driving themselves because of the risk to other road users.

**Actions in transit**

- Aspirin (300mg) should be given unless already taken or contraindicated (grade A recommendation), and should preferably be given early (e.g., by emergency or ambulance personnel) (grade D recommendation). Oxygen should also be given (grade D recommendation).

- Glyceryl trinitrate and intravenous morphine should be given as required (grade D recommendation).

- Where appropriate, a 12-lead ECG should be taken en route and transmitted to a medical facility (grade B recommendation).

- Receiving medical facilities should be given warning of incoming patients in whom there is a high suspicion of ACS, particularly STEMI, or those whose condition is unstable (grade B recommendation).

- Where formal protocols are in place, prehospital treatment should be given, including fibrinolysis in appropriate cases (grade A recommendation). See section on management of patients with STEMI (page 187) for further discussion of prehospital fibrinolysis.

**On arrival**

- All patients presenting with suspected ACS should be subject to ongoing surveillance and have an ECG completed within 5 minutes of arrival at the medical facility (grade A recommendation). The ECG should be assessed promptly by an appropriately qualified person (grade D recommendation).

- Oxygen and pain control should be given as required (grade D recommendation).

**Key Messages**

- People experiencing symptoms of ACS should seek help promptly and activate emergency medical services.

- The most important initial requirement is access to a defibrillator to avoid early cardiac death from reversible arrhythmias.

- Aspirin should be given early (e.g., by emergency or ambulance personnel) unless already taken or contraindicated.

---

This narrative text is marked-up using GEM Cutter II (Yale School of Med, 2006).

It is represented in XML format, in Figure 4.3, after marking up. Figure 4.3 also shows another ‘Knowledge Component’ element that has been annotated from the same CPG.
According to Encoding strategy, each Recommendation element must be marked up. Its annotated text could be taken from guideline and some additional words may be added to specify the meanings. Recommendation tag in Figure 4.2 is indicating annotated text. Figure 4.4 shows the annotated text for the recommendation element where some additional words, based on guideline recommendation context, have been added by clinician.

![Figure 4.4 Marking up of Recommendation element](image)

It has been described in section 4.3.1.1 that marking up of Conditional element is not mandatory, it may and may not be marked. In Figure 4.2 “Getting to hospital” has been tagged as annotated text for conditional element. Figure 4.5 shows the Conditional element marking up process output with and without annotated text.

![Figure 4.5 Marking up of Conditional element](image)

In Figure 4.2, annotated text for Decision Variable, Value and Decision Variable Description elements have been identified. It is shown in Figure 4.6 that how simply these content from guidelines are transformed into XML format during mark up process.
It can be noted that no ‘Atomization’ or ‘Normalization’ is needed for these elements’ content.

In Figure 4.7, Action, Action Description, Recommendation Strength and Reason elements are shown with their annotated text after marking up. It can be noticed from Figure 4.2 that after identifying the “action statements” from guideline, marking up is very straightforward. It is evident in Figure 4.7 that every Action element has Recommendation Strength and Reason elements, even though if there is nothing in the guideline to be annotated for these elements. This technique makes it easy for computer-
In our technique, Linkage element is marked up to link one or more Knowledge Components or Recommendations. In Figure 4.8, it is shown that the Knowledge Component represented by “Management of patients with ST-segment-elevation myocardial infarction” is related with the Knowledge Component shown in Figure 4.3.

Figure 4.8 Mark up of ‘Linkage’ element

In this example, we have shown that how simple a guideline text is marked using our Encoding strategy to produce GEM-encoded guidelines. Examples of complete GEM-encoded knowledge components representing CPG guideline content are given in appendix A, Table A.1.

In the next section, we describe the structure and functionality of our ‘Extended-Knowledge Component Ontology’ (Ex-KC-O). This ontology is used for further processing of ‘marked-up’ CPG.

4.3.4 Extended-Knowledge Component Ontology

In our computerization technique, GEM model is extended at ‘Knowledge Component’ level. This extension is performed by using an ontology termed as “Extended-Knowledge Component Ontology” (Ex-KC-O). We have developed Ex-KC-O
according to our objective for CPGs computerization framework. The fundamental functions of this ontology are:

- to help create additional elements and their attributes within ‘Knowledge Component’ elements hierarchy
- to restructure ‘Knowledge Component’ elements in more organized-fashion that comply with our CPG computerization requirements.

A structure of Ex-KC-O and the relationships between its elements are shown in Figure 4.9. In Figure 4.9 shaded area indicates restructuring or addition of elements or attributes. Rounded-shape boxes in Figure 4.9 indicate those elements that have sub-elements (children). Following, we describe elements and their attributes in Ex-KC-O in top to bottom fashion.

The top level element of Ex-KC-O is a *Knowledge Component* (KC). It has one attribute ‘ID’ that represents unique identity of the corresponding KC within a knowledge base. A KC element has three sub-elements or child-elements:

(i) KC.name,

(ii) KC.MedTerm, and

(iii) Recommendation.

The relationship of ‘KC’ element with its sub-elements is “Has” relationship. A *KC.name* element represents “Knowledge Component name” and it stores original annotated text of *Knowledge Component* element from GEM-Encoded CPG (that has been marked-up using our ‘Encoding strategy’). It can also be noticed from Figure 4.9.
that Ex-KC-O puts the limit on \textit{KC.name} element that only one such an element is allowed in one \textit{Knowledge Component}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Extended-Knowledge Component Ontology' elements structure and relationships}
\end{figure}

The \textit{KC.MedTerm} element represents ‘Knowledge Component medical term’. This term is identified and extracted (using computerized information processing)—see section 6.3.1—from the original annotated text of \textit{Knowledge Component} element. One \textit{Knowledge Component} may have none, one or more \textit{KC.MedTerm} element(s). Each \textit{KC.MedTerm} element has six attributes pertaining to corresponding medical term. These attributes are called “meta-information”:

- CW
- ID
ST
VS
ON
Scor

$CW$ is a contextual weight of the medical term. It defines the contextual impact of a medical term within the Knowledge Component. $ID$ represents a unique identity of the medical term within a Knowledge Component. It is generated automatically based on our ID scheme (see section 4.3.5). $ST$ is a ‘semantic type’ of a medical term. $VS$ represents vocabulary source that tells which thesaurus a medical concept is coming from. $ON$ represents original medical term name that is used to find the best mapping of the medical concepts from UMLS for knowledge standardization. $Scor$ is a UMLS score, which is “the strength/confidence” of the mapping of the original phrase to the corresponding SNOMED CT term or MeSH term.

The third sub-element of Knowledge Component is Recommendation (Rec) element. This Recommendation element has one ‘ID’ attribute for its unique identification within ‘KC’. The Recommendation element has five sub-elements:

(i) Rec.name,

(ii) Rec.MedTerm,

(iii) Imperative,

(iv) Conditional,

(v) Linkage.
The *Rec.name* element stores the original annotated text of a *Recommendation* element from GEM-Encoded CPG. The *Rec.MedTerm* element represents ‘Recommendation medical term’ that has been identified and extracted from Recommendation element annotated text. It contains similar six attributes called meta-information which have been defined for the *KC.MedTerm* element.

According to the Encoding strategy, *Linkage* element stores information about related recommendation(s) or knowledge component(s) for a corresponding recommendation. It has been restructured under *Recommendation* element in the Ex-KC-O. There could be none, one or more ‘Linkage’ element(s) under ‘Recommendation’.

Recommendations in CPGs could be conditional or imperative. The *Imperative* element represents imperative recommendation and it has one attribute ‘ID’. This element represents the entire imperative statement from GEM-Encoded CPG. Each Imperative element has sub-element i.e. an *IM.MedTerm*. There could be none, one or more occurrences of this element.

The *IM.MedTerm* element represents a medical term identified and extracted from the annotated text of the *Imperative* element. It has similar six meta-information attributes as defined for the *KC.MedTerm* element’s meta-information.

The *Conditional* element represents a conditional recommendation in CPG. It has one attribute *ID* and has three sub-elements:

(i) C.MedTerm,

(ii) Decision Variable,

(iii) Action.
The *C.MedTerm* element represents a medical term identified and extracted from the annotated text of *Conditional* element. It has six similar meta-information attributes as of the *KC.MedTerm* element.

The *Decision Variable* (DV) element has one attribute *ID* and has four sub-elements. The four sub-elements are:

(i) DV.name,
(ii) DV.MedTerm,
(iii) DV.Value,
(iv) Decision Variable Description.

The *DV.name* element represents the original annotated text of *Decision Variable* from GEM-Encoded CPG. It has only one occurrence under one *Decision variable* element.

The *DV.MedTerm* element represents the medical term/concept/phrase that is the best mapping of the medical term from SNOMED CT and MeSH. It has six meta-information attributes similarly like the *KC.MedTerm* meta-information attributes. It is different in terms of its number of occurrences from the ‘KC.MedTerm’ (or the elements same in nature). There must be one or more occurrences of this element if recommendation is conditional in nature.

The *DV.Value* element represents the annotated text of *Value* element from Gem-Encoded CPG that signifies the *Decision Variable*. It has only one sub-element i.e. the *Dvalue.MedTerm*. It represents the identified and extracted medical concept from the *DV.Value* element and it has similar six meta-information attributes as of the *KC.MedTerm* element.
The *Decision Variable Description* element represents the same annotated text of Decision Variable Description from GEM-Encoded CPG.

The *Action* (AC) element has one attribute *ID* and it has five sub-elements:

(i) AC.text,

(ii) AC.MedTerm,

(iii) Action Description,

(iv) Recommendation Strength,

(v) AC.Reason.

The AC.text element represents an annotated text of *Action* element from GEM-Encoded CPG and it must have only one occurrence for each *Action* element.

The AC.MedTerm represents a identified and extracted medical term from annotated text of *Action* element. It has similar six meta-information attributes as of KC.MedTerm element.

The *Action Description* element represents the same annotated text of *Action Description* from GEM-Encoded CPG.

According to Encoding strategy, each Action must have Redenomination Strength and Reason elements. So, Recommendation Strength element and Reason elements have been re-structured under Action element. The *Recommendation Strength* element represents recommendation strength of an Action from GEM-Encoded CPG. There must be at least one Recommendation Strength element under an Action element.
The Reason element represents annotated text of reason element from GEM-Encoded CPG. The EX-KC-O makes sure that there must be at least one Reason element under Action element.

We have described the elements structured in EX-KC-O. The Ex-KC-O is used to help create the instances of ‘Extended-Knowledge Component’ (Ex-KC) from GEM-Encoded CPG. The slots for values of Ex-KCs’ elements are filled during the instantiation of Ex-KC by using different modules (as shown in Figure 4.1). In the next section we define our ID scheme that is used to assign ID to different elements (automatically) within Ex-KC.

### 4.3.5 Extended-Knowledge Component Elements ID Scheme

An *Extended-Knowledge Component* is represented in XML format, where its elements are organized in a hierarchical structure. The Ex-KC has different segments which are labeled by semantic tags derived from the GEM model and the Ex-KC Ontology (Ex-KC-O). We designed our *Element ID Scheme* based on the information provided by semantic tags and hierarchical structure of the Ex-KC. It helps keep track on elements effectively and search information within the Ex-KC efficiently. The elements’ IDs are generated automatically.

In our technique, IDs are generated only for those elements that have sub-elements and some of their sub-elements take the corresponding generated IDs. It can be noticed in Figure 4.9 that only those elements (in round-shape boxes) have been assigned ID attribute that have direct sub-elements. The reason behind is that these elements have significant importance semantically and contextually within Ex-KC whereas some of their sub-elements have been derived to support operations on their corresponding parent
elements. For example, *Decision Variable* sub-elements DV.MedTerm, DV.Value.MedTerm (see Figure 4.9) are used to support operation (like, searching, processing, etc) on Decision variables. So, these sub-elements are identified with their *Decision Variable* element’s corresponding ID. For instance, ID for a Decision Variable element is:

\[
<	ext{DecisionVariable} \text{ ID}="\text{ACS}_{-06}\_\text{b.xml}_{-}\text{KCl}_{-}\text{R1}_{-}\text{C1}_{-}\text{dv1}"\text{, so the ID of DV.MedTerm will be:}
\]

\[
<\text{DV.MedTerm} \text{ CW=}\text{"1.0" ID=}"\text{ACS}_{-06}\_\text{b.xml}_{-}\text{KCl}_{-}\text{R1}_{-}\text{C1}_{-}\text{dv1"}}\text{.}
\]

We describe our ID scheme with the following example. The Ex-KCs are derived from a GEM-Encoded CPG by applying the EX-KC-O. A GEM-Encoded CPG has unique name that is given during marking-up process and it has number of ‘Knowledge Components’ ranging from ‘1 to n’. During the process of Ex-KCs creation, GEM-Encoded CPG ‘filename’ and the number of ‘Knowledge Components’ (that is being transformed to Ex-KCs) are taken for the ‘ID’ generation. We illustrate ‘ID’ scheme with the GEM-Encoded CPG “for the management of acute coronary syndromes (ACS 06)” (Heart Foundation Australia, 2006). The file name of this GEM-Encoded CPG is “ACS_06_a.xml”. This file name will be used with all the Ex-KCs created from GEM-Encoded CPG of ‘ACS 06’. During the transformation of first Ex-KC, the elements IDs are given as below.

The ‘KC’ representing *Knowledge Component* is a top level element in an Ex-KC, which is assigned ‘ID’=“ACS_06_a_KCl”. Its sub-elements that do not have further sub-elements are given the same ‘ID’ (see table 4.1). The first ‘Recommendation” element within this EX-KC is assigned ‘ID’=“ACS_06_za.xml_KCl_R1” which is read as “first
Recommendation element of first Knowledge Component of GEM-Encoded CPG ‘ACS 06’”. Other elements are assigned ‘IDs’ same as of ‘Recommendation’ element based on their position in Ex-KC hierarchical structure.

For example “ACS_06_za.xml_KC1_R1_C1_dvl” is an ‘ID’ of a first ‘Decision Variable’ of a first ‘Conditional’ element of a first ‘Recommendation’ of a first ‘Knowledge Component’ of the GEM-Encoded CPG (ACS 06). Table 4.1 elaborates other elements of the Ex-KC at different position in a hierarchical structure. In the next section we describe the “Contextual Weight” that defines the contextual impact of different elements within Ex-KC.

Table 4.1 Elements IDs of Ex-KC1 from ‘ACS_06_za.xml’ GEM-Encoded CPG

<table>
<thead>
<tr>
<th>Element</th>
<th>ID</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>KC</td>
<td>ACS_06_za.xml_KC1</td>
<td>First ‘Knowledge Component’ from GEM-Encoded CPG</td>
</tr>
<tr>
<td>KC.MedTerm</td>
<td>ACS_06_za.xml_KC1</td>
<td>Sub-element of ‘KC’ element</td>
</tr>
<tr>
<td>Recommendation</td>
<td>ACS_06_za.xml_KC1_R1</td>
<td>First ‘Recommendation’ of KC1</td>
</tr>
<tr>
<td>Conditional</td>
<td>ACS_06_za.xml_KC1_R1_C1</td>
<td>First ‘Conditional’ element of R1</td>
</tr>
<tr>
<td>DecisionVariable</td>
<td>ACS_06_za.xml_KC1_R1_C1_dvl</td>
<td>First ‘Decision Variable’ of C1</td>
</tr>
<tr>
<td>Action</td>
<td>ACS_06_za.xml_KC1_R1_C1_A1</td>
<td>First ‘Action’ element of C1</td>
</tr>
</tbody>
</table>

4.3.6 Contextual Weight

In our work, Knowledge Components are modeled into Ex-KCs through extension of GEM. The Ex-KC has different segments that represent different contextual importance within it. In consultation with healthcare practitioners and undertaking analysis of CPG models and CPGs content, we have divided the Ex-KC structure into five segments categories based on their contextual importance. Each segment category is assigned contextual weight ($\eta$) to determine contextual impact of corresponding element and its content. A contextual weight is assigned on the scale of 0.25 to 1.00 where 1.00
being the highest contextual weight. Those elements that are considered more significant and specific in the Ex-KC are given higher weight than those which are considered relatively less significant and specific. For example, the Decision Variable element is more specific in defining the meaning and scope of its medical concepts than the Knowledge Component element. Following, we define these categories and corresponding contextual weight ($\eta$).

(i) ‘Decision Variable’ element: if a medical phrase or medical term or related info is a part of this element, it is given a contextual weight, $\eta = 1.0$, 

(ii) ‘Recommendation’ element: if a medical phrase or medical term or related info is a part of this element then it is assigned a $\eta = 0.75$, 

(iii) ‘Knowledge Component’ element: if a medical phrase or medical term or related info is a part of this element then it is given a $\eta = 0.5$, 

(iv) ‘Imperative’ element: if a medical phrase or medical term or related info is a part of this element then it is given a $\eta = 0.5$, 

(v) If a medical phrase or medical term or related info is a part of other than these four elements, it is given a $\eta = 0.25$. 

This contextual impact factor of the different elements within an Ex-KC has important implications. For example it is used in a technique for linking relevant medical literature with Ex-KCs and in the retrieval of relevant computerized CPGs segments (see chapter 5 and 7).
In the final process of CPGs computerization, annotated text of Ex-KC elements is analyzed to find medical terms, their related meta-information and to perform their standardization. In next section, we describe this process that is performed by *UMLS-MMTX Manager*.

### 4.3.7 UMLS-MMTX Manager

The ‘UMLS-MMTX Manager’ is a module that implements the following processes:

- Identification and extraction of medical concepts from the annotated text of the corresponding elements,

- Filtering out redundant medical concepts/terms/phrases,

- Standardizing medical concepts from MeSH and SNOMEDCT, and

- Retrieving meta-information related to extracted medical concepts/terms/phrases.

The UMLS-MMTX Manager receives annotated text from an *Ex-KC instance creator*, extracts medical terms, retrieves corresponding meta-information and sends them back to the *Ex-KC instance creator*. The *Ex-KC instance creator* stores the received information into corresponding elements’ slots of each Ex-KC. All created Ex-KCs are stored in the *Ex-KC knowledge base* (Ex-KC KB). We elaborate these above defined processes in details with our algorithms and engineering techniques in the following sections.
4.4 Implementation of the CPGs Computerization Framework

In this section, we describe and illustrate knowledge engineering techniques and algorithms developed for the implementation of CPGs computerization framework. We will first explain an algorithm for medical terms identification, extraction and retrieval of corresponding meta-information. Next, we describe an algorithm of redundant medical terms filtration. In further section, we discuss the engineering process for ‘Ex-KC instance creator module. The output of our CPGs computerization is presented as the final section.

4.4.1 Medical Terms Extraction and Meta-Information retrieval Method

This section describes the process of identifying and extracting medical terms from annotated text of the GEM-Encoded CPGs and retrieving meta-information. The organization of extracted and retrieved information into data suture is also elaborated.

The Ex-KC instance creator module extracts relevant annotated text of Knowledge Component elements from the GEM-Encoded CPGs for identifying and extracting medical terms and phrases. The flow chart of this process is shown in Figure 4.10. The Ex-KC instance creator module sends annotated text to the UMLS-MMTX Manager to perform medical concepts identification, extraction and meta-information retrieval. The text is converted to MMTX compatible document by using MMTXLite API (MetaMap Transfer, 2007). In the next step sentence-extraction is performed. This process identifies and extracts the number of sentences in a document (denoted by ‘S’ in Figure 4.10). For each sentence, medical phrases are identified (denoted by ‘P’ in Figure 4.10).
Figure 4.10 Flow chart of the medical terms identification, extraction, and meta-info retrieval.
The process of final best medical mapping is performed for each extracted medical phrase to find the best medical mapping candidates from the UMLS, as shown in Figure 4.10.

A decision check is applied to find out if there is any best medical mapping candidate(s). In case of “No”, next phrase is taken and sent to find its best medical mapping. In case of “Yes”, the best medical mapping candidates are retrieved for the phrase (denoted by ‘C’ in Figure 4.10). For every best-medical-mapping candidate, the following meta-information attributes are retrieved from UMLS: (i) Preferred medical concept or name, (ii) Vocabulary sources, (iii) Semantic type, (iv) UMLS score, and (v) Original name of the medical phrase. The “phrase loop” is terminated, once all the ‘best-medical-mapping’ candidates for a particular phrase are processed, and the next phrase is taken for processing. The “sentence loop” is terminated, once all the phrases for a particular sentence are processed, and a next sentence is sent to be processed. After processing all the sentences, medical phrases are sent back to the ‘Ex-KCs instance creator’.

Figure 4.11 shows the data structures used (i) to store the above retrieved information for the ‘best-medical-mapping’ candidates, (ii) to group all the ‘best-medical-mapping’ candidates of a phrase, and (iii) to group all the medical phrases of a sentence.

The ‘best-medical-mapping’ candidate with its all meta-information is stored in an object of a class “CandidateValues”, represented as <CandValues> in Figure 4.11 (Left side) and its object is represented by “Cand”. The medical phrase is conceptualized by a class “MedicalTerm” represented as <MedTerms> and its object is represented by ‘MT’
The “MedicalTerm” class has an array to store “CandidateValues” class objects that is represented by “MT.CandObj”.

![Diagram](image)

**Figure 4.11** (Left side): “CandidateValues” class structures and (Right side): Arrays for ‘MT’ and ‘MX’ objects

All retrieved medical phrases are stored in a “UMLS-MMTX” manager class that is represented by <MMtXLite> in Figure 4.11. This class stores all “MedicalTerm” class objects in the array. This array is represented by a “MX. MTarray”.

Table 4.2 illustrates the outputs (results) of ‘medical terms identification, extraction and meta-information retrieval’ processes. The annotated text from an Action element of a Knowledge Component is processed and medical terms from the annotated text are identified and extracted along with corresponding meta-information. During the process of medical terms identification and extraction using the UMLS, we have to deal with UMLS redundant behavior. Because UMLS uses more than one thesaurus, it generates redundant medical terms. In the next section, we describe this redundant behavior and the design of a filter for redundant medical terms filtration.
<table>
<thead>
<tr>
<th>Text from CPG</th>
<th>Extracted Medical Terms</th>
<th>Meta-Information Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reperfusion may be obtained with fibrinolytic therapy</td>
<td>Reperfusion Therapy</td>
<td><strong>OrgName</strong> Therapeutic Procedure or Preventive Procedure</td>
</tr>
<tr>
<td>Thrombolytic Therapy</td>
<td></td>
<td><strong>Vocabulary-sources</strong> SNOMED CT, MSH</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Semantic Type</strong> Therapeutic Procedure or Preventive Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>UMLS Score</strong> 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Contextual Weight</strong> 0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ID</strong> ACS_06_b_KC2_R1_C1_A1</td>
</tr>
</tbody>
</table>

**Table 4.2 Results from the medical Terms identification, extraction and meta-info retrieval processes**

4.4.1.1 Experiments for Determining the UMLS Redundant Behavior

Based on the nature of our work, we are using two thesauruses from UMLS, MeSH and SNOMEDCT, for the *medical term identification and extraction* process. We conducted experiments to analyze the redundant behavior of the UMLS so that an efficient filter could be developed. Experiments were conducted in three settings, one with MeSH thesaurus only, one with SNOMEDCT only and one with MESH and SNOMEDCT.

A sentence with only one medical term “pain” was submitted to our implemented framework with MeSH thesaurus. We found two ‘best final mappings candidates’:

(i) ‘Pain Clinics’ with *[semantic type: “Manufactured Object”]*

(ii) ‘Pain’ with *[semantic type: “Sign or Symptom”]*.

Other attributes’ details like UMLS score etc are omitted here because those are not important for this section. Another sentence, with only “Symptoms” as a medical concept in a sentence, was submitted to the same setting and only one ‘best final mapping candidate’ was found:
A new sentence with both medical concepts “pain” and “Symptoms” was submitted, and four mapping candidates were retrieved:

(i) ‘Pain Clinics’ with [semantic type- “Manufactured Object”],

(ii) ‘Symptoms aspect’ with [semantic type- “Functional Concept”],

(iv) ‘Pain’ with [semantic type- “Sign or Symptom”], and

(v) ‘Symptoms aspect’ with [semantic type- “Functional Concept”].

It is evident that more than one medical concept in a sentence may produce redundant information during medical term identification and extraction process that would affect the efficacy of our system.

In the second setting, we repeated the same experiment with SNOMEDCT thesaurus. For “Symptoms” medical term in a sentence, only one ‘best final mappings candidate’ was retrieved:

(i) ‘Symptoms’ with [semantic type- “Sign or Symptom”].

For “pain” medical term in a sentence, only one ‘best final mappings candidate’ was retrieved:

(i) ‘Pain’ with [semantic type- “Sign or Symptom”].

For a sentence having both medical concepts in it, the best final mapping retrieved two candidates:

(i) ‘Symptoms’ with [semantic type- “Sign or Symptom”] and

(ii) ‘Pain’ with [semantic type- “Sign or Symptom”].
We conducted the same experiment in the third setting, using MeSH and SNOMEDCT thesauruses. We submitted a sentence with both medical concepts “Pain” and “Symptoms” to the third setting that retrieved eight ‘best final mapping candidates’ out of which four were redundant medical concepts.

We conducted additional experiments with more medical concepts and discovered that the number of redundant concepts may be increased with different medical concepts using MeSH and SNOMEDCT thesauruses. As our implemented framework works automatically and no manual intervention is performed during this process, it was required to develop an algorithm to filter out the redundant medical concepts. With the above described experiments, we determined the functional requirements of that algorithm. In the next section, we explain our ‘Medical Terms Filtration’ algorithm.

4.4.2 Medical Terms Filtration Algorithm

The filtration of redundant medical concepts generated during ‘medical terms identification and extraction’ process is not straight forward. We developed a special heuristic to be used for medical terms filtration. The filtration process is performed once all the medical phrases are stored in a “UMLS-MMTX” manager class object’s array (MX. MTarray). The steps of the filtration algorithm are shown in Figure 4.12. The Figure 4.13 depicts the flow of the filtration mechanism. The steps of the filtration algorithm are described as follows.
1. Take ArrayList of `<MedTerms>` objects (MT)
2. Create three object references of `<MedTerms>`
3. Ma, Mb Medf
4. if Size of arraylist `<MedTerms>` objects > 1
5. {
6. initiate Ma = MT at index (0)
7. Mb = MT at index (1)
8. Pass both Ma and Mb to Compareobj (Ma, Mb) method that
9. return `<MedTerms>` objects (MT)
10. Initiate Medf = with returned `<MedTerms>` objects (MT)
11. for (b = 2 and b less than Size ArrayList of `<MedTerms>` objects (MT))
12. do {
13. 
14. initiaite Mb = MT at index (2)
15. pass Mb and Medf to compare object method and iniate
16. Medf object
17. Medf = Compareobj (Ma, Mb)
18. }
19. 
20. else{
21. ArrayList of `<MedTerms>` objects (MT) size = 1
22. initialize Medf = MT at (0)
23. }
24. return Medf

Figure 4.12 Medical candidates filtration algorithm

Figure 4.13 The flow of “MedicalTerm” objects filtration
All “MedicalTerm” class objects are taken from a UMLS-MMTX manager object which are stored in a temporary array. At this point the array size is measured to evaluate if the filtration process will proceed. In case of ‘Yes’, three objects from “MedicalTerm” class are created which are represented as “Ma, Mb and Medf” in Figure 4.13.

Next, ‘Ma’ object is initialized with a “MedicalTerm” (MT) object at index = 0 of the array and a ‘Mb’ object is initialized with a ‘MT’ object at index = 1 of the array.

Both objects ‘Ma’ and ‘Mb’ are passed to a “MedicalTerm” objects’ comparison method. This method takes two ‘MT objects’ and compares their ‘medical-candidates’ and returns a new “MedicalTerm” object. The resulting “MedicalTerm” object does not have similar medical candidates.

In the next step, a ‘Medf’ object is initialized with resulting “MedicalTerm” object and a ‘Mb’ object is initialized with “MedicalTerm” (MT) object at index = 2 of the array. These two objects ‘Medf’ and ‘Mb’ are now compared using “MedicalTerm” objects’ comparison method that returns new “MedicalTerm” object with non-similar medical candidates.

The resulting “MedicalTerm” class object is again used to initialize ‘Medf’ object and ‘Mb’ object is initialized with “MedicalTerm” (MT) object at index = 3 of the array. The same process is repeated until all “MedicalTerm” objects in the array are processed.
In Figure 4.14, the process of filtering the redundant medical terms is illustrated. Medical concepts from GEM-encoded CPG are processed for ‘medical term identification and extraction’. This process generates eight standardized medical concepts containing redundant medical concepts. Our filtration algorithm is applied to filter out the redundant medical terms. In the filtration process “MedicalTerm” objects are compared with each other. Each “MedicalTerm” object contains the “Candidate Values” <CandValues> objects. In next section we present our algorithm for the ‘Medical Terms Objects’ comparison.

4.4.3 Algorithm for Medical Terms Objects Comparison

The “MedicalTerm” class objects comparison algorithm is not a simple object comparison, as each object has different arrays holding objects of different types (like CandidateValues objects). So, it is required to go at low level, inside the objects, to compare two “MedicalTerm” class objects. The heuristic of this algorithm is given in Figure 4.15. Following are the steps of this algorithm.
1. Take two <MedTerms> objects M1, ans M2
2. Create <CandValues> Object CandRM
3. Create boolean flag variable 4. for all <CandValues> objects in M1.CandObj arraylist
5. (m = 0 run till m is less than M1.CandObj array size)
6. do 1 run another loop7. for all <CandValues> objects in M2.CandObj arraylist
7. (n = 0 run till n is less than M2.CandObj array size)
8. do 1
9. Send <CandValues> objects at index M1.CandObj(m) and at index M2.CandObj(n) to CompCand method
10. set flag = CompCand(M1.CandObjat(m), M2.CandObjat(n))
11. if ( flag = true) 1
12. Remove <CandValues> object from index M2.CandObj(n)
13. } }}
14. if M2.CandObj arraylist of <CandValues> objects is not empty
15. { for(i = o i till i is less than M2.CandObj array size)
16. do
17. add <CandValues> object index M2.Candobj(i) to
18. <CandValues> objects in arraylist of M1.Candobj object 1}
19. return M1 <MedTerms>

Figure 4.15 “MedicalTerm” objects comparison algorithm

- Two “MedicalTerm” class objects, represented by ‘M1’ and ‘M2’, (Figure 4.15) are taken to be compared.

- Two loops are initiated to compare all the “CandidateValues” class objects contained in a ‘M1’ and a ‘M2’.

- The First loop (outer-loop-steps: 4-15, Figure 4.15) runs through all the “CandidateValues” objects in ‘M1’ and second loop (inner loop-steps: 7-15, Figure 4.15) runs through all the “CandidateValues” objects in ‘M2’.

- The first “CandidateValues” object in ‘M1’ is compared against all “CandidateValues” objects in ‘M2’. If the first “CandidateValues” object in ‘M1’ exists in ‘M2’, it is removed. The Index then moves to the second “CandidateValues” object in ‘M1’ and it is compared against all remaining
“CandidateValues” objects in ‘M2’. This process continues until all “CandidateValues” objects in ‘M1’ are compared.

- At the end of the comparison process, all remaining “CandidateValues” objects in ‘M2’ are added to ‘M1’. The resulting ‘M1’ has all unique “CandidateValues” class objects.

The functional flow of this process is depicted in Figure 4.16. The comparison of two “CandidateValues” class objects is performed at attribute level to compare attributes of two “CandidateValues” objects. Each attribute of a “CandidateValues” object is matched against corresponding “CandidateValues” object attribute to find, if both are same or not. Figure 4.17 explains this process. In next section we describe the engineering process and algorithms for the ‘Ex-KC instance creator’ module.

```
'MedicalTerms' objects comparison Process

Figure 4.16 The flow of “MedicalTerm” objects comparison process
```
4.4.4 Engineering process of the Extended-Knowledge Component Instance Creator

The “Ex-KC instance creator” module performs different tasks. One of the main tasks is to create the Extended-Knowledge Component (Ex-KC) instances according to the Extended-knowledge Component Ontology (Ex-KC-O). The challenges in performing its tasks are (i) to read GEM-Encoded CPG and extract only those elements, which are required according to Ex-KC Ontology, (ii) to process the content of ‘Knowledge Component’ elements, to identify and extract medical terms, and to retrieve related meta-information, (iii) to create structure of the Ex-KCs and (iv) to store all the information and data into Ex-KCs structure.

To cope up with these challenges, we represented all necessary elements in ‘objects of different class types’ and group these objects according to the Ex-KC Ontology. Figure 4.18 shows the Ex-KC instance knowledge representation at object level. We, describe the data structure for the elements of an Ex-KC instance in following steps.
The ‘Knowledge Component’ element is represented by a <KC> class object. It contains two objects: (i) a “Recommendation” element object and (ii) an “Other data” object.

The “Recommendation” element is represented by a <RC> class object. The “Other data” object contains a “MedicalTerm” class object, a UMLS score, a Semantic Type, and the other meta-info.

The <RC> class object contains: (i) a “Conditional” element object, (ii) an “Imperative” element object, (iii) a “Linkage” element object, and (iv) an “Other data” object.
• The “Imperative” element is represented by an <IMP> class object. The “Linkage” element is represented by a <Link> class object. The Conditional” element object is represented by a <Cond> class object.

• The <Cond> class object contains three objects: (i) a “Decision variable” element object (ii) an Action element object and (iii) an “Other data” object. The “Decision variable” element object is represented by a <DV> class object and an Action element object is represented by a <Action> class object.

• The <DV> class object contains: (i) a “DecisionValue” (Dval) object, (ii) a “Description” (Descrp) object, and (iii) an “Other data” object.

• The <Action> class object contains: (i) a “Recommendation strength” (RecST) Object, (ii) a “Reason” (Resn) object, (iii) a “Description” (Descrp) object, and (iv) an “Other data” object.

In Figure 4.19 the heuristic of the Ex-KC instance creator algorithm is presented. The “Ex-KC instance creator” module accepts the GEM-encoded CPG file. It processes only those elements and their related information which have been specified in our ‘Encoding Strategy’. First, the “Knowledge Component” element is processed (see Figure 4.20). Upon receiving this event, signal flag is set to ‘true’ so that the data related to the “Knowledge Component” element could be retrieved. In the next step, it creates a <KC> class object and stores data in this object.
1. Read GEM-Encoded XML file
2. Catch events
3. if (knowledgeComponent) then
   do {set its flag for characters data true and create KC object
   add data to this object. Send data to Medical Terms processor
   add results from Processors to KC object and set flag false}
4. if (Recommendation) then
   do{ Set its flag true, Create RC object get character add data to
   RC object, send it to Medical term processor add result to RC object
   and add this RC object in KC object arraylist of type <RC>}
5. if (conditional) then do{ repeat above process for conditional and
   add 'Cond object' <Conditional> to RC object}
6. Repeat above process for all the events defined in Ex-KC-Ontology and
7. Structure (add) event-based objects according
to Ex-KC-Ontology defined-objects relationships
8. Return KC (knowledgeComponent) object that has
   all the defined-objects and info

Figure 4.19 Ex-KC instance creator algorithm

This data from the "Knowledge Component" element is sent to the 'UMLS-MTX Manager'. The results received from the 'UMLS-MTX Manager' are added to this <KC> class object and a signal flag is set to 'false'. The next element, Ex-KC instance creator retrieves from GEM-Encoded CPG file is the "Recommendation" element and the same process is repeated. At the end, its corresponding object i.e. <RC> class object is added to the <KC> class object. The next element to be retrieved is either 'Conditional' or 'Imperative'. Both elements are processed with the same procedure as described above and their corresponding class objects, which are, <IMP> class object and <Cond> class object are added to the <RC> class object. All other elements defined in the Extended-Knowledge Component Ontology (Ex-KC-O) are processed and structured in the form of objects as shown in Figure 4.19.

In our technique one complete 'Knowledge Component' block is processed in a way that all elements and their corresponding information reside in one object i.e. <KC> class object. In the same manner all "Knowledge Component" elements in a GEM-
encoded file are processed and stored in <KC> class objects. At the end, the Ex-KC instance creator retrieves <KC> class objects one by one and creates “Ex-KC-instances” according to the Ex-KC Ontology. It uses the “Contextual Weight” module to assign ‘contextual weight’ and “Elem-ID” module to assign ‘element id’ for each element in the “Ex-KC-instances”.

These newly transformed Ex-KCs constitute the ‘Ex-KC knowledge base’ that represents the computerized CPGs. We have shown in Figure 4.2 (section 4.3.3) an excerpt from Australian CPG for Management of acute coronary syndrome that is marked up using our ‘Encoding strategy’ and the output of this process is shown in Figure 4.20. Once Clinical practice guideline is marked up it is processed to create ‘Extended-Knowledge Components’ (Ex-KCs). The result of our CPGs computerization technique is shown in the Figure 4.21. The Figure 4.21 shows excerpt of the CPG segment that is computer-interpretable.
Acute management of chest pain

Recommendation
Acute management of chest pain to go to hospital

<DecisionVariable source="nd">chest pain
  <Value source="nd">at rest or for a prolonged period</Value>
  <DecisionVariableDescription source="nd">at rest or for a prolonged period (more than 10 minutes, not relieved by sublingual nitrates)</DecisionVariableDescription>
</DecisionVariable>

<DecisionVariable source="nd">syncope
  <Value source="nd"></Value>
  <DecisionVariableDescription source="explicit">Chest discomfort at rest or for a prolonged period (more than 10 minutes, not relieved by sublingual nitrates), recurrent chest discomfort, or discomfort associated with syncope or acute heart failure are considered medical emergencies.</DecisionVariableDescription>
</DecisionVariable>

Action
People experiencing such symptoms should seek help promptly and activate emergency medical services to enable transport to the nearest appropriate health care facility for urgent assessment.

ActionDescription
Ideally, transport should be by ambulance. However, where ambulance response times are long, alternatives may need to be considered. Patients should be strongly discouraged from driving themselves because of the risk to other road users.

Reason

RecommendationStrength grade D recommendation

Figure 4.20 Excerpt of GEM-Encoded CPG Knowledge Component
In this chapter, we provided the approaches of our CPGs Computerization. After explaining the underlying rationale of the CPGs computerization, we explained its functional flow. The CPGs computerization framework consists of different modules and
techniques such as ‘Encoding strategy, Extended-Knowledge Component Ontology, Extended-Knowledge Component instance creator, Contextual Weight module, Element ID module, UMLS-MMTX Manager. We discussed and explained the techniques behind the functionalities of these modules. We put forward the details of the ‘Encoding strategy’ that was illustrated with a working example. Outputs of the different modules during the processing of CPGs computerization were shown and explained. The implementation of our CPGs computerization framework was discussed with our engineering techniques and algorithms developed specifically for this purpose. At the end, output of the clinical practice guideline modeled in Extended-Knowledge Component (Ex-KC) was shown and illustrated i.e. computer-interpretable.
CHAPTER 5

LINKING MEDICAL LITERATURE WITH COMPUTERIZED CPGs SEGMENTS

5.1 Introduction

With the omnipresence of internet, a large number of online medical knowledge sources are available to provide the insights and evidences about the medical and clinical research. Healthcare practitioners quite often seek online medical literature to find relevant evidences and related information pertaining to their information needs to support their decisions at point of care. In order to seek relevant medical literature, they have to formulate a focused and context-specific query that not only describes the problem at hand but also reflect their intentions (Ely, et al., 2005; Olena, 2005). It has been found that clinicians face many problems while seeking relevant medical literature to meet their information needs (Covell, et al., 1985; Ely, et al., 1999; Gorman and Helfand, 1995; Gorman, et al., 1994; Hersh, et al., 1996; Timpka, et al., 1989; Wanda and Henry, 2000). Such problems demand to find efficient ways to help fulfill healthcare practitioners’ information needs (Chambliss and Conley, 1996; Dina and Jimmy, 2007; Ely, et al., 2005).

In this chapter, we present a “Context Specific Query Generation Framework” (CQGF). The CQGF automatically generates query from segments of computerized clinical practice guidelines (Ex-KCs) and determines its query type to reflect the
This chapter is structured as follows. In section 5.2, we present an overview of clinical query models and revisit associated problem with corresponding research question. In section 5.3, we describe, in details, the algorithm, techniques and functionalities of different modules of the CQGF framework. We explain the CQGF framework process flow and outputs of its different modules by working example in section 5.4. An overview of implementation specification is put forward in section 5.5. Section 5.6 summarizes this chapter.

5.2 An Overview of Clinical Query Models, Query Types and Related-Problems

With the paradigm shift of evidence-based clinical practice, clinicians/healthcare practitioners are more inclined to make use of the clinical practice guidelines (CPGs). During their practice of the CPGs, they tend to search relevant online medical literature to clarify the ambiguities present in the CPGs, to validate the CPGs, to supplement their understanding of the CPGs content, to acquire insights into diversity of opinions, past clinical trials, clinical evidence pertaining to specific recommendation/procedure/observation of the CPGs (Abidi, et al., 2005 a; Dina and Jimmy, 2007). There are many hurdles in seeking such evidence at point of care such as exponentially expanding medical literature, unawareness of online literature search facilities, time constraints, inexperience and inability to formulate the right search query that can result in overwhelming number of documents, etc (Abidi, et al., 2005 b; Dina and Jimmy, 2007; Gorman and Helfand, 1995; Gorman, et al., 1994; Wanda and Henry, 2000). There is a
need to find efficient ways to address these problems in order to fulfill healthcare practitioners’ information needs and clinical inquires at point of care (Chambliss and Conley, 1996; Dina and Jimmy, 2007; Ely, et al., 2005).

In order to minimize such hurdles, several studies have investigated clinicians’ information seeking behavior and their information needs. Based on their finding different medical query models have been proposed to formulate clinical query and to represent its intentions (Chambliss and Conley, 1996; Cimino, et al., 1993 a; Covell, et al., 1985; Ely, et al., 1999; Ely, et al., 2000; Haynes, et al., 1990; Mendonça and Cimino, 2000; Timpka, et al., 1989). Following, we present a brief description of these studies and methods.

Cimino et al. (1993 a) developed a set of general-purpose questions called generic queries. User could select generic queries in one of two ways: (i) user may type in question that is analyzed to identify most relevant generic query or (ii) user may indicate patient data of interest, from a admission profile of patient record system, and then pick one of several potentially relevant questions. The rationale of this approach is that clinician’s needs are matched to one of a set of general query types for which retrieval strategies have been developed in advance (Cimino, et al., 1993 a). Cimino and Barnett (1993 c) presented a method to facilitate knowledge extraction that was based on laborious and time consuming tasks of executing searches and analyzing their results (Mendonça and Cimino, 2000). Zing and Cimino (1998) proposed a method that uses the co-occurrence of medical concepts as a basis of determining search query.

Haynes et al. (1994) proposed a method to develop optimal MEDLINE search strategies for the retrieval of sound clinical studies of the four query types: etiology,
prognosis, diagnosis, treatment. Their method has been incorporated at PubMed as built-in clinical query search filters. Along the same lines of the work (Haynes, et al., 1994; Wilczynski, et al., 2001), Mendonca and Cimino (2001) studied MEDLINE MeSH terms associated with the four basic clinical tasks: diagnosis, etiology, prognosis, therapy. Their goal was to automatically categorize citations for task-specific retrieval.

SUMSearch is a system that combines meta-searching and contingency searching in order to automate searching for the medical evidences. It provides additional search filters like ‘physical findings’, ‘Adverse treatment effects’, ‘Screening /prevention’ (SUMSearch, 2000). Wanda and Henry (2000) proposed a query model for a system “QueryCat”. Their model consists of ten query types. In their work clinical queries are categorized into their defined query types by lexical and Semantic analysis techniques.

Price et al. (2002) developed a system ‘SmartQuery’ that provides context-sensitive query from ‘Electronic Patient Record’ to relevant medical knowledge sources. Their work resembles the MEDLINE button and subsequent Infobuttons developed by (Cimino, 1996; Cimino, et al., 1997; Cimino, et al., 1992), in which they used generic queries whereby Price et al. (2002) tried to elicit the terms that described clinical situation.

Sibanda et al. (2006) proposed a statistical semantic category recognizer to identify eight semantic categories in discharge summaries. Abidi et al. (2005 a) described a method to generate query from ‘Computerized-CPG text’ selected by medical practitioner, to retrieve medical literature from PubMed. Zhenyu and Wesley (2007) described knowledge-based query expansion method. This method uses UMLS
Knowledge source to enhance the original query with additional terms that are specifically relevant to query’s scenario(s).

In the methods and techniques described above, concepts like scenario specific query, query types, tasks oriented query, query templates, and query categories have almost similar meaning. Recent studies reveal that almost 60% of clinicians query center on specific query types or scenarios (Ely, et al., 1999; Ely, et al., 2000; Haynes, et al., 1990; Hersh, et al., 1996; Zhenyu and Wesley, 2007). Cimino et al. (2002) showed that clinicians’ questions are predictable that center on specific tasks. It indicates that focused and context-sensitive medical query either from healthcare practitioner or system-generated should appropriately frame context, problem and clinical intention. It can be done by using a list of specialized problem-specific medical-terms/concepts as the search query (Abidi, et al., 2005 b; Olena, 2005).

There is an ongoing research to find efficient ways to fulfill healthcare practitioners’ information needs and clinical inquires at point of care (Chambliss and Conley, 1996; Dina and Jimmy, 2007; Ely, et al., 2005). It has been investigated and shown (Abidi, et al., 2005 a) 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 (Abidi, et al., 2005 a; Demner-Fushman, et al., 2008; Seol, et al., 2004). 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 (Abidi, et al., 2005 b; Hashmi, et al., 2009 a). Consequently, automatically linking online medical literature
to the knowledge content of CPGs would provide potential benefits to healthcare practitioners.

Automatically linking current clinical evidences to concise segments of computerized clinical practice guidelines imposes certain research issues that need to be answered. The research problem pertaining to automatically linking clinical evidences with concise segments of computerized clinical practice guidelines is described in section 1.4.1 and has been formulated in the following research question:

- How to link relevant current medical evidences with corresponding segments of computerized clinical practice guidelines?

In next section, we address the above defined research question by presenting our “Context Specific Query Generation Framework” (CQGF). We elaborate on the rationale, the solution design, and the functional flow of “Context Specific Query Generation Framework”.

5.3 CQGF: Context Specific Clinical Query Generation Framework

To address our research question, we have devised a Context Specific Query Generation Framework (CQGF) that is based on our technique to retrieve and link relevant medical literature with ‘Extended-Knowledge Components’. This framework automatically generates queries from refined medical terms and phrases derived from the computerized CPGs content (Hashmi, et al., 2009 d; Hashmi, et al., 2009 c) and determines its query type. 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 with the corresponding segments of computerized CPGs.

The underlying rationale of our approach is that:

1. Retrieved literature should be focused and relevant to Extended-Knowledge Components (Ex-KCs), since Ex-KCs are highly focused for specific clinical problems/procedures.

2. The search query is objectively and automatically derived from the content of computerized CPGs (i.e. Ex-KCs).

3. Ex-KC syntactic structure (elements tags), contextual, semantic and other meta-information should be exploited to formulate more focused search query.

4. As most of the clinicians’ search queries center on specific scenarios or query types, a query model that captures the scope and intentions of Ex-KC should be used.

5. Relevant medical literature should automatically be retrieved through online Webservises using system-generated focused query and linked to corresponding Ex-KCs in the Ex-KCs Knowledge Base.

To achieve the above defined tenets of our approach, we have developed a technique that (a) extracts potentially important medical terms and phrases based on their contextual importance (b) uses different filters to filter out insignificant medical terms (that have less impact in searching the medical literature within the context), (c) exploits the semantic types to find association between medical terms and phrases, (d) categorizes
search query based on the priori defined clinical query intention (defined in query model). The resulting context-specific query is used to query MEDLINE using “webservice” provided by PubMed’s E-utilities. After retrieving relevant medical articles, they are linked to the corresponding Ex-KC in the Ex-KCs Knowledge Base. In the following section we describe our algorithm that performs above defined tasks.

5.3.1 Context Specific Clinical Query Generation Technique

The technique of context specific query generation from Extended-Knowledge Components is shown with flowcharts in Figures 5.1 (a, b, and c). Following, we elaborate the steps that are involved in specific tasks. In Figure 5.1 a, potentially significant medical concepts (medical terms and phrases) are extracted from an Ex-KC. These medical concepts are passed through redundancy filter to filter out redundant medical concepts. Remaining medical concepts are evaluated based on their significance of semantic types. This is done by applying a semantic type filter to remove those medical concepts that are semantically less significant. The resulting medical concepts are then analyzed and additional meta-information (such as Term frequency, Term Total weight, etc) is attached to each concept.

In the next step, disease ontology filter is applied on each potentially significant medical concept to further specify its meaning and to remove the contextual-noise. The resulting medical concepts from a disease ontology filter are processed to find Semantic Relation Score (SMR score) (see section 5.3.2.6) based on their semantic types.
Figure 5.1 (a) Flowchart of search query generation from Ex-KC

1. Extraction of significant medical terms from Ex-KC
2. Redundancy Filter
3. Redundant Term
4. Retrieve Semantic types of medical terms
5. Semantic Type Filter
6. Filterable Semantic Type
7. Remove Terms
8. Term Frequency analysis
9. Total Weight Calculation
10. Term to Object Conversion
11. Adding meta-info to Term Object
12. Calculate SMR score of Terms
13. Set of Terms from KC-element of Ex-KC
14. Total number of Filtered Terms \leq \text{Threshold}
15. Final Query Set
16. Proceed to Query Type Finder
17. Proceed to Common Terms Analysis
18. Create set of Terms sorted by SMR score
19. Create set of Terms sorted by Total Weight
20. Diseases Ontology Filter
21. Yes
22. Term Found
23. Remove Terms
24. No
Create set of Terms sorted by SMR score

Create set of Terms sorted by Total Weight

Set of Terms from KC-element of Ex-KC

Common Terms analysis for highest value of Total Weight and SMR score along with KC-element terms

Yes

Final Query Set

Number of final set Terms = Threshold

No

Move the Common terms Found to potentially final set. Remove the common Terms from the input sets

Change the Threshold of SMR score and Total Weight to 50% of their highest value

Common Terms analysis for Threshold of 50% of the highest value of Total Weight and SMR score along with KC-element terms

Yes

Final Query Set

Number of final set Terms = Threshold

No

Change the Threshold of SMR score and Total Weight to 25% of their highest value

Move the Common terms Found to potentially final set. Remove the common Terms from the input sets

Common Terms analysis for Threshold of 25% of the highest value of Total Weight and SMR score along with KC-element terms

Yes

Final Query Set

Number of final set Terms = Threshold

No

Complete the number of remaining terms in final set based on terms Total weight and KC-element Terms

Final Query Set
In the next step, three sets of potentially significant medical concepts are created.

- Total Weight set (TW set),
- SMR set,
- KC set

**Total Weight set (TW set)**—In this set all the filtered potentially significant medical concepts are sorted based on their total weight (see section 5.3.2.4 for total weight calculation). **SMR set**—This set has the same medical concepts as in TW set; the only difference is that medical concepts are sorted based on *Semantic Relation Score (SMR score)*. **KC set**—It has those medical concepts which are found under the *KC element* of Ex-KC.

To formulate the Final query set, a threshold is applied that is used to determine the number of medical concepts for a Final query set. If the number of potentially significant medical concepts is less than or equal to the threshold then all concepts are members of the Final query set. The medical concepts are used to find the query type (see section 5.3.2.8). If the number of medical concepts is greater than the threshold then a Common Term Analysis is applied to the above defined three sets.

Figure 5.1 b 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 are identified (see section 5.3.2.7).
The threshold values are based on the highest value of the Total Weight and the highest value of the SMR score of the concepts in TW set and SMR set. The goal here is to find a concept that is common in at least two sets and has highest total weight score and SMR score. This medical concept becomes a member of the Final query set. The Common Term Analysis is continued until the number of concepts in the Final query set reaches the threshold. Once the Final query set is created the medical concepts are used to find the query type (see Figure 5.1 c).

Having presented the steps required to perform the tasks of our technique, we describe, in next section, the functional flow of the CQGF framework and functionalities of its modules in details.

5.3.2 Functional Flow of CQGF Framework

In this section we describe the functional flow of our CQGF framework and details of its each module. The functional-overview of CQGF is shown in Figure 5.2. It has been described in chapter 4 that after computerization, CPGs are transformed into ‘Extended-
Knowledge Components’ (Ex-KCs) that constitute the ‘Ex-KC knowledge base’. Each Ex-KC is taken one by one to be processed by CQGF framework so that ‘Context-specific’ query from an Ex-KC can be generated. The resulting query is submitted to retrieve the relevant medical articles pertaining to the Ex-KC. These medical articles are linked with the corresponding Ex-KC.

5.3.2.1 Medical Terms Extractor

The ‘Medical Terms Extractor’ module extracts medical concepts from an Ex-KC. This module extracts only those medical concepts from the Ex-KC that have specific “contextual impact” within the Ex-KC. The significant contextual segments are (i) KC element, (ii) Recommendation element, (iii) Decision Variable element, (iv) Imperative element, and (v) Action element. The medical concepts within these segments in the Ex-KC have relatively high impact factor than other elements. So, these have more potential to be representative of an Ex-KC’s query. The ‘Medical Terms Extractor’ sends the extracted medical concepts to a ‘Redundancy Filter’.

5.3.2.2 Redundancy Filter

There is a possibility of redundancy in a set of extracted medical concepts due to their belonging to different segments of the Ex-KC. It is necessary to remove all the redundant medical concepts from the potential set of a search query. The ‘Redundancy Filter’ removes all the redundant medical concepts. The set of remaining filtered medical concept is sent to ‘S.T Filter Module’.
5.3.2.3 Semantic Type (S.T) Filter

The Semantic Type Filter module filters out those medical concepts whose semantic types are not considered significant enough, e.g. ‘Reptile’, ‘Intellectual Product’, ‘Professional or Occupational Group’ etc (Mendonça and Cimino, 2000). This module retrieves the ULMS semantic types for all the medical concepts. Then it uses a list of “permissible” semantic types to select those medical concepts that are relevant for the domain. We have generated a list of semantic types relevant for Cardiology after consulting specialists in cardiology domain and after a qualitative analysis of the usage of these semantic types. With additional experiments conducted using clinical practice guidelines (CPGs) for heart disease we find out which semantic types best represent CPGs (see Appendix C). More semantic-types may be added based on the requirements.
5.3.2.4 Info-Manager

The ‘Info-Manger’ module performs following tasks. It adds related information to the potentially significant medical concepts. For each medical concept the following is added: ‘Term Frequency’, ‘Total Contextual Weight’ ‘Total weight’ and ‘Semantic Type’.

‘Term Frequency’ is calculated by a frequency analysis of a medical concept to find the number of times this medical concept appeared in the corresponding Ex-KC. After frequency analysis a ‘Total Contextual Weight’ of each medical concept is calculated using the Contextual Weight.

Every element in an Ex-KC refers to a special context e.g. Recommendation, Decision Variable etc. Such contexts of different elements have been quantified based on their impact factor into contextual weight represented as $\eta$ (see section 4.3.6). The medical concept may appear in different elements within the Ex-KC, so, the aggregation of its Contextual Weights is calculated. The Total Contextual Weight of any medical concept is a summation of Contextual Weights of the medical concept within an Ex-KC that is defined in equation (5.1). In equation 5.1, $T_{\eta_{mp, a_i}}$ is a “Total Contextual Weight” of a medical concept (mp) in ‘ith’ number of an Ex-KC ($\alpha$). The “$\eta_{mp, a_i}^{CTS}$” is defined as the Contextual Weight of ‘jth’ mp in ‘ith’ number of an Ex-KC ($\alpha$) at context-structure (CTS). ‘N’ represents the number of time that mp is found in $\alpha$.

$$T_{\eta_{mp, a_i}} = \sum_{j=1}^{N} \eta_{mp, a_i}^{CTS}$$  \hspace{1cm} (5.1)
After calculating a Total Contextual Weight of a medical concept, a Total Weight of a medical concept is calculated. The Total Weight of a medical concept is calculated as a summation of two factors (a) Term-Frequency and (b) Total Contextual Weight. If Total Weight is represented by ($\psi$) and Term Frequency is represented by ($\lambda$) then Total Weight can be formally defined in equation 5.2 where $T\eta_{mp,a}$ is a Total Contextual Weight of a medical concept as defined in the equation 5.1 (section 5.4 presents a working example that also illustrates these calculations with actual numbers).

$$\psi = \lambda + T\eta_{mp,a}$$

(5.2)

5.3.2.5 Diseases Ontology Filter

‘Diseases Ontology’ is a domain knowledge source that provides relationship and information about different entities such as diseases, symptoms, etc. Such domain knowledge plays important role in identifying the essence of different medical concepts. We have developed a prototype of diseases ontology for our framework that is shown in Figure 5.3. Diseases Ontology could be for any domain in general or for specific domain in particular.

Since our working domain, in healthcare setting, is cardiology, we have developed a “Heart Diseases Ontology” (HDO). The HDO contains heart diseases their symptoms, therapies, diagnostic procedures, and associated medications. It also defines the relationship among these entities. The structure and relationships among the entities in HDO are shown in Figure 5.3. Instances of HDO entities are connected through dotted- arrows in Figure 5.3. In HDO, we use ‘Has’ relationship between diseases, symptoms, therapies, and diagnostic procedures. For example, *st segment elevation myocardial
Infraction has symptoms—chest pain, arm pain etc. A disease, in HDO, has an “associated-with” relationship with medications. We used ICD10, SNOMEDCT, MeSH and the information provided by cardiology specialists to develop HDO (see Appendix B). The HDO prototype has been implemented using relational database approach in MYSQL. All different entities (diseases, symptoms, etc) are defined in separate tables. Each table of corresponding entity keeps standardized instances and their semantic types. Primary key and foreign keys are used to define the relation between diseases table and other tables. SQL queries have been created to make sense out of relations (“Has” and “Associated-with”) between the entities.

![Figure 5.3 Heart Diseases Ontology structure and relationship](image)

The function of HDO-Filter is to specify the meaning of the medical concepts and to remove the contextual-noise. The “HDO-Filter” module uses HDO and filters out those medical concepts that do not conform to ontology.
5.3.2.6 Semantic Relation Score

The UMLS provides the relation between different semantic types. We use "associated-with" relation (see Appendix D) that determines whether one semantic type is associated with other semantic type. As every medical concept belongs to a semantic type, so, by analyzing ‘associated with’ relation of the semantic types, we determine whether the medical concept is associated-with ‘none’, ‘one’, or ‘many’ medical concepts. The rationale behind this is “the more associated a ‘medical concept’ is with other medical concepts, the more potential it has to be included in a Final query set”.

We quantify the associated-with relation by giving the score of “1”, if a medical concept is found ‘associated-with’ other medical concept. The score of “0” is given if ‘associated-with’ relation is not found with other medical concept. It is formally represented by equation 5.3, where ‘associated with relation’ between two medical concepts (mp, and mpj) in Ex-KC (α) is represented by (SMRRel = ξ).

\[
SMR_{Rel} = \begin{cases} 
1 & \text{if } mp_{i}^{a} \text{ (associated - with) } mp_{j}^{a} \\
0 & \text{if } mp_{i}^{a} \text{ (Not associated - with)} mp_{j}^{a} 
\end{cases} 
\]

\[
SMR_{score} = \sum_{i=1}^{N} \xi
\]  

The summation of “SMR_{Rel} = ξ” score for a medical concept is called as Semantic Relation score (SMRscore) defined by equation 5.4. In equation 5.4, ‘N’ represents the total number of medical concepts in a query set. The value of ‘SMRscore’ for a particular medical concept could be within the range of [0 to (n-1)] where ‘n’ is the total
number of medical concepts. The ‘SMRscore Module’ calculates the ‘SMRscore’ for each medical concept.

5.3.2.7 Final query set Creator

The main function of the ‘Final query set Creator’ (FQSC) is to finalize the medical concepts for the search query that will be used to query MEDLINE database. It has been investigated (Abidi, et al., 2005b) that search query should include only small numbers of highly significant medical terms. A set of large number of medical concepts tend to make query too specific that normally results in with no medical articles from knowledge source. We conducted experiments in querying MEDLINE with several search queries, different in length of medical terms and phrases to find optimum number of medical concepts for a search query. Based on our experiments results, we concluded that five is the maximum numbers of medical concepts in a query and we use it as a threshold for Final query set (THFQS).

Upon receiving medical concepts, Final query set Creator (FQSC) module creates three sets of medical concepts: (i) one set of the received medical concepts sorted by Total Weight (TW set), (ii) one set of the received medical concepts sorted by ‘SMRscore’ (SMR set), and (iii) one set of medical concepts found under ‘KC’ element of Ex-KC (KC set). The rationale behind the third set is that ‘KC’ element within an Ex-KC represents Ex-KC title. So, the medical concepts under a KC element have some significance. If a KC element medical concept is found in a TW set or a SMR set, it enhances the strength of the corresponding medical concept to be the member of a Final query set. Medical concepts for a Final query set are determined as follows.
Final query set Creator evaluates the number of received filtered medical concepts against the threshold for Final query set (THFQS). If the number of received medical concepts is less than the threshold value \( \text{th} = 5 \), it tries to find those medical concepts from a ‘KC set’, which do not appear in the \textit{TW set} or the \textit{SMR set}. If such medical concepts are found, these medical concepts are used to complete the number of medical concepts of the Final query set otherwise \textit{Final query set} contains the received elements.

The selection of medical concepts from \textit{KC set} is based on the ‘Total weight’ (assigned by info-manager—section 5.3.2.4). The resulting Final query set is sent to \textit{Query Type Finder}.

In case, if number of received medical concepts is more than the threshold for Final query set, then Common Term Analysis is applied. The \textit{Common Term Analysis} is used to determine which medical concept will be included in the \textit{Final query set}. In Common Term Analysis three different threshold values (\textit{Common Terms Threshold (CTH) values}) are used: 100%, 50%, and 25% of the highest value in \textit{TW set} and \textit{SMR set}. All concepts that satisfy the condition of the first threshold are taken for \textit{Final query set}. If more concepts are needed, then second and third threshold are applied. \textit{Note that if medical concepts found common in ‘SMR set’ and ‘FW set’ but not in KC set (within the range of CTH) then these are also taken to the Final query set.}

If number of Final query set medical concepts is less than the threshold for Final query set, FQSC completes the remaining number of medical concepts based on the \textit{Total Weight}.

5.3.2.8 \textit{Query Type Finder}
Different models have been proposed to categorize the intentions of clinicians for their information needs (see section 5.1). In our work, we are using PubMed to find relevant medical literature. PubMed has adopted the query model proposed by (Haynes, et al., 1994; Haynes, et al., 2005) for the clinician query classification. In line with the query classification at PubMed, we categorize 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 predication or forecast of a course of disease or condition, (iv) Therapy: literature related to therapy, prevention, or rehabilitation. In case, the search query falls into more than two categories or does not fall into any above defined category, it is assigned “General” type.

We categorize a search query automatically into above defined five categories. Figure 5.1 (c) shows the process for determining a query type. A set of semantic types is assigned to each query type. Each semantic type in the set is a triggered semantic type for the corresponding query type. So, we have four sets of semantic types, each corresponds to their query type. In consultation with cardiologists following the example provided in literature (Bodenreide and McCray, 2003; Haynes, et al., 1994; Long, 2005; Sibanda, et al., 2006; Zhenyu and Wesley, 2007), we derived these sets of semantic types. Each medical concept is associated at least one semantic type in UMLS. So, if the medical concept belongs to a semantic type that is a part of one of the four query types’ sets, it implies that the medical concept is associated to the corresponding query type. The query type for a search query is determined by finding the maximum numbers of medical concepts that belong to a specific query type. In case of having same number of medical concepts belonging to two query types, both query types are assigned (Abidi, et al., 2005
a) to the final search query. If more than two query types are found or no query type is found then final search query is assigned “General” query type. “Query Type Finder” performs the above described task and sends the query type to “Final query set Creator” module (see Figure 5.1 c).

5.3.2.9 PubMed Webservice Connector

The “PubMed Webservice Module” serves as a channel to connect to PubMed to send search query and to receive the retrieved medical articles. It uses a Pubmed E-utilities feature that is AXIS 2 Webservice. It complies and implements the specifications defined by PubMed. This module transforms a final search query into the format defined by PubMed and implements our heuristic to retrieve and link medical literature to corresponding Ex-KCs. In case, if medical literature is not retrieved for the submitted query, this module removes the least significant medical concept from the query based on the Total Weight and SMRscore values and re-submits the new query to PubMed. This process continues until medical literature is retrieved or query terms are finished.

In the following section, we present a working example of the CQGF framework processing that will elaborate its functionalities.

5.4 Working Example of CQGF Framework

In this section, through a working example, we present the CQGF framework functionalities. We take a segment of knowledge content from the clinical practice guideline (CPG) “for the management of acute coronary syndromes 2006” (ACS 06) (Heart Foundation Australia, 2006). This segment from ACS 06 CPG is shown in Figure 5.4 that is entitled “Reperfusion Therapy”. This segment of CPG presents the procedures
that can be used to perform reperfusion therapy for the management of patients with ST-segment-elevation myocardial infarction. This segment of CPG knowledge is processed to be computerized by our CPG computerization framework (see chapter 4) that transforms it into Ex-KC. Figure 5.5 shows the excerpt of “Reperfusion Therapy” Ex-KC.

Management of patients with ST-segment-elevation myocardial infarction

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 5.4 Segment of content from ‘ACS 06’ CPG entitled reperfusion therapy
Reperfusion Therapy

Methodology

ST segment elevation

Percutaneous coronary intervention

Assay of fibrinolysis

Figure 5.5 Excerpt of reperfusion therapy Ex-KC
CQGF takes "Reperfusion Therapy Ex-KC" to find the potentially significant medical concepts as described in section 5.3.2.1. The extracted medical concepts are listed in Table 5.1. The extracted medical concepts are processed by Redundancy Filter to filter out the redundant concepts (as described in section 5.3.2.2). The remaining medical terms and phrases are listed along with their semantic types in Table 5.2. It can be seen that all redundant concepts have been filtered out. These medical phrases and terms are sent to Semantic Type Filter that filters out the insignificant concepts based on the semantic types (as described in section 5.3.2.3). The resulting filtered medical concepts from Semantic Type Filter are shown in Table 5.3, which are sent to ‘Info-Manager’ to add related information (see section 5.3.2.4 for description). The number of potentially significant concepts is now down to ‘11’.

Table 5.1 Potentially significant extracted medical phrases and medical terms

<table>
<thead>
<tr>
<th>Medical Phrase/Terms</th>
<th>Medical Phrase/Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>methodology</td>
<td>operative surgical procedures</td>
</tr>
<tr>
<td>procedures</td>
<td>patients</td>
</tr>
<tr>
<td>reperfusion therapy</td>
<td>percutaneous coronary intervention</td>
</tr>
<tr>
<td>methodology</td>
<td>protein c inhibitor</td>
</tr>
<tr>
<td>myocardial infarction</td>
<td>reperfusion therapy</td>
</tr>
<tr>
<td>procedures</td>
<td>science of anatomy</td>
</tr>
<tr>
<td>reperfusion therapy</td>
<td>surgery specialty</td>
</tr>
<tr>
<td>st segment elevation</td>
<td>surgical aspects</td>
</tr>
<tr>
<td>anatomy</td>
<td>thrombolytic therapy</td>
</tr>
<tr>
<td>anatomy aspects</td>
<td>transplantation</td>
</tr>
<tr>
<td>assay of fibrinolysis</td>
<td>transplanted tissue</td>
</tr>
<tr>
<td>combined</td>
<td>using</td>
</tr>
<tr>
<td>coronary artery bypass surgery</td>
<td>utilization qualifier</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>graft material</td>
</tr>
<tr>
<td>percutaneous coronary intervention</td>
<td>obtained</td>
</tr>
</tbody>
</table>

These concepts are further processed by ‘Diseases Ontology Filter’ to filter out relatively insignificant medical terms and phrases. After Diseases Ontology filtration,
seven medical phrases are left which are shown along with their Total weight and Total Contextual Weight in Table 5.4.

Table 5.2 Remaining medical phrases and terms after ‘Redundancy Filter’

<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>Intellectual Product</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>

These medical concepts are sent to SMRscore module to analyze and calculate semantic relation score for each medical concept (as described in section 5.3.2.6). These medical concepts along with their SMRscore are listed in Table 5.5. In Table 5.5 medical concept “fibrinolysis” has SMRscore “0”, because, its semantic type is “Physiologic Function” that does not have associated-with relation with Therapeutic or Preventive Procedure, Disease or Syndrome, and Finding semantic types in UMLS. The remaining
total number of medical terms and phrases for final query are greater than the pre-defined THFQS threshold that is ‘5’. So, three sets \textit{TW set}, \textit{SMR set} and \textit{KC set} are created and \textit{Common Term Analysis} is performed (as described in section 5.3.2.7). \textit{KC set} medical terms are “reperfusion therapy, methodology and procedures”. Table 5.6 shows the \textit{KC set} medical concepts and their Total Weight.

After \textit{Common Term Analysis} five medical terms and phrases are selected for the Final query set. These medical terms and phrases are:

\textit{[reperfusion therapy, percutaneous coronary intervention, myocardial infarction, st segment elevation, thrombolytic therapy]}

<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>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>myocardial infarction</td>
<td>Disease or Syndrome</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>procedures</td>
<td>Health Care Activity</td>
</tr>
<tr>
<td>st segment elevation</td>
<td>Finding</td>
</tr>
</tbody>
</table>

Query type for the ‘Final query set’ is determined (as described in section 5.3.2.6). The query type for this query is “Therapy”. So, the final query submitted to the PubMed is \textit{[reperfusion therapy, percutaneous coronary intervention, myocardial infarction, st segment elevation, thrombolytic therapy]} with query type “Therapy”. The medical articles retrieved and linked to \textit{Reperfusion Therapy Ex-KC} are shown in Table 5.7.
Table 5.4 Resulting medical terms/phrases from Diseases Ontology Filter

<table>
<thead>
<tr>
<th>Semantic Type</th>
<th>TCW</th>
<th>Medical Phrase/Term</th>
<th>TW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapeutic or Preventive Procedure</td>
<td>1.75</td>
<td>reperfusion therapy</td>
<td>4.75</td>
</tr>
<tr>
<td>Therapeutic or Preventive Procedure</td>
<td>0.50</td>
<td>percutaneous coronary intervention</td>
<td>2.5</td>
</tr>
<tr>
<td>Finding</td>
<td>0.75</td>
<td>st segment elevation</td>
<td>1.75</td>
</tr>
<tr>
<td>Disease or Syndrome</td>
<td>0.75</td>
<td>myocardial infarction</td>
<td>1.75</td>
</tr>
<tr>
<td>Therapeutic or Preventive Procedure</td>
<td>0.50</td>
<td>thrombolytic therapy</td>
<td>1.5</td>
</tr>
<tr>
<td>Physiologic Function</td>
<td>0.50</td>
<td>fibrinolysis</td>
<td>1.5</td>
</tr>
<tr>
<td>Therapeutic or Preventive Procedure</td>
<td>0.50</td>
<td>coronary artery bypass surgery</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 5.5 Medical concepts with SMR score value

<table>
<thead>
<tr>
<th>Medical Phrase/Term</th>
<th>SMR score</th>
</tr>
</thead>
<tbody>
<tr>
<td>myocardial infarction</td>
<td>5.0</td>
</tr>
<tr>
<td>reperfusion therapy</td>
<td>1</td>
</tr>
<tr>
<td>st segment elevation</td>
<td>1</td>
</tr>
<tr>
<td>thrombolytic therapy</td>
<td>1</td>
</tr>
<tr>
<td>percutaneous coronary intervention</td>
<td>1</td>
</tr>
<tr>
<td>coronary artery bypass surgery</td>
<td>1</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.6 KC set medical concepts

<table>
<thead>
<tr>
<th>KC set Phrase/Term</th>
<th>TW</th>
</tr>
</thead>
<tbody>
<tr>
<td>reperfusion therapy</td>
<td>4.75</td>
</tr>
<tr>
<td>methodology</td>
<td>3.25</td>
</tr>
<tr>
<td>procedures</td>
<td>1.25</td>
</tr>
</tbody>
</table>
## 5.5 Implementation Specifications

In this section we provide an overview of the implementation specifications of the CQGF framework. The CQGF framework is implemented into web application using java 1.5 jdk. Every module of CQGF has been implemented in java classes. Diseases
ontology has been implemented in XML. We use SAX parser to extract information from the Ex-KCs and 'Disease ontology'. We use AXIS 2 (webservice) client to connect to PubMed through java servlet. We are using MYSQL 5.0 to store the Ex-KCs and its retrieved relevant medical articles. The medical articles for the Ex-KCs are retrieved and linked upon instantiation of Ex-KCs from our CPGs computerization framework. The retrieval and linking of the medical articles can also be done at real time usage of the application but it will delay the process, as some of the factors for delay are the limitation of time and usage of E-utilities, imposed by the PubMed E-utilities.

5.6 Summary

Given the significance of linking online medical literature with clinical practice guidelines to assist healthcare practitioners at their decision points, we have described the functionalities and techniques in detail for our “Context Specific Query Generation Framework” (CQGF). The CQGF generates clinical query objectively from the content of an Ex-KC and determines its query type. It submits the query to PubMed through ‘webservice’ to query and retrieve relevant article. These retrieved articles are linked to their corresponding Ex-KC.

We provided an overview of existing clinical query models that have been proposed to reflect clinicians’ intentions for their information needs. We defined the steps of our algorithm to objectively generate a clinical query and a query type in the form of flowcharts.

We have presented the architecture of the CQGF and explained its functional flow. The detailed functionalities of each module have been explained and illustrated. A
working example to illustrate the processing of the CQGF and the outputs of its different modules has been presented. The working example takes the Ex-KC for ‘Reperfusion Therapy’ segment from the clinical practice guideline (CPG) “for the management of acute coronary syndromes”. It illustrates the functionality of each ‘CQGF modules’ and shows the corresponding outputs. Once the query is generated and a query type is determined, it is sent to PubMed. The retrieved articles relevant to the Ex-KC are presented. These articles are stored together with the Ex-KC.
Medical information pertaining to patients is normally represented in narrative text. Different forms of patient medical documents like, discharge summaries, electronic medical record, medical reports, etc, contain important information about patients. Diverse applications require automatic processing of important information in medical documents, such as information retrieval (Jones and Staveley, 1999; Li, et al., 2004), browsing interfaces in digital libraries (Gutwin, et al., 1999), vocabulary construction (Kosovac, et al., 2000), summarizing medical article (National Library of Medicine, 2006), indexing medical literature (Clifford, et al., 2005), document classification and clustering (Jonse and Mahoui, 2000), Knowledge mapping (Shepherd, et al., 2006), knowledge discovery, text mining (Ng and Wong, 1999; Ono, et al., 2001; Tanabe, et al., 1999) and many more. Computerized processing of medical documents requires challenging tasks, such as, extraction of medical facts, deriving their meaning and relationships, classification to further specifying meaning etc (Krauthammer and Nenadic, 2004; Quanzhi and Yi-Fang, 2006). Techniques and methods from natural language processing, knowledge management, text mining etc are being investigated to automatically process medical information.
We present an “Automatic Medical Information Processing Framework” (AMIPF) to help facilitate formulation of focused customized query from referral letter to access the evidence-based clinical knowledge. In this chapter, we discuss our method and technique for AMIPF in detail. We first present an overview of medical information processing and revisit associated problem with corresponding research question. In next section, we define the underlying rationale of AMIPF. We describe and discuss the detailed functionalities and technique of each module of AMIPF. Proceeding to next section, we present a case study in a form of working example to comprehensively illustrate the computerized-processing of referral letter by AMIPF. Towards the end, we summarize this chapter by highlighting the key points.

6.2 An Overview of Medical Information processing

In medical text and information processing applications, concept identification is very crucial. Concepts are normally represented by noun phrases that make identification of noun phrases one of the fundamental problems for applications in mining medical documents (Quanzhi and Yi-Fang, 2006). In the noun phrases, a critical information is called keyphrases and identification of keyphrases has been challenging research problem, since keyphrases are more domain-oriented and selective as opposed to noun phrases (Quanzhi and Yi-Fang, 2006). In medical documents noun phrases and keyphrases contain important information that may require key processes in computerized information processing like recognition, extraction, semantics-processing, mapping and classification, which are challenging tasks (Krauthammer and Nenadic, 2004).
Various techniques have been proposed for medical information processing for terms, phrases, and key-phrase detection. They can be broadly grouped into four types of approaches: Dictionary-based, Rule-based, Machine-learning and statistical, and Hybrid (Krauthammer and Nenadic, 2004).

Dictionary-based methods normally use existing vocabularies to locate terms in the text. However studies have demonstrated that many terms may not be recognised if a straightforward dictionary look-up technique is employed (Gaizauskas, et al., 2000; Hirschman, et al., 2002; Krauthammer and Nenadic, 2004; Tuason, et al., 2004).

Rule-based approaches normally try to identify terms by re-establishing associated term formation patterns (Collier, et al., 1999). In these approaches rules are developed to describe the common naming structures for certain term classes using either orthographic or lexical information. In practice these techniques are extremely time-consuming and rules are generally inflexible as in essence they are normally specific (Krauthammer and Nenadic, 2004).

Statistical approaches are usually deployed for recognition of general terms, i.e. keywords (Andrade and Valencia, 1998). Machine learning approaches are normally developed for specific entities and are integrated with term recognition and term classification (Sibanda, et al., 2006). Machine learning approaches are based on training data to learn the prominent features for term recognition and classification. One of the main problems in machine learning is the reliable training resources, which are not widely available (Benson, et al., 2000; Krauthammer and Nenadic, 2004). Other challenges are related to selection of discriminating features and detection of term boundaries (Krauthammer and Nenadic, 2004).
Hybrid approaches combine different methods, for example, rule-based and statistical techniques. Such approaches also used various resources, such as pre-compiled lists of specific terms or words for the term recognition task (Krauthammer and Nenadic, 2004; Tanabe and Wilbur, 2002).

One of the medical documents that contain important information about on going investigation of patient is known as “Referral and Response Letters”. Referral letters are a common means by which healthcare practitioners exchange information relevant to patient care (Tattersall, et al., 2002). The content of letters can range from straightforward single problems, e.g. a sebaceous cyst or lipoma to complex cases in which extensive details need to be communicated in both direction (John, et al., 1992; Westerman, et al., 1990). Both specialists and GPs are united in their desire to exploit evidence-based healthcare based on knowledge from ‘Clinical practice guidelines’ and other reliable online sources such as ‘Systematic reviews’, ‘Randomised controlled trials’, ‘Case-control studies’ etc. However, formulating correct and accurate key-terms is necessary for performing electronic searches (Olina, 2005). Poorly formulated and unfocused queries have been identified as a major hurdle in accessing relevant medical information and knowledge (Ely, et al., 2005). The narrative text in medical referral letters contains important medical information related to the patient’s medical problems and is consequently a good source of information for formulating a query to access the latest evidence-based practice. The main challenge is developing a method to automatically analyse and extract the relevant information. This includes finding semantic relationships among the terms in the letter, semantically classifying and standardising them, and specifying their meaning. A suitable technique needs to be
developed for processing the information in medical letters. It could be used to formulate queries to represent the focused information needs of GPs or other healthcare practitioners. The research problem related to medical information processing of referral letters has been described in section 1.4.1 in detail and has been framed in the following research question:

- How to enable computerized processing and analysis of medical referral and response letters to provide comprehensive information-view for customized query formulation?

In next section, we address the above defined research question by presenting a framework for automatic medical information processing of referral letters (AMIPF). We describe rationale of our approach, elaborate on architecture and functional flow of AMIPF.

6.3 AMIPF: Automatic Medical Information Processing Framework

To address our research question, we have developed an Automatic medical Information Processing Framework (AMIPF) that is based on our method of computerized processing and analysis of medical information in referral and response letters.

The AMIPF is designed to facilitate optimal querying of evidence-based knowledge bases and hence helps meet the information needs of GPs while dealing with referral letters and their responses. The techniques and strategies developed for the AMIPF are based on a hybrid approach, since they are drawn from knowledge management, ontology modeling, and natural language processing. The AMIPF analyzes letters to detect
symptoms, diseases, diagnostic procedures, therapies, and automatically generates potential query terms. The query type, which is also determined can be for Diagnosis, Etiology, Prognosis, or Therapy (Haynes, et al., 1994). The AMIPF also detects critical situations that require attentions. The results of AMIPF analysis are intended to help GPs and other healthcare practitioners in formulating customized queries.

The underlying rationale of our approach is that medical letter contains important information related to a patient’s medical problems that could be used to help healthcare practitioners formulate a focused clinical query. This medical information should be analyzed, extracted and identified into general and domain-specific classification to provide a meaningful view of the information in a letter. Negated sentences along with their negated terms should be identified to enhance the quality of automatic generated query terms and customized query. Potential query terms should be generated, automatically, containing semantically and statistically important medical terms that will further help in customizing focused query formulation.

6.3.1 Functional Flow and Techniques of Automatic Medical Information Processing Framework

The functional-overview of the AMIPF is shown in Figure 6.1. The computerised processing of a referral letter by the AMIPF can be grouped into four major modules, where each module consists of a set of smaller processing components. The four modules include:

- Noun-Phrases and Concepts Recognition - recognises medical noun-phrases and terms and finds the related concepts in the UMLS (Unified Medical Language System) meta-thesaurus.
• Classification of Noun-Phrases and Concepts - groups “noun-phrases and concepts” into different medical categories.

• Information Mapping - maps classified noun-phrases and concepts to domain knowledge to refine their meaning.

• Clinical Query Formulation - automatically generates a query and provides an information-view from the referral letter to assist healthcare professionals in formulating a customised query.

In the following sections, we define the techniques and detailed functionalities of modules and components of the AMIPF framework.

6.3.1.1 AMIPF Input Pre-processing

The referral letters to be processed are in the form of a free text document, it can be a Microsoft Word document, PDF or any other format, in our case in Microsoft Word. To process such a letter it is necessary to extract the text from the ‘letter-specific-format’. The “Extraction text from letters” module implements a software scheme to extract text from a letter. After extraction the text is sent to the ‘Noun-phrase and concept recognition’ module.
6.3.1.2 Noun-phrase and Concept Recognition

The ‘Noun-phrase and concept recognition’ (NPCR) consists of a set of smaller modules. Figure 6.2 shows the flowchart for the functional steps of the NPCR. The technologies used are:

(i) A POS (part of speech) tagger and ‘SPECIALIST lexicon’ provided by the National Library of Medicine (National Libraray of Medicine, 1994; Smith, et al., 2004).

(ii) A MMTX API (Application programming interface) with Java 1.5 to implement the steps required for ‘Noun-phrase and Concept Recogniser’.

(iii) SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) and MeSH (the National Library of Medicine's controlled vocabulary for indexing articles) thesauri from UMLS (National Library of Medicine, 2001).
Figure 6.1 Architecture for automatic medical information processing framework, where POS = Part of speech, SMR = Semantic relation score, S.T= Semantic Types, and UMLS = Unified Medical Language System.
Extracted text from previous module is processed by ‘Parser-Tokenizer’. This text is parsed into sections consisting of sentences and tokens. The ‘Part of speech tagger’ (POS Tagger) takes the tokens and tag them to identify which part of speech the token belongs to. We are using POS tagger developed by (Smith, et al., 2004). The tagged-tokens are sent to ‘Lexical Look up’. The ‘Lexical Look up’ finds out if any of the tagged-token belongs to a particular lexicon. We are using ‘SPECIALIST lexicon’ provided by NLM (National Library of Medicine, 1994). The results of this module are the lexical elements which are passed to ‘Noun-phrase parser’. The ‘Noun-phrase parser’ is a barrier category parser that uses the part of speech categories from lexical elements and the part of speech tags from a tagger (Guy, 2003).

These noun-phrases are sent to Variant Generator to calculate variants. “Concept Extractor” uses noun-phrases and their variants to match phrases/concepts from UMLS Metathesaurus. We are using SNOMED CT and MeSH thesauri from UMLS. The small modules for ‘Tokenization’, ‘Variant generation, ‘Noun-phrase parser’ ‘phrases/concepts extraction’ are freely available in java from (National Library of Medicine, 2001). These modules can also be accessed using MMTX API (MetaMap Transfer, 2007). We are using MMTX API with java 1.5 to implement the steps required for ‘Noun-phrase and Concept Recognizer’. The extracted UMLS medical concepts (medical phrases and medical concepts) are sent to ‘Meta-info Extractor’.

174
Figure 6.2 Flowchart for the functional-steps of Noun-phrase and Concept Recognition, UMLS = Unified Medical Language System.
6.3.1.3 Classification of Noun-Phrases and Concepts

6.3.1.3.1 Meta-Info Extractor

Each medical noun phrase and medical term is enriched with related meta-information by the Meta-info Extractor module. The meta-information, retrieved from the UMLS, includes: a) Semantic Type, b) UMLS score, which is “the strength/confidence of the mapping of the original phrase to the corresponding SNOMED CT term”, c) Original medical phrase or medical term, and d) Vocabulary source. This meta-information and its corresponding medical concept are sent to the ‘Indexer’ module for indexing.

6.3.1.3.2 Indexer

The “Indexer” module creates an index of the extracted medical concepts along with the retrieved meta-information. We refer to this index as “Referral-Letter index” (R-L index). It also creates an index based on the frequency of each medical concept its SMR score and Total Weight (see section 6.3.1.5.1) within a letter. Another function of the ‘Indexer’ is to keep track of all sentences along with their corresponding ‘UMLS concepts’. This is achieved by indexing sentences, as a full string, along with its ‘ID’ and corresponding ‘UMLS concepts’.

6.3.1.3.3 Critical Situation Sensor

The main task of the ‘Critical Situation detector’ is to identify a medical concept, or a combination of medical concepts that need attention. This is because specialists feel that GPs often ignore or fail to urgently act on critical symptoms because they are unaware that they may be circumstantial. Specialists would like to highlight to GPs that when patients develop specific symptoms or combinations of symptoms e.g. syncope in a
patient with aortic stenosis, they should be referred urgently rather than routinely. In consultation with Cardiologists, we have developed taxonomy along with some rules that are applied to “R-L index” to identify critical situations.

6.3.1.3.4 Redundancy Filter

For automatic query generation and customized query formulation, redundant medical concepts are required to be filtered out. The ‘Redundancy Filter’ is applied to remove the repetition of medical concepts. The redundancy is determined by comparing medical concepts and their meta-information attributes. The resulting set of medical concepts is sent to the ‘Semantic Type Filter’.

6.3.1.3.5 Semantic Type Filter

During the extraction process, all medical noun-phrases and concepts in a referral letter are extracted. Some of the extracted concepts will have less medical significance given the context of the referral letter and the task of query formulation. We have developed a “Semantic Type Filter” (S.T Filter) to filter out those medical concepts that seem less significant based on their semantic types. The list of semantic types to be used in S.T Filter was generated, with a Cardiologist and using the results of a qualitative analysis of the usage of these semantic types.

6.3.1.3.6 Negation Detector

Identification of the absence of a particular medical finding enhances the information-view for ‘customised query formulation’ and the context-sensitivity for ‘automatic query generation’. For example, in a referral letter, a sentence such as “He has good effort tolerance, playing tennis each week without any symptoms of chest
discomfort, dyspnoea, or palpitations” contains the medical findings: discomfort, dyspnoea, and palpitations. These findings will be extracted and remain in the set of possible significant ‘medical concepts’ after the filtration processes. However although these symptoms appear in the text of the letter, the patient does not actually have these symptoms. Eliminating such “negative” medical findings is an important step for query formulation.

The Negation detector uses the NEgEx negation detection algorithm for finding negated medical concepts (Chapman, et al., 2001). “NegEx’s input is a sentence with indexed findings and diseases, while its output is whether the indexed terms are explicitly negated in the text”(Chapman, 2003). NegEx works on a sentence and finds one negated medical concept at a time (see Figure 6.3). If the UMLS concept is found negated, it is tagged and stored as a negated medical phrase or concept along with its ‘sentence ID’ by the ‘Indexer’. In this manner, all sentences in a referral letter are processed to find negated UMLS concepts.
Figure 6.3 Flowchart of sentence analysis and negation detection in a referral letter
6.3.1.3.7 Phrase and Concepts Classifier

The role of term classification is to pinpoint the specific type of a domain concept to specify the meaning of a term (Krauthammer and Nenadic, 2004). We classify medical concepts from referral letter to specify the meaning of medical concepts. To provide a comprehensive and focused view of information for formulating a query, medical concepts are classified into five categories:

- Diseases
- Symptoms
- Diagnostic procedures
- Therapies
- Medications

Each category has a set of semantic types that serve as ‘category-trigger’ for a corresponding category. These category-triggers are chosen from the list of ‘semantic types’ selected for the ‘Semantic Type Filter’. The grouping of category-triggers is based on a definition of semantic types (Bodenreide and McCray, 2003; Haynes, et al., 1994; Long, 2005; McCray, et al., 2001; Sibanda, et al., 2006; Zhenyu and Wesley, 2007). For example, semantic types [“Organic Chemical” and “Pharmacologic Substance”] are the ‘category-trigger’ for the ‘Medications’ category.

6.3.1.4 Information Mapping

One of the tasks of the Information Mapping is to remove the ‘contextual-noise’ by removing those ‘UMLS concepts’ that do not conform to a ‘Disease Ontology’. The
second task is to map the significant medical phrases and medical concepts to ‘Disease Ontology’ to find more specific meanings.

6.3.1.4.1 Diseases Ontology

We have developed a prototype of a *Heart Diseases Ontology* that has been explained in chapter 5 (see section 5.3.2.5) and shown in Figure 5.3. In general, ‘Diseases Ontology’ contains domain knowledge that provides relationships and information about different entities such as diseases, symptoms, diagnosis procedures, etc. Such domain knowledge plays an important role in identifying the type and relationships of different medical terms.

6.3.1.4.2 Info-Mapper

During the extraction of UMLS concepts from letters, some of the extracted UMLS concepts are not contextually present in the letter. Such UMLS concepts are considered ‘contextual-noise’ that affects the efficacy and accuracy of the query formulation process. For example, after processing the sentence: “Patients with ST segment elevation myocardial infarction usually have a completely occluded coronary artery with thrombus at the site of a ruptured plaque Restoring coronary patency.....” the term ‘plaque’ will result in extraction of the following UMLS concepts: (i) Dental Plaque [with semantic type ‘Disease or Syndrome’], (ii) Plaque (lesion) [with semantic type ‘Body Substance’], and (iii) Senile Plaques [with semantic type ‘Disease or Syndrome’]. UMLS concepts ‘Plaque (lesion)’ and ‘Senile Plaques’ are filtered out by the Semantic Type Filter, but ‘Dental Plaque’ remains the member of the set of significant medical phrases and concepts. In reality, ‘Dental Plaque’ is not related to the context in the referral letters of
cardiology domain. The Info-Mapper uses the Heart Diseases Ontology to remove such contextual-noise and to specify the meaning within the cardiology domain.

6.3.1.5 Query Formulation

The “Query Formulation” module automatically generates a query and comprehensive information-view from a letter. This assists GPs and healthcare professionals in choosing key terms for formulating customized queries.

6.3.1.5.1 Automatic Query Generation

The ‘Automatic Query Generation’ module generates potential query terms from medical concepts that belong to the following five categories: diseases, symptoms, diagnosis, therapies, and diagnostic procedures. The UMLS concepts that belong to medications currently are not used (but may be added during customized query formulation). The reason is that articles in PubMed are indexed based on the MeSH terms found in titles, keywords, and abstracts that normally contain very few medications.

This module is actually used to finalize the medical concepts for the search query that will be automatically generated. It has been described earlier in chapter 5 that search queries should only include small numbers of highly significant medical terms. We identified five as being the maximum number of medical terms/phrases that should be used in a query (see section 5.3.2.7). We have used this number to define the threshold for a Final query set (THFQS).

The medical concepts for the Final query set are determined as follows. Upon receiving medical concepts, the ‘Total Weight’ and the ‘SMRscore’ modules are used to generate two sets of medical concepts:
a) Medical concepts sorted by Total Weight (TW set)

b) Medical concepts sorted by ‘SMRscore’ (SMR set).

The ‘Total Weight’ module calculates the *Total Weight* for each medical concept as a summation of the frequency of each medical concept in a referral letter and its *UMLS score* (see equation (6.1)). The medical concept frequency refers to the “number of time” a medical concept found in the referral/response letter. The ‘UMLS score’ refers the strength/confidence of the mapping of the original phrase to the corresponding UMLS term. Total Weight is formally defined in equation 6.1, where \( \Psi \) is a Total Weight of a medical candidate, \( \lambda \) is a concept frequency and \( \tau \) is a ‘UMLS score’.

\[
\Psi = \lambda + \tau \tag{6.1}
\]

The SMRscore module determines the semantically related UMLS concepts and quantifies their semantic relationship i.e. called as ‘SMRscore’. Its functionality and algorithm are same as defined in section 5.3.2.6.
The Automatic Query Generation (AQG) module evaluates the number of received medical concepts against the threshold for the Final query set. If the number of received
medical concepts is less than or equal to the threshold value of 5, it takes these medical concepts as the *Final query set* and determines its query type. In the event that the number of received medical concepts is more than the threshold value of 5, then a Common Term Analysis is performed to determine which medical concept will be included in the *Final query set*. In the Common Term Analysis three different threshold values are used: 100%, 50%, and 25% of the highest value in the *TW set* and the *SMR set*. All concepts that satisfy the condition of the first threshold are included in the *Final query set*. If more concepts are needed, then the second and third thresholds are applied.

If the number of Final query set medical concepts is less than the threshold for the Final query set, AQG completes the remaining number of medical concepts based on the Total Weight.

The queries are categorized automatically into four categories as used in PubMed (Haynes, et al., 1994). Figure 5.1(c) shows the process for determining a query type as described earlier. The query type for the search query is determined by finding the maximum numbers of medical terms/phrases that belong to a specific query type. In the event that the same number of medical concepts belong to two query types, both query types are assigned (Abidi, et al., 2005a) to the automatically generated query. If a query has more than two query types then a General query type is assigned. Figure 6.5 shows the flowchart of the automatic query generation procedure for Final query set.

6.3.1.5.2 Information-View: Customized Query Formulation

The “Customised Query Formulator” provides an information view of the processed referral/response letter to help GPs and other healthcare practitioners to formulate a customized query. The information view includes medical information
classified into five general categories, five domain-specific (cardiology) categories, 
negated sentences along with their negated terms, automatically generated query terms, 
and alerts for critical situations. Healthcare practitioner is also presented with the view of referral/response letter.

The classification into general categories represents medical information in general, 
whether they belong to cardiology or not. These general categories are: symptoms, 
diseases, diagnosis, therapies, and medications. The classification into domain-specific 
(cardiology) categories represents that medical information that belongs to cardiology 
only. These categories are: heart disease symptoms, heart diseases, diagnostics 
procedures for heart diseases, heart diseases therapies, and heart diseases medications (see 
Appendix F). All these information are presented to help formulate a focused customized 
query so that relevant information form evidenced-based knowledge source could be 
accessed.

In next section, we present a case study through a working example using a 
response letter to a referral letter to comprehensively illustrate the functionality of the AMIPF framework.

6.4 A Case Study: Working Example of the AMIPF

A ‘de-identified’ medical response letter to a referral is shown in Figure 6.6 (see 
Appendix G for more letters). The text of this letter is extracted and is processed by the "Noun-phrase concept recognition" module. This module generates UMLS medical 
concepts along with their variants. For this letter 184 UMLS medical concepts are 
generated. The ‘Meta-info Extractor’ extracts the meta-information for each UMLS
candidate. Some of extracted UMLS concepts are shown in Table 6.1 along with their original name and semantic types (see Appendix E, Table E.1 for all extracted medical concepts). These UMLS concepts are indexed and then sent to ‘Redundancy Filter’ module. Redundancy filter module filters out the redundant medical concepts, which results in 145 medical concepts.

Thank you for asking me to see Mrs. — whom I saw today the 16th November, 2007. She recently injured her right foot with fractured toes and has been relatively immobile for a while and during this time has developed ankle oedema bilaterally. The oedema is there most of the day and does not appear to become worse during the day. Also however she has noticed dyspnoea going up stairs, but she has also gained a significant amount of weight after being treated with Arimidex for breast cancer. Mrs. — has no history of ischaemic heart disease but she did have rheumatic fever aged eight.

She had a breast carcinoma in January this year with a limited resection being followed by radio therapy and now long term hormonal therapy. Risk factors for heart disease include hypertension and hypercholesterolemia.

Her medications currently include Atacand 16mg mane, 8mg nocte, Lipitor 10mg mane, Astrix 100mg mane, Lasix 40mg b.d. (recently increased), Arimidex 1mg, Nexium 40mg mane, Rani 2 and Alodorn.

On examination she weighed 94.5kg, the pulse was regular at 50 beats per minute, blood pressure 130/80, the JVP was elevated slightly and there was very mild ankle oedema limited just to the ankles and feet. On auscultation there was a grade 2/6 systolic ejection murmur at the base. The chest was clear.

ECG showed sinus bradycardia with non-specific ST T-wave changes.

Echocardiography today showed normal left ventricular and right ventricular size and function. She has mild aortic stenosis and mild mitral stenosis. She has mild tricuspid regurgitation with mild pulmonary hypertension.

The increased dose of Lasix recently has not improved her dyspnoea or her ankle oedema, so I have asked her to reduce the dose back to 40mg mane and to see you regularly for checks of potassium. Overall I feel her heart condition is mild and not likely to be contributing particularly significantly to her oedema, which is more likely to be related to a recent immobility and her weight gain. I suggested that antibiotic prophylaxis should be used and regular exercise once she is able would be helpful. In terms of sinus bradycardia, at this stage she seems to be asymptomatic, but if she has episodes of presyncope or syncope in the future we should reconsider this as a problem.

I have organised to review her again in six month’s time.

With kind regards,

Yours sincerely,

Figure 6.6 De-identified response letter to a referral

These medical concepts are sent to ‘S.T Filter’ that filters out the medical concepts based on their semantic types. For example ‘Neoplastic Process’, ‘Receptor’, and ‘Organism Attribute’, ‘Intellectual Product’, ‘Functional Concept’ etc are not considered
significant for our method. Table 6.1 shows some of the medical concepts, which are removed, with light shaded cells. After this process, 46 remaining medical concepts are sent to the ‘Negation detector’. ‘Negation detector’ detected one medical concept in the letter within the sentence “Mrs. ---- has no history of ischaemic heart disease but she did have rheumatic fever aged eight”. The negated medical concept is “ischaemic heart disease”. After the removal of the negated medical concept, the total number of concepts is 45. These concepts are classified by the ‘Noun-phrases and concepts classifier’. The classified medical concepts into general categories are shown in Table 6.2. All medical concepts are sent to the “Info-Mapper” except those medical concepts that belong to the ‘Medications’ category.

<table>
<thead>
<tr>
<th>Weighing patient</th>
<th>Weight</th>
<th>Diagnostic Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treats</td>
<td>Treats</td>
<td>Classification</td>
</tr>
<tr>
<td>Arimidex</td>
<td>Arimidex</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>Breast Cancer</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>Mineralocorticoid receptor</td>
<td>MR</td>
<td>Receptor</td>
</tr>
<tr>
<td>History of present illness</td>
<td>History, NOS</td>
<td>Organism Attribute</td>
</tr>
<tr>
<td>Medical history</td>
<td>History</td>
<td>Finding</td>
</tr>
<tr>
<td>Myocardial ischemia</td>
<td>Ischaemic heart disease</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>Rheumatic fever</td>
<td>Rheumatic Fever</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>Eight</td>
<td>Eight</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>Breast Carcinoma</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>Year</td>
<td>year</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>Limited</td>
<td>Limited</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>Resection</td>
<td>Resection</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>Radio communications</td>
<td>Radio</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>Therapeutic aspects</td>
<td>therapy</td>
<td>Functional Concept</td>
</tr>
</tbody>
</table>

Table 6.1 A table with medical concepts to be filtered out based on their semantic types. Medical concepts with light shaded cells are filtered out as their semantic types of not being significant.
<table>
<thead>
<tr>
<th>SYMPTOMS</th>
<th>DIAGNOSTIC PROCEDURE</th>
<th>THERAPIES</th>
<th>MEDICATIONS</th>
<th>DISEASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immobile</td>
<td>Weighing patient</td>
<td>Excision</td>
<td>Arimidex</td>
<td>Rheumatic Fever</td>
</tr>
<tr>
<td>Swollen ankle</td>
<td>Physical Examination</td>
<td>Diagnostic procedure</td>
<td>Atacand</td>
<td>Heart Diseases</td>
</tr>
<tr>
<td>Edema</td>
<td>Pulse taking</td>
<td>Endocrine therapy</td>
<td>Lipitor</td>
<td>Hypercholesterolemia</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>Auscultation</td>
<td>Antibiotic Prophylaxis</td>
<td>Lasix</td>
<td></td>
</tr>
<tr>
<td>Medical History</td>
<td>Electrocardiogram</td>
<td>Physiologic pulse</td>
<td>Nexium</td>
<td></td>
</tr>
<tr>
<td>Systemic arterial pressure</td>
<td>Echocardiography</td>
<td>Physiological aspects</td>
<td>Alodorm</td>
<td>Aortic Valve Stenosis</td>
</tr>
<tr>
<td>Heart murmur</td>
<td>Magnetic Resonance Spectroscopy</td>
<td></td>
<td>Base</td>
<td>Mitral Valve Stenosis</td>
</tr>
<tr>
<td>Ejection murmur</td>
<td></td>
<td>MEDICATIONS</td>
<td>Tricuspid Valve Insufficiency</td>
<td>Hypertension, Pulmonary</td>
</tr>
<tr>
<td>(D) Sinus bradycardia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Related personal status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight Gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Able</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptomatic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presyncope</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syncope</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The resulting 37 medical concepts are processed by “Info-Mapper” to determine those medical concepts that have significance with respect to Cardiology. After processing 37 medical concepts by “Info-Mapper”, only 15 medical concepts remained. Table 6.3 shows the remaining medical concepts as an output of “Info-Mapper”, where column for therapies is empty because no cardiology-specific therapy is identified. These 15 candidates are sent to “Query Formulation Module”. The “Total Weight” and “SMRscore” for each medical concept are calculated using equation 6.1 and 5.4. Table 6.4 shows the UMLS medical concepts, the original name of the medical concepts (as found in the letter), their semantic types, Total Weight and “SMRscore”.

189
Table 6.3 Medical concepts after ‘Info-Mapper’ processing

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Diagnostic procedure</th>
<th>Therapies</th>
<th>Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyspnea</td>
<td>Auscultation</td>
<td></td>
<td>Tricuspid Valve Insufficiency</td>
</tr>
<tr>
<td>Syncope</td>
<td>Electrocardiogram</td>
<td></td>
<td>Mitral Valve Stenosis</td>
</tr>
<tr>
<td>Heart murmur</td>
<td>Echocardiography</td>
<td></td>
<td>Aortic Valve Stenosis</td>
</tr>
<tr>
<td>Sinus bradycardia</td>
<td></td>
<td></td>
<td>Hypercholesterolemia</td>
</tr>
<tr>
<td>Ejection murmur</td>
<td></td>
<td></td>
<td>Rheumatic fever</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hypertensive disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hypertension, Pulmonary</td>
</tr>
</tbody>
</table>

At this stage, the “Automatic Query Generation” module compares the number of significant medical concepts with the ‘concepts-threshold’, a number defined for the length of a query, to decide what to do next.

Table 6.4 Medical concepts with their ‘Final Weight’ and SMR ‘score’

<table>
<thead>
<tr>
<th>UMLS Medical Candidates</th>
<th>Original Name</th>
<th>Semantic Type</th>
<th>Total Weight</th>
<th>SMR score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aortic Valve Stenosis</td>
<td>Aortic Stenosis</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Auscultation</td>
<td>Auscultation</td>
<td>Diagnostic Procedure</td>
<td>862</td>
<td>8</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>Dyspnea</td>
<td>Sign or Symptom</td>
<td>863</td>
<td>8</td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td>ECG</td>
<td>Diagnostic Procedure</td>
<td>1001</td>
<td>8</td>
</tr>
<tr>
<td>Echocardiography</td>
<td>Echocardiography</td>
<td>Diagnostic Procedure</td>
<td>695</td>
<td>8</td>
</tr>
<tr>
<td>Ejection murmur</td>
<td>Ejection murmur</td>
<td>Finding</td>
<td>820</td>
<td>8</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>Hypercholesterolemia</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Hypertensive disease</td>
<td>Hypertension</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Mitral Valve Stenosis</td>
<td>Mitral Stenosis</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Heart murmur</td>
<td>Murmur</td>
<td>Sign or Symptom</td>
<td>796</td>
<td>8</td>
</tr>
<tr>
<td>Hypertension, Pulmonary</td>
<td>Pulmonary hypertension</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Rheumatic Fever</td>
<td>Rheumatic Fever</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Sinus bradycardia</td>
<td>Sinus bradycardia</td>
<td>Pathologic Function</td>
<td>1002</td>
<td>14</td>
</tr>
<tr>
<td>Syncope</td>
<td>Syncope</td>
<td>Sign or Symptom</td>
<td>1001</td>
<td>8</td>
</tr>
<tr>
<td>Tricuspid Valve Insufficiency</td>
<td>Tricuspid Regurgitation</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
</tbody>
</table>

It can be seen from Table 6.4 that the numbers of medical concepts are more than our ‘concepts-threshold’, so, the ‘Automatic Query Generation’ applies the ‘Common
Term Analysis’ to select the medical concepts for the Final query set. Table 6.5 shows the medical concepts in light shaded cells that are selected for the Final query set. The query type of the final query is ‘diagnoses’. The reason for this is that four medical concepts belong to ‘Disease or Syndrome’ semantic type, which is a trigger-semantic type for ‘Diagnosis’.

The automatically generated query terms are Sinus bradycardia, hypercholesterolemia, Hypertensive disease, Rheumatic Fever, Aortic Valve Stenosis, and its query type is Diagnosis.

The ‘Customised Query Formulator’ provides an information view of the processed referral letter to help GPs and other healthcare practitioners to formulate a customized query (see Appendix F, Figure F.7, F.8, and F.9).

<table>
<thead>
<tr>
<th>UMLS Medical Concepts</th>
<th>Semantic Type</th>
<th>Final Weight</th>
<th>SMR score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinus bradycardia</td>
<td>Pathologic Function</td>
<td>1002</td>
<td>14</td>
</tr>
<tr>
<td>hypercholesterolemia</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Hypertensive disease</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Rheumatic Fever</td>
<td>Disease or Syndrome</td>
<td>1001</td>
<td>14</td>
</tr>
<tr>
<td>Aortic Valve Stenosis</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Mitral Valve Stenosis</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Hypertension, Pulmonary</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Tricuspid Valve Insufficiency</td>
<td>Disease or Syndrome</td>
<td>902</td>
<td>14</td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td>Diagnostic Procedure</td>
<td>1001</td>
<td>8</td>
</tr>
<tr>
<td>Syncope</td>
<td>Sign or Symptom</td>
<td>1001</td>
<td>8</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>Sign or Symptom</td>
<td>863</td>
<td>8</td>
</tr>
<tr>
<td>Auscultation</td>
<td>Diagnostic Procedure</td>
<td>862</td>
<td>8</td>
</tr>
<tr>
<td>Ejection murmur</td>
<td>Finding</td>
<td>820</td>
<td>8</td>
</tr>
<tr>
<td>Heart murmur</td>
<td>Sign or Symptom</td>
<td>796</td>
<td>8</td>
</tr>
<tr>
<td>Echocardiography</td>
<td>Diagnostic Procedure</td>
<td>695</td>
<td>8</td>
</tr>
</tbody>
</table>
6.5 Summary

Computerized processing of medical information from narrative text medical documents is a challenging task. Such a computerized processing of medical documents involves many tasks like medical terms recognition, extraction, classification, identification, negation detection etc. These tasks are crucial for different medical applications. We have defined the significance of computerized processing and analysis of medical referral letters to help formulate an optimal clinical query.

We have presented an Automatic medical information processing framework (AMIPF). The AMIPF processes a referral/response letter: (i) to help formulate customized query by providing comprehensive information-view of analyzed letter and (ii) to provide alerts for critical medical situation.

Computerized processing of a letter involves (i) Extraction of text from specific format, (ii) recognizing and extracting the medical concepts from narrative text, (iii) extraction of meta-information pertaining to the medical concepts, (iv) indexing of important medical information, (v) elimination of redundant medical concepts, (vi) filtration of non-significant medical concepts, (vii) identification and removal of negated medical concepts, (viii) classification of the medical concepts to derive the meaning, (ix) information mapping to further specify the meaning and removal of contextual-noise (x) generation of an automatic potential clinical query, and (xi) creation of comprehensive information-view for the formulation of a customized clinical query.

We have explained the functionalities of the AMIPF framework. We have explained the processing of a letter by describing the techniques of each module.

192
have presented a case study through a working example of the response letter processing
to comprehensively illustrate our method and techniques.
CHAPTER 7

KNOWLEDGE RETRIEVAL FROM COMPUTERIZED CLINICAL GUIDELINES

7.1 Introduction

Effectively, searching and retrieving to access the relevant content of the documents has become crucial with the exponential growth in digital literature, in general and the medical literature in particular (Druss and Marcus, 2005). Clinical practice guidelines (CPGs) are meant to support clinical decisions and to help manage patients at point of care (Robert and Yuval, 2009). The requirements for efficient access and retrieval of the relevant and context-specific medical knowledge of the clinical practice guidelines are also crucial in their usage at point of care. A recent paradigm shift geared towards evidence-based medicine and provision of automated support to guideline-based clinical care has emphasized such requirements.

Techniques and algorithms for the effective access and retrieval need to be developed, which must consider the special requirements and information introduced by the process of computerizing and structuring guidelines for the purpose of automated support to guideline-based clinical care (Robert and Yuval, 2009; Sonnenberg and Hagerty, 2006). It has been stated that exploiting the potential existence of rich-domain specific knowledge, (which is common in medical domain), especially, when using clinical practice guidelines would enhance the efficacy of the search and the retrieval (Robert and Yuval, 2009). In this chapter, we present and discuss our technique for the
contextually relevant retrieval and access of concise segments of computerized clinical practice guidelines. We first present an overview of computerized clinical practice guidelines (C-CPG) knowledge access at point of care and related issues in section 7.2, proceeding to section 7.3, we present Content and Knowledge Retrieval (CKR) framework and discuss its each module in detail. Indexing strategy and formal representation of the query and the knowledge base are also discussed in this section. In section 7.4, we present Clinical Knowledge Assistance system architecture and describe its functionalities. We summarize this chapter in section 7.5.

7.2 An Overview of Computerized Clinical Practice Guidelines Content and Knowledge Access at Point of Care

It has been discussed that clinical practice guidelines (CPG) are useful at the point of care and substantial evidence exists that conforming to clinical practice guidelines improves the quality of care (Grimshaw and Russel, 1993; Micieli, et al., 2002; Qualigini, et al., 2004). Currently, CPGs are represented in free text that poses limitation for effective use of CPGs and provides less support for computerized applications of CPGs at point of care (Robert and Yuval, 2009). Many approaches for structured representation format have been proposed to transform CPGs into computer interpretable format (see section 2.2 for details). Once CPGs have been transformed into computer interpretable format, their content is made available at point of care by the ‘Guideline Execution’. The ‘Guideline Execution’ is a process by which management of guideline information is enacted so that information could be accessed. There are normally two types of execution engine approaches for clinical guidelines (i) event-based approach (EB) and (ii) rule-based approach (RB) (David and Antonio, 2008). The event-based execution engines can be
used in a continuous system where events are handled asynchronously as they appear (e.g. the arrival of patient’s results etc). The rule-based execution engine is monitored by another operator that supervises and controls the rules that can be triggered in any time in a synchronous manner (David and Antonio, 2008). Such approaches to access guidelines content require patient data integration with the workflow in form of EMR or EHR. Following, we provide a brief overview of existing approaches for accessing CPGs content.

GLARE, Guideline Acquisition, Representation and Execution, is a system to acquire and execute clinical guidelines (Terenziani, et al., 2001). This is based on rule-based execution approach. CPGs in GLARE do not use any external or standard representation rather their authors have defined a proprietary graph-based representation. In such representation each action is represented by node and control relations are represented by arcs. For each action there is a set of preconditions that need to be matched for execution and a set of conclusions that are presented after the execution (David and Antonio, 2008).

GLEE, Guideline Execution Engine, is used for executing guidelines that are encoded in GLIF model version 3 (GLIF3) (Wang, et al., 2004; Wang and Shortliffe, 2002). GLEE is based on event-based approach for guidelines execution. In GLIF3 CPGs are represented as flowcharts of temporally ordered nodes called guidelines steps. These steps store actions (Action_Steps), decisions (Decsion_Steps), and clinical states of the patient (Patient-Clinical_States). GLEE execution model adopts an approach known as “system suggests user controls”. It supports an event-driven execution once it is linked to the clinical event monitor in a local environment (David and Antonio, 2008).
NewGuide is a framework for modeling and executing clinical practice guidelines (Ciccarese, et al., 2004; Ciccarese, et al., 2005). It is a rule-driven guideline execution system. In ‘NewGuide’ CPGs are represented using a representation language called GUIDE that is based on Petri Nets (Quaglini, et al., 2000). Its inference engine consists of three main components: a general manager, a message manager, and an instance manager. The inference engine is invoked by healthcare practitioners and it creates automatically an instance of CPG for the management of the patient. Once the guideline is loaded, the instance manager collects patient’s data stored in patient record. The execution engine goes step-by-step recommending actions (David and Antonio, 2008).

There are some other CPGs execution engines like SpEM (Dube, et al., 2005), HeCase2 (Isem and Moreno, 2004), SAGE (Berg, et al., 2004), etc that are based on event-driven approaches while DeGeL (Young, et al., 2007), Arezzo (Fox, et al., 2006) etc are based on rule-driven execution engines. Most of the approaches for accessing and utilizing the content of computerized guidelines center on execution engine.

There are few other approaches to access and navigate the content of CPGs. Poon et al. (2006) developed a method for the retrieval and navigation of CPGs based on ‘infobuttons’ approach. They developed a list of question that were categorized according to the classification proposed by (Ely, et al., 2000). They achieved the navigations through hyperlinks from each question to relevant parts of the guidelines. Guidelines were tagged using “named destinations” to provide a quick navigation to the specific portion of the document containing the answer. So each question or set of question were defined for certain infobutton that represent the problem from EMR. Each question or set of questions were linked to the specific portion of guidelines (as guidelines were tagged).
Selecting the question or set of questions and clicking the infobutton for a specific problem in EMR would take the corresponding guideline that has hyperlink to the specific portions of guideline.

Another approach for the retrieval of clinical practice guideline is developed by (Robert and Yuval, 2009). They developed a concept-based clinical guideline search engine “Vaidurya” to retrieve a set of clinical guidelines from digital library of guidelines (a repository of guidelines).

One of the requirements for execution-engine-based approaches is the integration of patient data via EMR etc or mechanism of accessing the patient data/decision variables that are required for a rule to be triggered or an event to be invoked. Other approaches like Poon et al. (2006) requires EMR and direct linking of question to the tagged guidelines. David and Antonio (2008) argued that the best possible representation of an EMR is in itself is exiting problem and has not been solved yet whereby applications are made ad hoc to fit a certain representation. They also stated that in most of the countries (including ones considered to be more technologically advanced), healthcare is not yet fully computerized that makes it hard to include automated guideline enactment systems in real clinical settings (David and Antonio, 2008). On the other hand, those clinical settings, where input would not be from EMR or EHR, require new methods and techniques to access clinical practice guideline content at point of care.

Clinical guidelines are normally transformed into computer interpretable format based on model representation or ontology (e.g asbru, GEM etc). This model or ontology is referred as contextual model for CPGs (Moskovitch, et al., 2007). Moreover, computerized clinical guidelines contain more semantically meaningful segments and
internal structure. Such segments and structure expose a sense of context for CPGs content (Hashmi, et al., 2009 b; Hashmi and Zrimec, 2008 b; Robert and Yuval, 2009). Additionally, meta-information and structural information (semi-structured document—XML format) introduced by computerized methods enrich CPGs content. Such information rich representation of computerized CPGs needs new approaches and techniques to access relevant CPGs content and knowledge in concise yet focused segments. The fourth research problem of this thesis has been articulated along the same lines in section 1.4.1 and has been framed into the following research questions:

- What approaches from knowledge management and information retrieval could be used to develop ‘concise content retrieval’ technique that exploits the information provided by the computerized clinical guidelines?

- How to develop information structure and indexing strategy to support effective retrieval of the semi-structured medical documents (in our case C-CPGs)?

In the following section, we present *Content and Knowledge Retrieval* (CKR) framework that addresses the above defined research questions.

### 7.3 Content and Knowledge Retrieval Framework

Since the CPGs computerization framework, presented in chapter 4, enriches CPGs content with additional meta-information, adds context, semantics and transforms CPGs content into a set of computer interpretable concise segments, represented in XML. The idea behind CKR framework approach is to exploit the additional information introduced to enrich computerized CPGs content. It will support the contextually-relevant retrieval
of concise and focused segments of CPGs along with the medical literature linked with these segments, at point of care, for the corresponding information needs.

The CKR framework is based on a technique for contextual and statistical analysis of medical concepts. The underlying rationale of CKR framework approach is that:

(i) the semi-structured representation of medical documents (in our case computerized clinical practice guidelines), semantic tags, contextual information, and additional meta-information are relevant to CPGs content, which should be exploited for contextually relevant retrieval of CPGs content,

(ii) medical concepts are represented by medical phrases that make them highly specific in meaning, so, medical concepts should be used as index terms; since highly specific index terms produce high precision (Salton and McGill, 1983),

(iii) exhaustive indexing produces high Recall (Salton and McGill, 1983), so, information-rich yet efficient indexing strategy should be incorporated,

(iv) different medical concepts may have same meaning, therefore medical concepts should be standardized to enhance the search accuracy,

(v) a statistical model provides statistical information to help determine the relevance of documents with respect to a query, so statistical analysis technique should be devised.
The CKR framework is applied on the Ex-KCs knowledge base, which is created by the CPGs computerization technique (see chapter 4) that transforms CPGs content into Extended-Knowledge Components. However, CKR is designed to be generic for search and retrieval within the collection of semi-structured, structured, and indexed medical documents, for which certain ontological assumptions are satisfied, a situation that is most common in medical domains.

### 7.3.1 Functional Flow of Content and Knowledge Retrieval Framework

The CKR framework and its functional workflow are shown in Figure 7.1. In this section we describe the functional workflow of CKR while techniques and functionalities of each module will be presented, in detail, in forthcoming sections.

*Index Manager*, in the CKR framework, has three indices and it indexes the information from semi-structured documents, in our case Ex-KCs, which are stored in the EX-KCs knowledge base. It also provides information to different modules in the framework. *Ex-KCs Clustering* module receives optimized query in a form of medical concepts (medical phrases) defining the information needs. It creates clusters of the Ex-KCs based on the elements of a query set.
All the Ex-KCs having the medical concepts found in the query set are associated with one or more clusters. Context Analyzer module analyzes contextual structure of a particular medical concept in the corresponding Ex-KCs to provide total contextual weight, to a requesting module. Ex-KC objs Merger module is used to merge different Ex-KCs objects having same “KC ID” from different query-element-based clusters. So
the Ex-KCs objects having the same “KC ID” are merged into one object such that newly formed object represents the corresponding Ex-KC segment of the CPG.

*Weight Manager* uses a defined weighting strategy to calculate weights of the medical concepts found in the Ex-KCs and a query set. *Similarity and Ranking* module receives the required information from other modules and calculates the similarity between a query set and potential Ex-KCs and ranks them according to a defined scheme. Following we describe and discuss the functionalities and techniques of each module in details.

### 7.3.2 Index Manager and Indexing Strategy for Extended-Knowledge Components

An Index Manager populates and manages three indices. It accesses each Ex-KC from “Ex-KC knowledge base”, extracts the required information and prepares corresponding indices (as shown in Figure 7.1). It controls the flow of information ‘to and from’ indices. It communicates with *Ex-KCs Clustering module, Context Analyzer,* and *Weight Manger module* to provide information about Ex-KCs corresponding to a request.

Since Extended-Knowledge Components are transformed in XML format based on the GEM model, we analyzed the problems and potential solutions related to the search and retrieval of semi-structured documents (see section 2.4) in crafting indexing strategy for CKR framework. It has been stated earlier that Ex-KCs are enriched with the information relevant to CPGs content like context, semantic tags, meta-information from domain ontology (UMLS) and information provided by XML structure etc. To exploit such relevant information in Ex-KCs for contextually relevant retrieval of computerized
CPGs segments, indexing strategies for the following indices have been developed to index the different information contained in the Ex-KCs. These three indices are:

1. Inverted-Ex-KC-Index (I-Ex-KC-I)
2. Inverted-Path-Index (I-P-I)
3. Meta-Info-Index (M-I-I)

7.3.2.1 Inverted-Ex-KC-Index

The objective of Inverted-Ex-KC-Index (I-Ex-KC-I) is to index standardized medical concepts from the Ex-KCs, to keep track on medical concepts within Ex-KCs, and to store information of the Ex-KC to which medical concepts belong. In I-Ex-KC-I, every medical concept is linked to its corresponding element-ID and Extended-Knowledge Component ID. Our Elem-ID scheme (see section 4.3.5) is designed to capture Ex-KC structure (like XML structure) so, there is no need to capture hierarchical Ex-KC structure separately to keep track on medical concepts.

Storing element-ID with medical concepts helps analyze its structure (location) and helps assign its corresponding contextual weight. Figure 7.2 provides an example of an I-Ex-KC-I and shows that how in real system medical concepts and related information are indexed.
In Figure 7.2, “M-Cand” represents medical concepts, ‘Elem-ID’ represents the node/element to which corresponding medical concept belongs, and ‘KC-ID’ represents the Ex-KC in which the medical concept appears. For instance, medical concept ‘therapeutic procedure’ belongs to ‘Recommendation 2’ element, which is found in Ex-KC that has “ACS_06_za_KC3” ID, and this Ex-KC is a third ‘Knowledge Component’ of GEM-Encoded file “ACS_06_za”.

7.3.2.2 Inverted-Path-Index

The objective of an Inverted-Path-Index (I-P-I) is to keep track on the physical paths of original clinical practice guidelines, the GEM-Encoded file, the Ex-KC file, and to link related Ex-KCs. Figure 7.3 provides an example of an I-P-I and shows that how in real system different objects are linked and indexed. The I-P-I has five index fields:

(i) **Org-CPG**: This field indexes (stores) physical path of the original CPG.

(ii) **GEM-Encoded-file path**: This field stores the physical path of GEM-Encoded CPG.

(iii) **Ex-KC file Path**: This field stores the file path of the ‘Extended-Knowledge Components’,
(iv) **KC-ID:** It is used to index Ex-KC IDs and link it to original Ex-KCs.

(v) **KC-Link:** It is used to keep the information of related Ex-KCs or Ex-KC’s recommendations.

![Table showing KC-ID, GEM_Encoded file path, Ex-KC file path, KC-ID, and KC-Link](chart.png)

*Figure 7.3 Inverted-Path-Index example*

In Figure 7.3, for example, `ACS_06_zb_KC3` is an ID of an Ex-KC that is related to an Ex-KC with an ID `ACS_06_za_KC2`, while the location of its xml physical path is `//path/ACS_06_za_KC2.xml`, which is a part of GEM-Encoded CPG at physical location `//path/ACS_06_za.xml`. This GEM-Encoded CPG was transformed from original CPG physically located at `//path/filename`.

### 7.3.2.3 Meta-Info-Index

The ‘Meta-Info-Index’ (M-I-I) is designed to index and keep track on meta-information pertaining to the medical concepts. The M-I-I indexes the following meta-information for a corresponding medical concept:

- Semantic Type (ST)
- UMLS Score (Score)
- Vocabulary Sources (Vsources)
- Contextual Weight Coefficient (CW)
- Original Name (Oname)
Figure 7.4 provides an example of the M-I-I and shows that how data for each field is indexed and linked. For instance, a medical concept that has been standardized to "Myocardial Infarction" and indexed in ‘I-Ex-KC-I’ index (see Figure 7.2) has its original medical name “Heart attack” (appeared in CPG) indexed in M-I-I (see Figure 7.4). It has been indexed with its contextual weight (CW) (see section 4.3.6 for details) within the Ex-KC i.e. ‘1’. It has semantic type ‘Disease or Syndrome’ and its UMLS score is ‘1000’. Its standardized medical name was mapped from ‘SNOMEDCT’ thesaurus of UMLS.

In next section, we describe the formal representation of the knowledge base for Extended-Knowledge Components and optimized query.

### 7.3.3 Representation of Knowledge base for Extended-Knowledge Components and Query

It has been described that clinical practice guidelines are transformed into Ex-KCs by CPGs computerization framework. These Ex-KCs constitute the knowledge base of computerized CPGs. As Ex-KCs are transformed from the CPGs so, we can formally represent the CPGs knowledge as a set of Ex-KCs. Lets CPG is represented by $\phi$ and Ex-KC is represented by $\alpha$ so CPG is defined by equation 7.1. If the knowledge base is
considered as a search space, say \( \delta \) for ‘\( m \)’ number of CPGs, so it can be defined in terms of search space by equation 7.2. The search space based on equation 7.1 can be represented in terms of Ex-KCs by equation 7.3.

\[
\phi = \{ \alpha_1, \alpha_2, \ldots, \alpha_n \} \tag{7.1}
\]

\[
\delta = \{ \phi_1, \phi_2, \ldots, \phi_m \} \tag{7.2}
\]

\[
\delta = \{ (\alpha_1 \ldots \alpha_n), (\alpha_1 \ldots \alpha_n) \ldots (\alpha_1 \ldots \alpha_n) \} \tag{7.3}
\]

Extended-Knowledge Components are semi-structured medical documents containing standardized medical concepts. So, an Ex-KC can be formally represented in terms of standardized medical concepts by equation 7.4. In equation 7.4, Ex-KC is represented by \( \alpha \), medical concepts are represented by \( mp \) and the superscript “s” represents the standardized medical concepts.

\[
\alpha = \{ mp, mp, \ldots, mp \} \tag{7.4}
\]

The CKR framework receives user information needs in a form of a query set that is optimized by AMIPF framework (see chapter 6). A query set consists of medical concepts as medical practitioners define medical problems in medical phrases/medical terms. Medical concepts of a query set are also standardized during a query optimization process, as, standardizing query set and Ex-KCs medical concepts would enhance the retrieval performance. A query set, after query modeling and standardization, can be formally defined by equation 7.5. In equation 7.5, \( Q' \) is standardized query, “q” is a member of this set. The same ‘Query set’ in terms of medical concepts can be defined by
equation 7.6. In equation 7.6, $Q'$ denotes standardized query, $mp$ represents medical concepts, and "s" indicates that medical concepts are standardized. Every query set has a query type ($Q_{TYPE}$) (see section 5.3.2.8 for details).

$$Q' = \{q_n, \ldots, q_{n} \}$$ \hspace{1cm} 7.5

$$Q' = \{ mp_i, \ldots, mp_{i} \}$$ \hspace{1cm} 7.6

### 7.3.4 Extended-Knowledge Components Clustering

The CKR framework receives the optimized query set as information specifications from external module that is represented in Figure 7.1 as a *Query modeling and optimization module*. In our case, query set is processed and sent by AMIPF framework (see chapter 6 for details). The formal representation of a query set is given in equations 7.5 and 7.6. In Figure 7.1, a query set is shown ($Q'$) as an input to Ex-KCs Clustering Module. A query set’s members are represented by ‘$q_1,q_2,\ldots,q_n$’ (in Figure 7.1), which are in essence medical concepts ($mp$).

The Ex-KCs Clustering Module takes a query set ($Q'$) and creates clusters of Ex-KCs based on the query set elements. So the elements of a query set serve as seeds for Ex-KCs clusters. The Ex-KCs Clustering module works with Index Manager to find the associated Ex-KCs to query set’s elements (indicated by ‘2a’ in Figure 7.1). Every single Ex-KC found to have a query set element is associated with a cluster of that particular query set element. The query set elements clusters are formed using equation 7. 7. In Equation 7.7, $q_{ci}^{CL}$ defines Ex-KCs cluster based on query element ($q_i$), which is a member of a query set ($Q'$), the ‘$mp_i$’ denotes standardized medical concepts (query
elements $q_i$) whereas $m_{p_{a}}$ indicates standardized medical concepts corresponding to Ex-KC, and CL denotes a cluster.

\[
d_{mCL}^{q} = \begin{cases} 
1 & \text{if} \quad m_{p_{a}} = m_{p_{b}} \\
0 & \text{if} \quad m_{p_{a}} \neq m_{p_{b}} 
\end{cases} \tag{7.7}
\]

In equation 7.7, “1” means that a query set element (medical concept) has been matched with one of the medical concept in Ex-KC and “0” means the vice versa. Every Ex-KC, that is found to have a query set element, is attached to a query element cluster by an object. The Ex-KCs Clustering Module creates an object that has Ex-KC ID, medical concept matched with a query element, contextual weight ($\eta$) of the corresponding medical concept within the Ex-KC. So, in essence every query element has Ex-KC objects representing an Ex-KC in a query set element clusters. In this way, if an Ex-KC happens to have two or more medical concepts, found in the query set, can be attached to different query set element’s clusters by different objects having the same Ex-KC ID with additional information.

![Figure 7.5 Example for explanation of Ex-KCs objects for query set elements clusters](image-url)
For example, in Figure 7.5, query set elements q1 (myocardial infarction) and q2 (chest pain) both are found in Ex-KC with ID = 1. So, one Ex-KC object having same Ex-KC ID, medical concept (Med-con = Myocardial infarction in Figure 7.5) and its contextual weight is attached to q1 cluster and the other Ex-KC object having same Ex-KC ID with medical concept (chest pain) and its contextual weight in Ex-KC is attached to q2. In this manner all elements in a query set have none or some Ex-KCs objects associated to them in their clusters. The Ex-KCs Clustering module communicates with Context Analyzer module (indicated by ‘2b’ in Figure 7.1) to find the contextual weight ($\eta$) for the medical concept found in Ex-KC.

### 7.3.5 Context Analyzer

The Context Analyzer module’s one of the main functions is to analyze the contextual structure (context at different location within Ex-KC) of a particular medical concept (mp) in a corresponding Ex-KC and to assign contextual weight accordingly. It also calculates total contextual weight ($T\eta_{mp_i,n}$) for a specific medical concept (mp) in a corresponding Ex-KC and sends results to a requesting module. It uses equation 7.8 to calculate a total contextual weight of a medical concept in an Ex-KC. In equation 7.8, $T\eta_{mp_i,n}$ is a total contextual weight of a medical concept (mp) in ‘ith’ number of an Ex-KC ($\alpha$). The $\eta_{mp_i,j}$ is defined as a contextual weight of ‘jth’ medical concept (mp) in ‘ith’ number of Ex-KC ($\alpha$) at contextual structure (CTS). ‘N’ represents the number of time a medical concept (mp) is found in $\alpha$. So, total contextual weight is a sum of all contextual weights (at different locations) of a medical concept in an Ex-KC.
The Context analyzer works with three modules (i) ‘Ex-KC Clustering module’, (ii) ‘Ex-KC objs Merger module’, and (iii) ‘Index Manager’ (indicated by ‘2b, 4a’ in Figure 7.1). It takes request from Ex-KC clustering module to find contextual weight of a particular medical concept. It takes request from Ex-KC objs Merger module to find a total contextual weight of a medical concept. It makes use of Index Manager during context analysis and in assigning contextual weight.

7.3.6 Extended-Knowledge Components Objects Merger

The Extended-Knowledge Components Objects Merger (Ex-KC objs Merger) module is used to merge Ex-KC objects, having the same ID, from different query set element clusters. We have developed the heuristic for this process and equation 7.9 formally defines the merging process of the Ex-KC objects. In equation 7.9 “merged Ex-KC object” is denoted by \( \alpha^{\text{mg}} \), where N is a number of query set element clusters, and ‘CL’ represents a cluster.

\[
\alpha^{\text{mg}} = \left\{ \bigcup_{j=N-1}^{j=N} ((q_{i\in Q}^{\text{CL}} \cap q_{2\in Q}^{\text{CL}}) \cap (q_{j\in Q}^{\text{CL}} \cap q_{j\in Q}^{\text{CL}})) \right\}
\]  

\[ T \eta_{mp \ a_i} = \sum_{j=1}^{N} \eta_{mp \ a_i}^{CTS} \]
A merging of ‘Ex-KC objects’ is performed based on ‘Ex-KC ID’. Merging of objects starts by picking any Ex-KC object in one cluster and its Ex-KC ID is searched in the Ex-KC objects of other clusters. If an ‘Ex-KC object’ is found with common ‘Ex-KC ID’, it is merged by taking its all information (medical concept, corresponding contextual weight).

After merging the resulting ‘Merged Ex-KC object’ is kept and the other Ex-KC objects are removed. In this manner all Ex-KC objects in different clusters with same ‘Ex-KC ID’ are merged into one ‘Merged Ex-KC object’. The Ex-KC objs Merger module also communicates with Context Analyzer module to add total contextual weight.
of each medical concept in a Merged Ex-KC object. Figure 7.6 illustrates a merging process of Ex-KC objects with reference to Figure 7.5. Merged Ex-KC object contains all the distinct medical concepts (mp) along with their contextual weight, total contextual weight and Ex-KC ID. So, merged Ex-KC object contains at least one and at maximum all query set elements (medical concepts). The collection of the merged Ex-KC objects gives potential Ex-KCs that may be retrieved for the query set. This collection of merged Ex-KC objects is sent to Weight Manager module.

7.3.7 Weight Manager

The Weight Manger main task is to calculate weights of the medical concepts found in the Ex-KCs and a query set. These weights are calculated by our developed strategy and heuristic. The Weight Manager receives all merged Ex-KC objects. It takes each object, extracts its ‘Ex-KC ID’, and assigns weights for all medical concepts found in that Ex-KC.

As each Ex-KC is considered as a separate entity mimicking a separate document, on the other hand, each Ex-KC has additional meta-information like context etc. During the weight assignment and similarity measure processes, we want to exploit the meta-information available for CPGs medical concepts. We have divided medical concepts (mp) of the potential Ex-KCs, which may be retrieved, in two categories for weighting process. Category I: medical concepts of the potential Ex-KCs that did not match with query set elements, and Category II: medical concepts of potential Ex-KCs that matched with query set element (those which are found in merged Ex-KC objects).

We first describe the weight calculation of those medical concepts that belong to Category I. Medical concepts belong to this category are assigned statistical weight
(weight calculated based on the statistics of medical concepts within Ex-KC). Statistical
weight of a medical concept is calculated using equation 7.10 that has been adapted to
suit Ex-KCs from (Salton and Buckley, 1991; Salton and McGill, 1983). In equation
7.10, $w_{mp,a}^s$ is a statistical weight where “s” represents the statistical weight of medical
concept (mp) in Ex-KC ($\alpha$), $f_{mp,a}$ is a frequency of the particular medical concept (mp) in
an Ex-KC ($\alpha$), and $i\alpha_f$ is the inverse Ex-Kc frequency. The frequency ($f_{mp,a}$) of a
medical concept (mp) in an Ex-KC ($\alpha$) is defined in equation 7.11 where N is a number
of occurrences of a mp in the $\alpha$. The inverse Ex-Kc frequency ($i\alpha_f$) is calculated by
equation 7.12.

$$W_{mp,a}^s = f_{mp,a} \times i\alpha_f$$  \hspace{1cm} 7.10

$$f_{mp,a} = N(mp,a)$$  \hspace{1cm} 7.11

$$i\alpha_f = \log\left(\frac{\lambda_a}{\gamma_{mp,a} + 1}\right) + 1$$  \hspace{1cm} 7.12

In equation 7.12, $\lambda_a$ is the total number of Ex-KCs in the ‘Ex-KCs knowledge
base’ and $\gamma_{mp,a}$ is the total number of Ex-KCs in which particular medical concept (mp)
from a query set is found. By using these three equations statistical weight for each
medical concept, which belongs to Category I, is calculated.

Medical concepts that belong to Category II are assigned “Hybrid Weight”. A
Hybrid Weight is a sum of total contextual weight and a statistical weight of a medical
concept. It is defined in equation 7.13 where “Hybrid Weight” of a medical concept (mp)
is represented by $W_{mp,a}^H$, “H” indicates “hybrid”, and $W_{mp,a}^s$ is a statistical weight (see
The total contextual weight \( (T\eta_{mp_a}) \) of a corresponding medical concept is calculated by equation 7.8.

\[
W^{H}_{mp_a} = T\eta_{mp_a} + W^{S}_{mp_a}
\]

Index Manager calculates weights of all medical concepts of both categories for all potential Ex-KCs (identified with merged objects Ex-KC ID). We are using vector space model for the retrieval of relevant Ex-KCs, so, Index Manager creates Hybrid weight vectors for all potential Ex-KCs. Weight vector represents a document (in our case Ex-KC) in weights of its content (in our case medical concepts). A Hybrid weights vector \( \alpha^{H,w} \) is defined in equation 7.14. In equation 7.14, \( \alpha^{H,w} \) is hybrid weight vector of an Ex-KC \( (\alpha) \), \( W^{H}_{mp} \) denotes a hybrid weight for ‘ith’ medical concept for Category II, ‘n’ is a number of medical concepts that belong to the Category II, while ‘k’ is a number of medical concepts that belong to the Category I and the statistical weight of this category medical concept is denoted by \( W^{S}_{mp} \).

\[
\alpha^{H,w} = \left\{ (W^{H}_{mp_1}, \ldots, W^{H}_{mp_n}), (W^{S}_{mp_1}, \ldots, W^{S}_{mp_k}) \right\}
\]

After calculating the Hybrid weight vectors for all potential Ex-KCs, Weight Manger sends them to ‘Similarity and Ranking module’.

### 7.3.8 Similarity and Ranking module

The ‘Similarity and Ranking’ module task is to find the similarity between a query set and the potential Ex-KCs and to rank them relatively. It receives Hybrid weight vectors of the Ex-KCs from Weight Manager and a query set \( (Q) \) from query modeling.
and optimization module. It communicates with the Weight Manager to calculate the weights of the query set’s elements (mp: medical concepts) and a corresponding Hybrid weight vector for a query set. The Hybrid weight vector of a query set represents it in terms of maximum Hybrid weights of the corresponding medical concepts. The query set’s Hybrid weight vector is transformed using equation 7.15. In equation 7.15, $Q^{i,H}$ represents the Hybrid weight vector of a query set ($Q'$), $N$ represents the number of query elements in a $Q'$, and $W_{q_i,a}^{M,H}$ represents the maximum Hybrid weight of the query set element ($q_i = mp =$ medical concept) from an Ex-KC ($\alpha$) that has the maximum hybrid weight for that particular $q_i$. For example, say one of the medical concepts of a query set is “chest pain” and three Ex-KCs are having “chest pain” with its Hybrid weight 1.75, 2.0, 1.50 respectively. So, 2.0 will be taken as a Hybrid weight for the query set medical concept “chest pain”.

$$Q^{i,H} = \left\{ \sum_{i=1}^{N} W_{q_i,a}^{M,H}, ..., W_{q_N,a}^{M,H} \right\}$$

The ‘Similarity and Ranking’ module implements our similarity measure algorithm given in Table 7.1.
Table 7.1: Similarity Measure Algorithm for CKR Framework

<table>
<thead>
<tr>
<th>Similarity Measure Algorithm for CKR Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Step1: Get ‘Hybrid weight’ vector $\alpha_i^{H.w}$ (defined in equation 7.14) for Ex-KC ($\alpha_i$),</td>
</tr>
<tr>
<td>• Step2: Calculate ‘Hybrid weight’ vector length of Ex-KC ($\alpha_i$) using an equation 7.16,</td>
</tr>
<tr>
<td>• Step3: Calculate ‘Hybrid weight’ vector length of $\mathcal{Q}_i$ using equation 7.17,</td>
</tr>
<tr>
<td>• Step4: Calculate dot product of ‘Hybrid weight’ vector of Ex-KC ($\alpha_i$) and ‘Hybrid weight’ vector of $\mathcal{Q}_i$ using equation 7.18,</td>
</tr>
<tr>
<td>• Step5: Calculate similarity between $\mathcal{Q}_i$ and $\alpha_i$ using equation 7.19.</td>
</tr>
</tbody>
</table>

\[ |\alpha_i|^H = \sqrt{\sum_{i=1}^{n} (W_{mp_i}^H)^2 + \sum_{j=1}^{k} (W_{mp_j}^S)^2} \]  

7.16

The equation 7.16 is used to calculate the length (modulus) of a Hybrid weight vector of an Ex-KC ($\alpha_i$). In the equation 7.16, $|\alpha_i|^H$ is a hybrid weight vector length of an Ex-KC ($\alpha_i$), $W_{mp_i}^H$ is a “Hybrid Weight” of a medical concept (mp) in the Ex-KC ($\alpha_i$), $W_{mp_j}^S$ is a statistical weight of a medical concept in the Ex-KC ($\alpha_i$).

\[ |\mathcal{Q}_i|^H = \sqrt{\sum_{i=1}^{N} (W_{m_i}^{M,H})^2} \]  

7.17
The equation 7.17 is used to find the modulus of the Hybrid weight vector of a query set \( (Q') \). In the equation 7.17 \( |Q'|^H \) is a ‘Hybrid weight’ vector length of a query set \( (Q') \) and \( W_{q_i, a}^{M, H} \) represents the maximum Hybrid weight of a query set element \( (q_i = mp = \text{medical concept}) \) from an Ex-KC \( (\alpha') \) that has the maximum Hybrid weight for that particular \( q_i \).

\[
|Q'|^H = \sum_{i=1}^{N} W_{q_i, q'}^{M, H} W_{mp, a}^H
\]

7.18

The equation 7.18 is used to find a value of the dot product for a Hybrid weight vector of an Ex-KC \( (\alpha') \) and a Hybrid weight vector of a query set \( (Q') \).

\[
\text{Sim}_{(Q', \alpha')} = \frac{|Q'.\alpha'|^H}{|Q'|^H |\alpha'|^H}
\]

7.19

The similarity between a query vector and an Ex-KC vector is determined by finding the angle between them, which is calculated by cosine measure. The equation 7.19 is used to find similarity between the query and the Ex-KC. The equation 7.20 is derived by putting values of \(|Q'.\alpha'|^H\), \(|Q'|^H\), and \(|\alpha'|^H\) in equation 7.19 for the similarity measure calculation. In equation 7.20, \( \text{Sim}_{(Q', \alpha')} \) represents the similarity value of both

\[
\text{Sim}_{(Q', \alpha')} = \frac{\sum_{i=1}^{N} W_{q_i, q'}^{M, H} W_{mp, a}^H}{\sqrt{\sum_{i=1}^{N} (W_{q_i, q'}^{M, H})^2} \sqrt{\sum_{i=1}^{N} (W_{mp, a}^H)^2 + \sum_{j=1}^{k} (W_{mp, a}^S)^2}}
\]

7.20
Hybrid weight vector of the Ex-KC ($\alpha$) and Hybrid weight vector of the query set $Q'$. The value of this similarity lies within the range of [0 to 1]. The value closer to 1 indicates more similarity of an Ex-KC ($\alpha$) to a query set ($Q'$).

Once all Ex-KCs are retrieved, they are ranked based on two factors, one is query type ($Q_{TYPE}$) of a query set (if query type is given for a query set) and second is the relative similarity measure values of the retrieved Ex-KCs. In case where a query type ($Q_{TYPE}$) is provided for a query set, all retrieved Ex-KCs that belong to that $Q_{TYPE}$ are taken first and ranked based on their relative similarity measure values. The rest of Ex-KCs (that do not belong to query type) are ranked based on their relative similarity measure values after the Ex-KCs that belong to $Q_{TYPE}$. If a query type is not provided for a query set, then its query type is considered as a general query type. All retrieved Ex-KCs for such a query set are ranked based on their relative similarity measure values. For example, an Ex-KC having similarity measure value closer to 1 is ranked number one and other Ex-KCs are ranked accordingly.

7.4 Clinical Knowledge Assistance System Architecture

Clinical Knowledge Assistance (CKA), a computer system, offers evidence-based clinical practice by providing relevant evidence-based medical information and knowledge to healthcare practitioners at their decisions points. It assists healthcare practitioners in better interpretation and understanding of clinical referral/response letters. The CKA system is based on an architecture that is presented in Figure 7.7. The architecture consists of eight layers and incorporates the techniques and strategies developed in four frameworks presented earlier in this thesis.
The eight layers are (1) Interface Layer, (2) Transport Layer, (3) Information Processing Layer, (4) Knowledge Description Layer, (5) Modeling and computerization Layer, (6) Knowledge Linking Layer, (7) Retrieval Layer, and (8) Object Layer.

- The Interface layer deploys a web interface for user to interact with underlying system.
- The Transport layer provides a channel for a web interface and rest of the system to send and receive data and messages.
The Information processing layer deploys the techniques and algorithms developed in AMIPF for medical information processing and standardization.

The Knowledge description layer deals with knowledge representation and formalization.

The Modeling and computerization layer deals with the clinical practice guidelines computerization.

The Knowledge linking layer deals with automatic generation of queries from computerized clinical guidelines segments and their linking to relevant evidence-based online medical literature.

The Retrieval layer deploys the techniques to provide access of relevant evidence-based knowledge for the information specifications of healthcare practitioners.

The Object layer defines the multi-modal knowledge and information organization for the CKA healthcare knowledge base.

7.4.1 Functional Overview of the CKA System Architecture

The functional overview of CKA system architecture is described along the lines of following processes.

- Computerization of CPGs,
- Linking of Ex-KCs to medical literature,
- Customized query formulation,
- Retrieval and presentation of computerized CPGs segments and linked literature to the user.

The CKA system architecture deploys CPGs computerization framework’s techniques at “Modeling and Computerization layer”. The CPGs are computerized to be transformed into Ex-KCs. In CKA architecture, modules at three layers are involved for the creation of Ex-KCs. In Figure 7.7, the “CPGs computerization” module (at Modeling and Computerization layer) works with GEM-encoded CPGs, Ex-KC Ontology, and the UMLS metathesaurus defined at “Knowledge Description layer”. It interacts with the UMLS metathesaurus via UMLS Manager. It stores and organizes Ex-KCs in the knowledge base at “Object layer”.

Once the Ex-KC knowledge base is populated, Ex-KCs are linked to the online relevant evidence-based medical literature. Linking of Ex-KCs with relevant online medical literature is performed by the modules at three layers. The module at “Knowledge Linking layer” deploys the techniques developed in the CQGF framework. It takes Ex-KCs (one by one) from the Knowledge base at “Object layer”, uses “Diseases ontology” at “Knowledge Description layer” to automatically generate a query from an Ex-KC. This generated query is sent through a “PubMed Webservice module” to retrieve medical literature from the MEDLINE and link them with the corresponding Ex-KC.

The customized query to represent the information need of a user is created by the modules located at three layers. A referral/response letter is sent by “web interface query specification” module (at interface layer) to the medical information processing module at “Information Processing layer”. The module at Information processing layer deploys the AMIPF framework techniques. This module uses Diseases ontology at “Knowledge
Description layer” and works with UMLS metathesaurus through UMLS Manager. It processes the medical information contained in a letter transforms them into standardized medical concepts and categorizes them into five classification. Once a letter is processed it is sent back to “Interface layer” module. If any new medical concept is submitted by user during the formulation of customized query, it is transformed to its standardized medical name through “Medical Information Processing and Standardization” module.

Access to relevant computerized CPGs segments and linked medical literature is made possible by the modules at two layers. The module at “Retrieval layer” deploys the techniques and algorithm developed in the CKR framework. It receives customized user query from a “Medical Information Processing and Standardization” module. It uses Index Manger at “Object layer” to retrieve relevant computerized CPGs segments and their linked medical literature. The Index Manger indexes all the important information of the Ex-KCs according to the indexing strategy defined earlier in this chapter. The retrieved Ex-KCs and linked medical literature are submitted to the Presentation module to format them according to the presentation scheme and send them to the module at “Interface layer”. Appendix F shows screen shots presenting the functionalities of the developed Clinical Knowledge Assistance system based on this architecture.

7.5 Summary

Digital literature in general and medical literature in particular is growing exponentially. Effective searching and retrieval for accessing relevant content from medical document is crucial at point of care. A new paradigm shift in the practice of evidence-based medicine has set the trend to use clinical practice guidelines at point of care in daily practice, which further emphasizes the provision of automated support to
guidelines-based clinical care. However, healthcare practitioners require access to the segments of knowledge content from CPGs that are relevant to the user information specifications at point of care.

We have developed and described our “Content and Knowledge Retrieval” (CKR) framework for retrieving and accessing the contextually relevant segments of clinical practice guidelines.

The underlying aim of the CKR framework is the retrieval of contextually relevant segments of semi-structured, indexed medical documents with high ‘Recall and high ‘Precision’. We have devised our indexing scheme that exploits the rich-information contained in the Ex-KCs to assist the contextually and semantically relevant search and retrieval. We have presented and discussed the technique and algorithms for each module in the CKR framework in details. The CKA system architecture has been presented and discussed that deploys the techniques and algorithms developed in the four frameworks.
CHAPTER 8

EVALUATION AND DISCUSSIONS

8.1 Introduction

In this chapter, we present evaluation of the work presented in this thesis. Evaluation for each framework described in chapter 4, 5, 6 and 7 has been performed along the lines of qualitative and quantitative measures. The CPGs Computerization Framework presented in chapter 4 has been qualitatively evaluated for knowledge modeling using Encoding Strategy and automatic processing of Gem-encoded guidelines to create Ex-KCs segments from CPGs content. The evaluation of Context Specific Query Generation Framework (CQGF) has been performed qualitatively and results have been quantified for the analysis. The CQGF presented in chapter 5 has been evaluated for generating relevant clinical queries, to find correct query types, and to find relevant medical literature from the MEDLINE. The Automatic Medical Information Processing Framework (AMIPF) has been evaluated along the lines of quantitative evaluation measures. The AMIPF evaluation has been performed for potentially significant medical concepts Filtration performance, General Classification performance, and Cardiology Classification performance of the filtered medical concepts. The Content and Knowledge Retrieval Framework (CKR) evaluation is presented along the lines of quantitative measures. The CKR evaluations has been conducted to measure the overall Recall and Precision of the retrieved Ex-KCs and to measure the relevancy of top ranking retrieved Ex-KCs to corresponding information needs. In next section, first CPGs Computerization
framework evaluation and discussion are presented, which is followed by the CQGF framework evaluation and discussion. Proceeding to the next section evaluation of the AMIPF is discussed and towards the end of this chapter, the CKR framework evaluation and discussion are presented.

8.2 Evaluation of Clinical Practice Guidelines Computerization Framework

The CPGs computerization framework transforms guidelines into Extended-Knowledge Components (Ex-KCs). This process of computerization mainly consists of two steps, an encoding step for the modeling of CPGs knowledge and an automatic processing step for creating Ex-KCs segments from encoded CPGs content. To evaluate our method of CPGs computerization, qualitative analysis was performed by domain expert at two levels. At the first level, an evaluation was performed for Encoding strategy used for knowledge modeling. At the second level, an evaluation was performed for automatic processing of CPGs computerization framework to create Extended-Knowledge Components.

Clinical practice guidelines that have been computerized using CPGs computerization framework are given in Table 8.1 with their full name, year and country of origin. The evaluation of CPGs computerization framework was performed by randomly selecting fifty Ex-KCs from the computerized guidelines.
Table 8.1 Clinical practice guidelines selected for computerization

<table>
<thead>
<tr>
<th></th>
<th>Guidelines for the management of acute coronary syndromes 2006</th>
<th>2006</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Hypertension Management Guide for Doctors</td>
<td>2004</td>
<td>Australia</td>
</tr>
<tr>
<td>3</td>
<td>Recommended Framework for Cardiac Rehabilitation</td>
<td>2004</td>
<td>Australia</td>
</tr>
<tr>
<td>4</td>
<td>Consensus statement for the prevention of vascular disease</td>
<td>2004</td>
<td>Australia</td>
</tr>
<tr>
<td>5</td>
<td>Primary Care Management of Chronic Stable Angina and Asymptomatic Suspected or Known Coronary Artery Disease: A Clinical Practice Guideline from the American College of Physicians</td>
<td>2004</td>
<td>USA</td>
</tr>
<tr>
<td>6</td>
<td>Evaluation of Primary Care Patients with Chronic Stable Angina: Guidelines from the American College of Physicians</td>
<td>2004</td>
<td>USA</td>
</tr>
<tr>
<td>7</td>
<td>Management of Patients With ST-Elevation Myocardial Infarction</td>
<td>2004</td>
<td>USA</td>
</tr>
<tr>
<td>8</td>
<td>Canadian recommendations for the management of hypertension</td>
<td>1999</td>
<td>Canada</td>
</tr>
<tr>
<td>9</td>
<td>Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and management</td>
<td>2006</td>
<td>Canada</td>
</tr>
<tr>
<td>10</td>
<td>Recommendations for the management of dyslipidemia and the prevention of cardiovascular disease: 2003 update</td>
<td>2003</td>
<td>Canada</td>
</tr>
</tbody>
</table>

At the first level of evaluation, we are looking for the following answers:

- Whether all important information was encoded during modeling process?
- Was the modeling using proposed Encoding strategy complex or not?
- Was the modeling process generic enough to encode different types of guidelines?
- Was the technical knowledge required during computerization of CPGs?

Results of the first level evaluation of the CPGs computerization framework are shown in Table 8.2. To determine measures of evaluation defined by the questions stated above, Table 8.2 has been divided into four major columns. Results in Table 8.2 shows that all important information in CPGs is encoded using Encoding strategy of CPGs computerization framework. The process of encoding CPGs is not complex and it is generic enough to encode different types of CPGs. There is no additional technical
knowledge required to encode CPGs, it means medical professionals without any computer technical background could easily encode CPGs.

Table 8.2 Evaluation Results of the encoding step for modeling of CPG knowledge

<table>
<thead>
<tr>
<th>No.</th>
<th>Guidelines for the management of acute coronary syndromes 2006</th>
<th>Complete Process</th>
<th>General Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agree</td>
<td>Yes</td>
<td>Agree</td>
</tr>
<tr>
<td>2</td>
<td>Hypertension Management Guide for Doctors</td>
<td>Yes</td>
<td>Agree</td>
</tr>
<tr>
<td>3</td>
<td>Recommended Framework for Cardiac Rehabilitation</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>4</td>
<td>Consensus statement for the prevention of vascular disease</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>5</td>
<td>Primary Care Management of Chronic Stable Angina and Asymptomatic Suspected or Known Coronary Artery Disease: A Clinical Practice Guideline from the American College of Physicians</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>6</td>
<td>Evaluation of Primary Care Patients with Chronic Stable Angina: Guidelines from the American College of Physicians</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>7</td>
<td>Management of Patients With ST-Elevation Myocardial Infarction</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>8</td>
<td>Canadian recommendations for the management of hypertension</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>9</td>
<td>Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and management</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
<tr>
<td>10</td>
<td>Recommendations for the management of dyslipidemia and the prevention of cardiovascular disease: 2003 update</td>
<td>Yes</td>
<td>Partially Agree</td>
</tr>
</tbody>
</table>

Evaluation for the second level was performed with the same data set. The objective of the second level evaluation was to measure:

- Was the CPGs content represented correctly in Extended-Knowledge Components (Ex-KCs) after computerization?
- Were significant medical concepts extracted and context, semantic and meta-information added to them?
- Were all the medical concepts standardized?
Results for the second level evaluation are shown in Table 8.3 to show the answers, which consists of three columns. The first column indicates the opinion of an expert on whether Ex-KCs represent CPGs content or not. The second column indicates whether, medical concepts are standardized and semantics, context, and meta-information are added or not. The third column shows if significant medical concepts are extracted or not during the creation of Ex-KCs. In this column significant medical concepts correspond to the medical terms in “Decision Variable Elements”, “Recommendation Elements” and “Imperative Elements” of the Ex-KCs.

Table 8.3 Evaluation Results of CPG Computerization Framework for Extended-Knowledge Components Creation

<table>
<thead>
<tr>
<th>No.</th>
<th>CPG-Content Representation in Ex-KC</th>
<th>Semantics, Context, Standardized Meta-Information</th>
<th>Significant Medical Concepts Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
<td>Partially Agree</td>
<td>Disagree</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Guidelines for the management of acute coronary syndromes 2006
- Hypertension Management Guide for Doctors
- Recommended Framework for Cardiac Rehabilitation
- Consensus statement for the prevention of vascular disease
- Primary Care Management of Chronic Stable Angina and Asymptomatic Suspected or Known Coronary Artery Disease: A Clinical Practice Guideline from the American College of Physicians
- Evaluation of Primary Care Patients with Chronic Stable Angina: Guidelines from the American College of Physicians
- Management of Patients With ST-Elevation Myocardial Infarction
- Canadian recommendations for the management of hypertension
- Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and management
- Recommendations for the management of dyslipidemia and the prevention of cardiovascular disease: 2003 update

230
8.3 Evaluation of Context Specific Query Generation Framework (CQGF)

The CQGF automatically generates queries from computerized clinical practice guidelines and determines query types for corresponding queries. These queries and query types are submitted to PubMed to retrieve and link relevant medical literature with the corresponding computerized CPGs segments. To evaluate CQGF technique, computerized CPGs in form of EX-KCs were processed with the CQGF framework to generate clinical queries, to find query types, to retrieve medical literature from MEDLINE and link them to the corresponding Ex-KCs. Domain expert (specialists from cardiology domain) was asked to analyze and evaluate the results in the following settings.

First, the generated clinical queries were evaluated based on two measures (i) Relevant and (ii) Not relevant, corresponding to their relevancy to Ex-KCs, ‘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 or is not relevant to the Ex-KC. Fifty Ex-KCs of above mentioned clinical guidelines (in Table 8.1) were taken for the evaluation. Table 8.4 shows the results of the evaluation of the, automatically generated queries from Ex-KCs for two measures of relevancy.

<table>
<thead>
<tr>
<th>Ex-KCs Queries numbers</th>
<th>Relevant</th>
<th>Not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Another evaluation for automatically generated queries was conducted along the lines of three measures of relevancy: (i) Relevant, (ii) Partially Relevant, and (iii) Not-
relevant. For this evaluation ‘Relevant’ indicates that generated query exactly represents the CPGs content, ‘Partially relevant’ indicates that generated query is relevant but medical terms may be added or deleted, and ‘Not relevant’ indicates that generated queries do not represent CPGs content at all. Table 8.5 shows the results for this evaluation.

Table 8.5 Results of automatic query generation for three measures of relevancy Ex-KCs

<table>
<thead>
<tr>
<th>Ex-KCs Queries numbers</th>
<th>Relevant</th>
<th>Partially relevant</th>
<th>Not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 28</td>
<td>% 56%</td>
<td>No. 22</td>
<td>% 44%</td>
</tr>
<tr>
<td>No. 0</td>
<td>% 0%</td>
<td>No. 0</td>
<td>% 0%</td>
</tr>
</tbody>
</table>

To evaluate the ‘Query Type’ generated by the CQGF for clinical queries, domain expert was asked to provide their feedback for two measures of evaluation: (i) Agree, and (ii) Disagree. Table 8.6 shows the results for this evaluation.

Table 8.6 Results of query type finding Process for two evaluation measures

<table>
<thead>
<tr>
<th>Total No. of query types</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10%</td>
</tr>
</tbody>
</table>

We conducted another evaluation for the “Query Type” for three measures of evaluation: (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 8.7 shows the results for this evaluation. The CQGF assigned some of the queries two query types e.g. Diagnosis and Therapy. In such cases, domain expert classified the Query Type under Agree with reservation category.
The 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 each query to find if they are relevant to corresponding CPGs segments. For this analysis maximum twelve retrieved medical articles for each query were analyzed based on their title and abstract. Retrieved medical articles were evaluated for two evaluation measures (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 the corresponding CPGs segments, and ‘Not-relevant’ indicates that retrieved articles are not relevant to the corresponding CPGs segments. Table 8.8 shows the results of this evaluation.

<table>
<thead>
<tr>
<th>Total No. of query types</th>
<th>Agree</th>
<th>Agree with reservation</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>50</td>
<td>39</td>
<td>78%</td>
<td>6</td>
</tr>
</tbody>
</table>

On reviewing the evaluation by domain experts, we found that queries and query types generated for the 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.
8.4 Evaluation of Automatic Medical Information Processing Framework (AMIPF)

The automatic medical information processing framework (AMIPF) processes medical referral/response letter information at three levels. The first level is Filtration of potential significant medical terms, the second level is classification of general medical terms, and the third level is Cardiology Classification of medical terms. The output from third level is the main results of the AMIPF.

To evaluate, the techniques of AMIPF, we performed evaluation at three levels.

- First at the Filtration level—filtration of correct medical concepts,
- Second at the General Classification level—classification of medical concepts into general five categories, and
- Third at Cardiology Classification level—classification of cardiology related medical concepts into five categories.

The Recall, Precision and F-measure were evaluation measures for each level. In AMIPF, we are using UMLS tool (MMTX) for extracting medical concepts from letters that extracts quite a large number of terms as medical terms. Some of those are not significant medical terms and some are not even medical terms. For Filtration level evaluation, medical terms, and phrases were manually identified and marked in each referral letter by domain expert. These manually marked medical terms and phrases were taken as a reference standard for filterable medical terms and phrases. This reference standard is used to determine the correct number of medical terms filtered by our filtration technique. These medical terms were matched against the AMIPF filtered
medical terms. Medical terms which were filtered correctly by AMIPF are counted as True Positive (TP), correct medical terms which were missed are counted as False Negative (FN), and medical terms which were filtered wrongly are counted as False Positive (FP).

Using TP, FP, and FN, three measures for evaluation were calculated for each letter, which are Recall, Precision and F-measures. Recall indicates sensitivity or true positive rate of correctly filtered medical terms in a letter. In other words it indicates how many correct medical concepts are filtered. It is defined by equation 8.1. Precision indicates the true predictive value of filtered medical concepts in a letter. In other words it tells that in filtered concepts how many are correct medical concepts and it is defined by equation 8.2. F-measure determines the harmonic mean of precision and recall that indicates the overall test accuracy of AMIPF filtration.

\[
\text{Recall} = \frac{TP}{TP+FN} \tag{8.1}
\]

\[
\text{Precision} = \frac{TP}{TP+FP} \tag{8.2}
\]

\[
\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{8.3}
\]

The referral letters were obtained by a tedious process of getting patient consents and by going through ethics approval procedure. On the other hand manually identifying and marking medical concepts to create reference standard and comparing them with system processed medical information was a time-consuming task. So, for this evaluation we randomly selected eighteen letters consisting of 28455 words.

The Filtration in AMIPF is performed on extracted medical concepts from the referral and response letters. If during extraction any medical concept was not extracted,
but it was present in a letter’s standard reference, it was counted as false negative (FN) for filtration process.

All medical referral letters were processed by the AMIPF and TP, FP, FN were counted. Recall, Precision and F-measure for filtration of each letter were calculated. We calculated Average Recall, Average precision and Average F-Measure for an estimation of AMIPF filtration performance. Table 8.9 shows the results of Recall, Precision, and F-measure of each letter for Filtration process and average Recall, average Precision, and average F-measure of Filtration process of the AMIPF. In table 8.9, Column “LT” shows the number of letter, “Ex” indicates the number of extracted terms from a letter as potential medical terms, TP shows the true positive, FP represents false positive, FN represents, false negative. Column “Rec” represents Recall, “Prec” represents Precision, “Fmes” represents F-measure and “Mx-ER FN” indicates error caused by MMTX during extraction along with false negative number. This error represents error in recall.

The AMIPF average Recall value for filtration is 0.92 (92%), average precision is 0.86 (86%) and the F-measure is 0.89 (89 %). The total Average error in Recall is 0.8 (8%). To estimate the error in Recall caused by MMTX (UMLS tool) and AMIPF filtration technique, we calculated the number of medical concepts not extracted by MMTX, say “M” and the number of extracted medical concepts did not filter by AMIPF filtration technique, say “Fl”. In Table 8.9, a column M-Er FN shows false negative (FN) number for each letter caused by the MMTX and corresponding error due to this FN. The average Recall error caused by MMTX is 0.078 (7.8%). It indicates a large number of Recall error has MMTX factor involved in the total Filtration Recall error.
For General classification evaluation, a reference standard was created by classifying medical terms (selected as filterable terms for Filtration evaluation) into five general categories for each referral letter.

The General classification evaluation of AMIPF was performed by using Recall, Precision, and F-measure. For General Classification, Recall measures the rate of medical terms classified for a corresponding category. It also indicates the number of medical terms missed from being classified. Recall is calculated for every category—General Symptoms, General Diseases, General medications, General Diagnosis and General Therapies, of each letter. Precision measures the correctly classified medical terms within classified medical terms for a corresponding category. F-measure indicates the accuracy of the General classification of the AMIPF. TP, FN, FP for the General classification were determined as: medial terms correctly classified are counted as TP, medical terms missed from being classified are counted as FN, and medial terms wrongly classified are counted as FP.

### Table 8.9 Results of Referral Letters Recall, Precision, and F-Measure and AMIPF Filtration Average

<table>
<thead>
<tr>
<th>#</th>
<th>Recall</th>
<th>Precision</th>
<th>F-Measure</th>
<th>AMIPF Filtering Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.84</td>
<td>0.90</td>
<td>0.87</td>
<td>0.14</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>0.87</td>
<td>0.875</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>0.91</td>
<td>0.89</td>
<td>0.89</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.11</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>0.71</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.84</td>
<td>0.89</td>
<td>0.87</td>
<td>0.16</td>
</tr>
<tr>
<td>7</td>
<td>0.91</td>
<td>0.86</td>
<td>0.89</td>
<td>0.09</td>
</tr>
<tr>
<td>8</td>
<td>0.84</td>
<td>0.90</td>
<td>0.87</td>
<td>0.11</td>
</tr>
<tr>
<td>9</td>
<td>0.92</td>
<td>0.79</td>
<td>0.85</td>
<td>0.08</td>
</tr>
<tr>
<td>10</td>
<td>0.92</td>
<td>0.85</td>
<td>0.88</td>
<td>0.08</td>
</tr>
<tr>
<td>11</td>
<td>0.95</td>
<td>0.81</td>
<td>0.88</td>
<td>0.025</td>
</tr>
<tr>
<td>12</td>
<td>1.00</td>
<td>0.77</td>
<td>0.87</td>
<td>0.00</td>
</tr>
<tr>
<td>13</td>
<td>0.93</td>
<td>0.85</td>
<td>0.89</td>
<td>0.07</td>
</tr>
<tr>
<td>14</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>15</td>
<td>0.86</td>
<td>0.91</td>
<td>0.89</td>
<td>0.14</td>
</tr>
<tr>
<td>16</td>
<td>0.94</td>
<td>0.84</td>
<td>0.89</td>
<td>0.06</td>
</tr>
<tr>
<td>17</td>
<td>0.90</td>
<td>0.83</td>
<td>0.86</td>
<td>0.10</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td>0.92 0.86 0.89 0.078</td>
</tr>
</tbody>
</table>

AVERAGE 0.92 0.86 0.89 0.078
counted as FP. Same equations 8.1, 8.2, and 8.3 for Recall, Precision, and F-measure are used for General classification evaluation of AMIPF.

After calculating Recall, Precision, and F-measure for each General Classification of each letter, we calculated average Recall, average Precision and average F-measure of each letter’s General Classification using equation 8.4, 8.5, and 8.7 (see Appendix E, Table E.3). In these equations, G.C indicates General Classification, G.Sym denotes General Classification of symptoms, G.Dis means General Classification of diseases, G.Med indicates General Classification of medications, G.Dig denotes General Classification of diagnosis, and G.The denotes General Classification of Therapies.

\[
\text{AVG}_{G.C, \text{Recall, Letter}} = \frac{G.\text{Sym}_{\text{Recall}} + G.\text{Dis}_{\text{Recall}} + G.\text{Med}_{\text{Recall}} + G.\text{Dig}_{\text{Recall}} + G.\text{The}_{\text{Recall}}}{5}
\]

8.4

\[
\text{AVG}_{G.C, \text{Precision, Letter}} = \frac{G.\text{Sym}_{\text{Precision}} + G.\text{Dis}_{\text{Precision}} + G.\text{Med}_{\text{Precision}} + G.\text{Dig}_{\text{Precision}} + G.\text{The}_{\text{Precision}}}{5}
\]

8.5

\[
\text{AVG}_{G.C, \text{F-measure, Letter}} = \frac{G.\text{Sym}_{\text{F-measure}} + G.\text{Dis}_{\text{F-measure}} + G.\text{Med}_{\text{F-measure}} + G.\text{Dig}_{\text{F-measure}} + G.\text{The}_{\text{F-measure}}}{5}
\]

8.6

To calculate the over all General Classification performance of AMIPF, we used the following equations.

\[
\text{AVG}_{G.C} - \text{Recall, AMIPF} = \frac{\sum \text{AVG}_{G.C} - \text{Recall, Letter}}{\text{number of all letters}}
\]

8.7

\[
\text{AVG}_{G.C} - \text{Precision, AMIPF} = \frac{\sum \text{AVG}_{G.C} - \text{Precision, Letter}}{\text{number of all letters}}
\]

8.8

\[
\text{AVG}_{G.C} - \text{F-measure, AMIPF} = \frac{2 \times \text{AVG}_{G.C} - \text{Recall, AMIPF} \times \text{AVG}_{G.C} - \text{Precision, AMIPF}}{\text{AVG}_{G.C} - \text{Recall, AMIPF} + \text{AVG}_{G.C} - \text{Precision, AMIPF}}
\]

8.9
The average Recall for General classification of AMIPF is 0.93 (93%), average Precision is 0.86 (86%) and average F-measure is 0.89. Figure 8.1, Figure 8.2, and Figure 8.3 show the average Recall, average Precision, and average F-measure results for General Classification of the AMIPF. A straight line, in all these figures, shows over all General Classification performance of the AMIPF in terms of average Recall, average Precision, and average F-measure. Figure 8.4 shows comparison of the AMIPF average Recall, average Precision, and average F-measure for the General Classification.

Figure 8.1 AMIPF overall average Recall for General classification indicated by thick straight line and average Recall for General classification of each letter indicated by thin curve
Figure 8.2 AMIPF overall average Precision for General classification indicated by thick straight line and average precision for General classification of each letter indicated by thin curve.

Figure 8.3 AMIPF overall average F-measure for General classification indicated by thick straight line and average F-measure for General classification of each letter indicated by thin curve.
For Cardiology classification evaluation, a reference standard was created from the letters by marking medical concepts that belong to cardiology domain, into five cardiology classification—cardiology symptoms, cardiology diseases, cardiology medications, cardiology diagnosis. Those medical concepts that did not belong to cardiology domain were not considered for cardiology classification reference standard. All letters were processed by AMIPF and true positive, false positive and false negative were calculated for every cardiology classification of each letter. Recall, Precision, and F-measure of each Cardiology Classification were calculated for each letter by using equations 8.1, 8.2, and 8.3. Next, average Recall, average precision, and average F-measure for each letter’s Cardiology Classification were calculated using equations 8.10, 8.11, and 8.12 (see Appendix E, Table E.4). In these equations, C.C indicates Cardiology Classification, C.Sym denotes Cardiology Classification of symptoms, C.Dis means Cardiology Classification of diseases, C.Med indicates Cardiology Classification of medications, C.Dig denotes Cardiology Classification of diagnosis, and C.The denotes Cardiology Classification of Therapies.
To evaluate the overall performance of Cardiology Classification of the AMIPF following equations were used.

\[ \text{AVG}_c.c.R_{\text{Recall, Letter}} = \frac{\text{C.Sym}_{\text{Recall, Letter}} + \text{C.Dis}_{\text{Recall, Letter}} + \text{C.Med}_{\text{Recall, Letter}} + \text{C.Dig}_{\text{Recall, Letter}} + \text{C.The}_{\text{Recall, Letter}}}{5} \]

\[ \text{AVG}_c.c.P_{\text{Precision, Letter}} = \frac{\text{C.Sym}_{\text{Precision, Letter}} + \text{C.Dis}_{\text{Precision, Letter}} + \text{C.Med}_{\text{Precision, Letter}} + \text{C.Dig}_{\text{Precision, Letter}} + \text{C.The}_{\text{Precision, Letter}}}{5} \]

\[ \text{AVG}_c.c.F_{\text{- meas, Letter}} = \frac{\text{C.Sym}_{\text{Precision, Letter}} + \text{C.Dis}_{\text{Precision, Letter}} + \text{C.Med}_{\text{Precision, Letter}} + \text{C.Dig}_{\text{Precision, Letter}} + \text{C.The}_{\text{Precision, Letter}}}{5} \]

The average Recall for cardiology classification of the AMIPF is 0.93 (93 %), average Precision is 0.996 (99.6%) and average F-measure of the AMIPF is 0.96 (96%). Figure 8.5, Figure 8.6 and Figure 8.7 show the average Recall, average Precision, and average, F-measure results for the Cardiology classification of the AMIPF. A straight line, in Figure 8.5, Figure 8.7 and a doted line in Figure 8.6, show over all Cardiology Classification performance of AMIPF in terms of average Recall, average F-measure, and...
average Precision. Figure 8.8 shows comparison of the AMIPF average Recall, average Precision, and average F-measure for the Cardiology classification.

Figure 8.5 AMIPF overall average Recall for Cardiology classification indicated by thick straight line and average Recall for Cardiology classification of each letter indicated by thin curve.

Figure 8.6 AMIPF overall average Precision for Cardiology classification indicated by thick straight line and average precision for Cardiology classification of each letter indicated by thin curve.
Figure 8.7 AMIPF overall average F-measure for Cardiology classification indicated by thick straight line and average F-measure for Cardiology classification of each letter indicated by thin curve.

Figure 8.8 Comparison of AMIPF average Recall, average Precision, and average F-measure for Cardiology classification.

Table 8.10 Average Recall, Precision, F-measure for AMIPF Filtration, General Classification and Cardiology Classification.

<table>
<thead>
<tr>
<th>AMIPF</th>
<th>Average Recall</th>
<th>Average Precision</th>
<th>Average F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtration</td>
<td>92%</td>
<td>86%</td>
<td>89%</td>
</tr>
<tr>
<td>General Classification</td>
<td>93%</td>
<td>86%</td>
<td>89%</td>
</tr>
<tr>
<td>Cardiology Classification</td>
<td>93%</td>
<td>99.6%</td>
<td>96%</td>
</tr>
</tbody>
</table>
In Table 8.10, performance of the AMIPF is summarized for the Filtration, the General Classification, and the Cardiology Classification techniques. In the Filtration technique, we wanted to have higher Recall to include as many significant medical concepts as possible at slight trade off for precision. The ideal Recall value is 1.00 or 100% but, as shown in Table 8.10, AMIPF Filtration Recall is 92%. The most of the Recall error (i.e. 7.8 %) in the Filtration process involves a factor of UMLS natural language processing tool (MMTX). A small Recall error of the Filtration process solely caused due to the AMIPF Filtration technique. For example, in most of the letters medical concept “smoking” was not filtered by AMIPF Filtration technique due to its semantic type “Individual Behavior” that is not in a list of filterable semantic types. If better language processing algorithm is used in AMIPF, the Filtration Recall can be increased.

The AMIPF Filtration average Precision value is 86%. One of the main reasons to affect the average precision of AMIPF Filtration is the association of filterable UMLS semantic types to those terms that were not considered medical concepts in standard reference of medial terms for the AMIPF Filtration. For example, terms “able” and “Problem” have semantic type “Finding” that is a filterable semantic type in the AMIPF for symptoms and syndromes. Filtration average precision was also affected due to some of the ambiguity terms in letters like “Mrs”, “came” etc, these terms were detected as medical terms. For example, a term “Mrs” was detected as a “Magnetic Resonance Spectroscopy” with a semantic type “Diagnostic Procedure”, term while “came” was detected as a medical term with semantic type “Organic Chemical”. The AMIPF average Precision error can be reduced by either having a list of stop-words or using machine learning algorithm to detect ambiguity terms and remove them from Filtration process.
The AMIPF General Classification average Recall and Cardiology Classification average Recall are almost same as the Filtration average Recall. The reason for this is the missing medical terms for the Filtration process, so this error propagated to the lower level. The average Recall of all processes can be further enhanced by using better natural language processing algorithm. The average General Classification Precision is the same as of average precision of the Filtration. Because, all the wrongly detected medical concepts during Filtration process propagated and due to their semantic types they were classified to one of the five general categories. On the other hand Cardiology Classification average Precision is 99.6%. The slight error in average precision is due to UMLS mapping process. For example, for some of the medical terms like “ejection murmur” UMLS mapping produced two terms like ejection murmur and murmur.

### 8.5 Evaluation of the Content and Knowledge Retrieval Framework

The content and knowledge retrieval (CKR) framework main task is to receive health professionals’ information needs in terms of query and retrieve relative computerized CPGs segments (Ex-KCs). The CKR framework is designed to obtain maximum Recall with high Precision by developing a technique that makes use of indexing strategy, phrase-based highly specific index terms, CPGs content meta-information, CPGs contextual model, knowledge standardization, and retrieval algorithm. It is based on a principle defined by (Saltón and McGill, 1983) that in order to obtain high Recall, an exhaustive indexing is required and to ensure high Precision, highly specific index terms should be used. It was also stated that terms should carry additional content indications.
The aim of CKR framework is to obtain relevant Ex-KCs in top ranking and to achieve maximum Recall that is all relevant Ex-KCs should be retrieved with slight trade off with Precision. The objective for the CKR framework evaluation is to find:

- Does CKR framework produce maximum Recall and relatively high Precision for all retrieved Ex-KCs?
- Are most of top ranking retrieved computerized CPGs segments (Ex-KCs) relevant?

To evaluate CKR framework, we performed evaluation at two levels. At the first level, we used a measure of evaluation to determine Recall and Precision for all retrieved (un-ordered) documents (Ex-KCs) for a set of queries. This evaluation addresses the first question. At the second level, we used the measure of evaluation, for ranked documents (Ex-KCs) to evaluate the Precision and Recall for $K$ ranked Ex-KCs for a set of queries. This evaluation addresses the second question.

In information retrieval evaluation “recall measurement requires knowledge of the total number of documents in the collection with respect to each query” (Salton and McGill, 1983). If the size of document collection is relatively small, it is often possible to obtain relevance judgment for all documents with respect to each query (Salton and McGill, 1983). In a case of large collection, it is necessary to estimate the total number of relevant documents in a collection (Salton and McGill, 1983).

For evaluation purpose, we consider every computerized CPGs segment (Ex-KC) as a document. For the evaluation a reference standard (gold standard) was created with sixty five randomly selected Ex-KCs. A set of fifteen queries were created and relevant
documents for each query in a set were identified. To perform evaluation addressing the second question, six top ranking retrieved Ex-KCs for each query were analyzed.

For first level of evaluation, Recall and Precision were calculated from the retrieved Ex-KCs for each query by using the equation 8.10 and 8.11.

\[
\text{Precision} = \frac{\#(\text{relevant Ex-KCs retrieved})}{\#(\text{total retrieved Ex-KCs})} \quad 8.16
\]

\[
\text{Recall} = \frac{\#(\text{relevant Ex-KCs retrieved})}{\#(\text{total relevant Ex-KCs})} \quad 8.17
\]

To find the overall Recall and Precision of the CKR framework, an average Recall and average Precision were calculated using equation 8.12 and 8.13. Figure 8.9 shows the results of Recall, and Precision for each query and Figure 8.10 shows the results of average Recall and average Precision of the CKR framework.

\[
\text{AVG}_{\text{Precision,CKR}} = \frac{\sum \text{Precision of retrieved Ex-KCs for each query}}{\text{Number of queries}} \quad 8.12
\]

\[
\text{AVG}_{\text{Recall,CKR}} = \frac{\sum \text{Recall of retrieved Ex-KCs for each query}}{\text{Number of queries}} \quad 8.13
\]
Figure 8.9 shows that CKR obtained maximum Recall for each query that is 1 (100%) while the Precision for each query was less than Recall but relatively high. Overall average Recall for CKR is 1 (100%) while overall average Precision is 0.89 (89%).
The second level evaluation was performed to address the relevancy of top ranked retrieved Ex-Cs for queries. For this evaluation, we performed information retrieval evaluation for ranked retrieval (Manning., et al., 2009; Salton and McGill, 1983) of top 6 ranked retrieved Ex-KCs. We calculated Precision and Recall of each ranked Ex-KC (from one to 6) for each query. An average Precision for six ranked (R1, R2, R3, R4, R5, and R6) Ex-KCs for all queries was calculated to analyze if the top ranked Ex-KCs retrieved by CKR are relevant to the corresponding queries. Figure 8.11 shows the average Precision for each ranked Ex-KCs. It can be seen that for all queries top three retrieved Ex-KCs were relevant producing 1.00 average Precision while for ranked 4 (R4) very few Ex-KCs were irrelevant. At rank 5 (R5), average Precision was 0.93 and at rank 6 (R6) average Precision was 0.86. This indicates that CKR produces relevant Ex-KCs to the corresponding information needs in the top ranking. Some of the main reasons of achieving high precision are contextual impact factor of medical concepts, indexing strategy to index highly specified medical terms as noun phrases,
standardization of medical concepts, and the algorithm ability to incorporate such important factors during retrieval process.

8.6 Summary

In this chapter, we have presented the evaluation of the techniques and methods comprising the four frameworks: (i) CPGs Computerization Framework, (ii) Context Specific Query Generation Framework, (iii) Automatic Medical Information Processing Framework, and (iv) Content and Knowledge Retrieval Framework. The evaluations of the frameworks were performed along the lines of qualitative and quantitative measures. Evaluation for the CPGs Computerization Framework was performed along the lines of qualitative measure. Evaluation for the CQGF was conducted qualitatively and results were quantified for analysis. The AMIPF and CKR were evaluated along the lines of quantitative measures.
CHAPTER 9
CONCLUSION

9.1 Preamble

This thesis addressed the research problems, in a clinical setting, related to the better interpretation and understanding, of the medical referral and response letters, exchanged between ‘Specialists’ and ‘General Practitioners’ (GP) that can help decision-making at point care. The research problems were implicitly associated with GPs information needs at point of care including access to evidence-based medical knowledge relevant to medical referral and response letters. The research problems presented (see section 1.4.1) in this thesis were addressed by designing and developing methods and techniques for a computer system that manages clinical knowledge, provides evidence-based information and knowledge to assist healthcare practitioners for better interpretation and understanding of clinical referral and response letters, and sends alert messages, for critical situations in letters, at point of care. The goal of the thesis was divided into five objectives.

This is a concluding chapter of the thesis, which summarizes the contributions of this thesis by revisiting them in section 9.2. It presents a review of the objectives and explains how these have been achieved in subsection of 9.2. It outlines the future directions in section 9.3 that emerged from the limitations of the work presented in this thesis.
9.2 Revisiting the Contributions

Earlier in this thesis, we stated the contributions in section 1.5, which are summarized as follows:

- The formulation of an Encoding strategy and a knowledge modeling methodology for the CPGs knowledge representation into Extended-Knowledge Components.

- The development of a CPGs computerization technique for creating Extended-Knowledge Components.

- An automatic query generation technique from computerized CPGs to link Computerized CPGs segments with online evidence-based medical literature.

- A method for computerized processing and analysis of medical referral and response letters to help formulate customized query for accessing evidence-based clinical practices.

- A technique to perform contextual and statistical analysis of medical concepts and an indexing strategy, for contextually relevant retrieval of the CPGs knowledge and corresponding evidence-based literature.

- The design of a knowledge assistance architecture and the implementation of a Clinical Knowledge Assistance computer system for better interpretation and understanding of clinical referral and response letters, sending alert messages for critical situations in letters, and supporting decision making at point of care. The CKA brings together the salient functions of CPGs computerization, linking
online evidence based literature, computerized analysis of medical letters, and contextual retrieval of evidence-based knowledge.

9.2.1 Revisiting the First Objective

The first objective (see section 1.3) of this thesis involves the development of a knowledge modeling and representation method for CPGs computerization into concise yet focused segments that incorporate context, semantics and other meta-information related to CPGs content. Additionally, this knowledge modeling technique should help healthcare professionals in participating with the CPGs computerization process without having technical knowledge. In pursuing this objective, we have developed a knowledge modeling methodology and a CPGs computerization technique (see section 4.3 and 8.2) that achieve this objective.

9.2.2 Revisiting the Second Objective

The second objective of this thesis is to develop a technique to automatically link computerized CPGs segments with relevant online evidence-based medical literature by using contextual, semantics and other information provided by computerized CPGs. In addressing this objective, we have developed a technique to generate automatic search queries (see section 5.3.1) for C-CPGs segments by exploiting the syntactic, contextual, semantic, and meta-information of C-CPGs content. These queries use webservice (see section 5.3.2.9) to retrieve relevant medical literature through online evidence-based knowledge source(s) and to link this literature to the corresponding C-CPGs segments (see section 8.3).
9.2.3 Revisiting the Third Objective

With regards to the third objective, we aim to design a method to perform computerized analysis of the referral and response letters to help healthcare practitioners formulate customized information specifications and to send alerts for critical conditions in letters. In achieving this objective, we have formulated a method that performs computerized processing and analysis of medical letters and provides comprehensive information-view of letters to help formulate customized query (see section 6.4 and section 8.4). This view provides a list of medical concepts classified into five general clinical categories, a list of medical concepts classified into five cardiology categories, a list of negated medical concepts and the corresponding sentences, a list of automatically generated potential query terms, and alerts for critical conditions.

9.2.4 Revisiting the Fourth Objective

In pursuing the fourth objective of this thesis, we have developed a contextual and statistical analysis algorithm and an indexing strategy for accessing contextually relevant evidence-based clinical knowledge at point of care (see section 7.3.1 and 8.5). Our retrieval technique exploits the semantics, context, structural, and meta-information to provide contextually relevant clinical knowledge at point of care.

9.2.5 Revisiting the Fifth Objective

In fulfilling the fifth objective, we have designed layered knowledge assistance system architecture and have implemented a Clinical Knowledge Assistance Computer System that brings together all the methods and techniques described in this thesis into a unified platform (see section 7.4). The layered architecture allows the various functions
and features to be distinguished as individual modules that can be implemented separately with clear and distinctive objectives.

9.3 Future Directions

There are few possible extensions to our work that would extend the functional limits of the proposed techniques and methods.

Our CPGs computerization technique is based on a document centric model of clinical practice guidelines. It uses the GEM as a document centric model for knowledge modeling. The way forward for a possible research direction would be to apply other document centric guideline models using our computerization approach. In our CPGs computerization technique, we have used some of the main elements defined under the Knowledge Component element of the GEM. A possible extension of this approach would be to analyze, identify, and define the scheme of using additional tags (elements e.g. sensitivity, specificity etc) defined in the GEM model for finding the potential tags to enhance context sensitivity and to enrich the CPGs content after computerization.

Our technique for linking computerized CPGs with online evidence-based literature by generating automatic query and determining its query type can be improved by developing a contextual filter and enhanced semantic filter to determine potential medical concepts for a query. Another possible and interesting research direction is to improve the query categorization scheme for finding query types. As our technique has been applied only on the MEDLINE, one of the future extensions would be to apply it on other online evidence-based medical knowledge sources.
The Automatic Medical Information Processing framework (AMIPF) developed in this thesis has been applied on referral letters. One of the future extensions of its application would be to apply it to perform computerized processing of other free text medical documents in different clinical setting for a specific goal. A possible future work would be the improvement of the general classification technique for medical concepts. Another potential extension of the AMIPF would be to enhance the information presented in the information-view. Additional features that can contribute to the information-view are, for example, presenting adverse events, identifying past history and current medical conditions, linking letters with other related past cases etc.

The contextual retrieval technique developed in this thesis can be improved by conducting future research in exploiting additional meta-information related to CPGs content, finding schemes to quantify them, and crafting strategies to use them.

The frameworks developed, in this thesis: CPGs computerization framework, Context Specific Query Generation framework, Automatic Medical Information Processing framework, and Content and Knowledge Retrieval framework, can work independently. One of the interesting future uses of these frameworks would be to apply them separately in other computer applications geared for delivery of medical information, knowledge, etc, in different clinical settings.

The Clinical Knowledge Assistance Computer system developed in this thesis uses a centralized knowledge base and all frameworks implementations reside on one Web server. As healthcare is distributed in nature, it would be advantageous to use distributed knowledge base that can be separated at distributed locations, while multi-agents
deploying the functionalities of each framework working in a distributed virtual setup of a multi-agent based Clinical Knowledge Assistance system.


Cimino, J. and Barnett, G. (1993 c) Automatic knowledge acquisition from MEDLINE, Methods of Information in Medicine, 32(2), 120-130.


Clayton, P., Pryor, T., Wigertz, O. and Hripcsak, G. (1989) Issues and structures for sharing knowledge among decision-making Approaches for creating computer-interpretable guidelines that


262


Guyatt, G.H., Sackett, D.L. and Cook, D.J. (1994) Users’ guides to the medical literature. ii. how to use an article about therapy or prevention. b. what were the results and will they help me in caring for my patients? evidence-based medicine working group, The Journal of the American Medical Association, 271, 59–63.


264


the guideline — implementation gap, Health Care and Informatics Review Online.


APPENDICES
This appendix provides a GEM-encoded CPG from Australian clinical practice guidelines for "Guidelines for the management of acute coronary syndromes 2006". Due to space only one part of GEM-encoded CPG is presented in Table A.1.

Table A.1 GEM-Encoded CPG part for Guidelines for the management of acute coronary syndromes 2006

<table>
<thead>
<tr>
<th>KnowledgeComponents source=&quot;nd&quot;</th>
<th>Acute management of chest pain.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;xml version=&quot;1.0&quot; encoding=&quot;UTF-8&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Developer source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;CommitteeName source=&quot;nd&quot;&gt;&lt;/CommitteeName&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;CommitteeExpertise source=&quot;nd&quot;&gt;&lt;/CommitteeExpertise&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;CommitteeMember source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;MemberExpertise source=&quot;nd&quot;&gt;&lt;/MemberExpertise&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/CommitteeMember&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/CommitteeName&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Funding source=&quot;nd&quot;&gt;&lt;/Funding&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;ComparableGuideline source=&quot;nd&quot;&gt;&lt;/ComparableGuideline&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;RoleOfSponsor source=&quot;nd&quot;&gt;&lt;/RoleOfSponsor&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;ConflictOfInterest source=&quot;nd&quot;&gt;&lt;/ConflictOfInterest&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/Developer&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Purpose source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;MainFocus source=&quot;nd&quot;&gt;&lt;/MainFocus&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Rationale source=&quot;nd&quot;&gt;&lt;/Rationale&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Objective source=&quot;nd&quot;&gt;&lt;/Objective&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;AvailableOption source=&quot;nd&quot;&gt;&lt;/AvailableOption&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;HealthOutcome source=&quot;nd&quot;&gt;&lt;/HealthOutcome&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Exception source=&quot;nd&quot;&gt;&lt;/Exception&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/Purpose&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;IntendedAudience source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Users source=&quot;nd&quot;&gt;&lt;/Users&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;CareSetting source=&quot;nd&quot;&gt;&lt;/CareSetting&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/IntendedAudience&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;TargetPopulation source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Eligibility source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;InclusionCriterion source=&quot;nd&quot;&gt;&lt;/InclusionCriterion&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;ExclusionCriterion source=&quot;nd&quot;&gt;&lt;/ExclusionCriterion&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/Eligibility&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/TargetPopulation&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;KnowledgeComponents source=&quot;nd&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>Acute management of chest pain.</td>
<td></td>
</tr>
</tbody>
</table>
<Recommendation source="nd">acute management of chest pain in going to hospital.
</Recommendation>

<Conditional source="nd">Getting to hospital
</Conditional>

<DecisionVariable source="nd">Chest discomfort
</DecisionVariable>

<Value source="explicit">at rest or for a prolonged period</Value>

<DecisionVariableDescription source="inferred">Chest discomfort at rest or for a prolonged period more than 10 minutes, not relieved by sublingual nitrates. </DecisionVariableDescription>

<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

<DecisionVariable source="nd">recurrent chest discomfort
</DecisionVariable>

<Value source="nd"></Value>

<DecisionVariableDescription source="nd"></DecisionVariableDescription>

<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

<DecisionVariable source="explicit">discomfort associated with syncope
</DecisionVariable>

<Value source="nd"></Value>

<DecisionVariableDescription source="nd"></DecisionVariableDescription>

<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

<DecisionVariable source="inferred">back pain
</DecisionVariable>

<Value source="nd"></Value>

<DecisionVariableDescription source="nd"></DecisionVariableDescription>

<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

<DecisionVariable source="inferred">neck pain
</DecisionVariable>

<Value source="nd"></Value>

<DecisionVariableDescription source="nd"></DecisionVariableDescription>

<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>
<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>
<DecisionVariable source="inferred">arm pain
  <Value source="nd"></Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
</DecisionVariable>
<DecisionVariable source="inferred">epigastric pain
  <Value source="nd"></Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
</DecisionVariable>
<DecisionVariable source="inferred">chest tightness
  <Value source="nd"></Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
</DecisionVariable>
<DecisionVariable source="inferred">dyspnoea
  <Value source="nd"></Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
</DecisionVariable>
<DecisionVariable source="nd">diaphoresis
  <Value source="nd"></Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
</DecisionVariable>
People experiencing such symptoms should seek help promptly and activate emergency medical services to enable transport to the nearest appropriate health care facility for urgent assessment.

Ideally, transport should be by ambulance. However, where ambulance response times are long, alternatives may need to be considered. Patients should be strongly discouraged from driving themselves because of the risk to other road users.
<Reference source="nd"></Reference>

<Imperative source="nd">
<Scope source="nd"></Scope>

<Directive source="nd">

<DirectiveBenefit source="nd"></DirectiveBenefit>

<DirectiveRiskHarm source="nd"></DirectiveRiskHarm>

<DirectiveDescription source="nd"></DirectiveDescription>

<DirectiveCost source="nd"></DirectiveCost>

<DirectiveValue source="nd"></DirectiveValue>

</Directive>

<Reason source="nd"></Reason>

<EvidenceQuality source="nd"></EvidenceQuality>

<RecommendationStrength source="nd"></RecommendationStrength>

<Flexibility source="nd"></Flexibility>

<Logic source="nd"></Logic>

<Cost source="nd"></Cost>

<Linkage source="nd"></Linkage>

<Reference source="nd"></Reference>

<Certainty source="nd"></Certainty>

<Goal source="nd"></Goal>

</Imperative>

</Recommendation>

<Recommendation source="nd">Important requirement to avoid early cardiac arrest</Recommendation>

<Conditional source="nd">

<DecisionVariable source="nd">cardiac arrest

<Value source="nd">occurring</Value>

<DecisionVariableDescription source="explicit">In the case of cardiac arrest occurring in a setting where a defibrillator is not immediately available</DecisionVariableDescription>

<TestParameter source="nd">

<Sensitivity source="nd"></Sensitivity>

<Specificity source="nd"></Specificity>

<PredictiveValue source="nd"></PredictiveValue>

</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

</DecisionVariable>

<DecisionVariable source="nd">defibrillator

<Value source="nd">not immediately available</Value>

<DecisionVariableDescription source="nd"></DecisionVariableDescription>

<TestParameter source="nd">

<Sensitivity source="nd"></Sensitivity>

<Specificity source="nd"></Specificity>

<PredictiveValue source="nd"></PredictiveValue>

</TestParameter>

<DecisionVariableCost source="explicit"></DecisionVariableCost>

</DecisionVariable>

<Action source="explicit">Cardiopulmonary resuscitation should be commenced immediately</Action>

<ActionBenefit source="nd"></ActionBenefit>

<ActionRiskHarm source="nd"></ActionRiskHarm>

<ActionDescription source="nd"></ActionDescription>

<ActionCost source="nd"></ActionCost>

<ActionValue source="nd"></ActionValue>

</Action>

<Reason source="nd"></Reason>

<EvidenceQuality source="nd"></EvidenceQuality>

<RecommendationStrength source="nd"></RecommendationStrength>
The most important initial requirement is access to a defibrillator to avoid early cardiac death from reversible arrhythmias. All Australian ambulances now carry defibrillators.

**Algorithm**

**DecisionVariable source="explicit">Chest discomfort**

**Value source="nd">**

**DecisionVariableDescription source="nd">**

**TestParameter source="nd">**

**Sensitivity source="nd">**

**Specificity source="nd">**

**PredictiveValue source="nd">**

**DecisionVariableCost source="explicit">**
<DecisionVariable source="explicit">acute heart failure</DecisionVariable>
<Value source="nd"><Value>
</Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>
<DecisionVariableCost source="explicit"></DecisionVariableCost>
</DecisionVariable>

<DecisionVariable source="explicit">chest tightness</DecisionVariable>
<Value source="nd"><Value>
</Value>
<DecisionVariableDescription source="nd"></DecisionVariableDescription>
<TestParameter source="nd">
  <Sensitivity source="nd"></Sensitivity>
  <Specificity source="nd"></Specificity>
  <PredictiveValue source="nd"></PredictiveValue>
</TestParameter>
<DecisionVariableCost source="explicit"></DecisionVariableCost>
</DecisionVariable>

<Action source="nd">Aspirin (300 mg) should be given unless already taken or contraindicated.</Action>
<ActionBenefit source="nd"></ActionBenefit>
<ActionRiskHarm source="nd"></ActionRiskHarm>
<ActionDescription source="inferred">should preferably be given early eg, by emergency or ambulance personnel grade D recommendation</ActionDescription>
<ActionCost source="nd"></ActionCost>
<ActionValue source="nd"></ActionValue>
</Action>

<Action source="inferred">Oxygen should also be given.</Action>
<ActionBenefit source="nd"></ActionBenefit>
<ActionRiskHarm source="nd"></ActionRiskHarm>
<ActionDescription source="nd"></ActionDescription>
<ActionCost source="nd"></ActionCost>
<ActionValue source="nd"></ActionValue>
</Action>

<Action source="inferred">Glyceryl trinitrate and intravenous morphine should be given as required.</Action>
<ActionBenefit source="nd"></ActionBenefit>
<ActionRiskHarm source="nd"></ActionRiskHarm>
<ActionDescription source="nd"></ActionDescription>
<ActionCost source="nd"></ActionCost>
<ActionValue source="nd"></ActionValue>
</Action>

<Action source="inferred">Where appropriate, a 12-lead Electrocardiography should be taken en route and transmitted to a medical facility.</Action>
<ActionBenefit source="nd"></ActionBenefit>
<ActionRiskHarm source="nd"></ActionRiskHarm>
<ActionDescription source="nd"></ActionDescription>
<ActionCost source="nd"></ActionCost>
<ActionValue source="nd"></ActionValue>
</Action>

<Action source="nd"></Action>
<ActionBenefit source="nd"></ActionBenefit>
<ActionRiskHarm source="nd"></ActionRiskHarm>
<ActionDescription source="nd"></ActionDescription>
<ActionCost source="nd"></ActionCost>
<ActionValue source="nd"></ActionValue>
</Action>
<Recommendation source="nd">high suspicion of Acute coronary syndrome Symptoms.
<Conditional source="nd">
<DecisionVariable source="nd">Acute coronary syndrome Symptoms
<Value source="nd">high suspicion</Value>
<DecisionVariableDescription source="explicit">whose condition is unstable</DecisionVariableDescription>
<TestParameter>
<Sensitivity source="nd"></Sensitivity>
<Specificity source="nd"></Specificity>
<PredictiveValue source="nd"></PredictiveValue>
</TestParameter>
<DecisionVariableCost source="explicit"></DecisionVariableCost></DecisionVariable>
<DecisionVariableSource="nd">ST segment elevation myocardial infarction
<Value source="nd"></Value>
<DecisionVariableDescription source="explicit">particularly STEMI</DecisionVariableDescription>
<TestParameter>
<Sensitivity source="nd"></Sensitivity>
<Specificity source="nd"></Specificity>
<PredictiveValue source="nd"></PredictiveValue>
</TestParameter>
Receiving medical facilities should be given warning of incoming patients in whom there is a high suspicion of ACS, particularly STEMI, or those whose condition is unstable.

Where formal protocols are in place, prehospital treatment should be given, including fibrinolysis in appropriate cases grade A recommendation.
Acute management of chest pain at hospital on arrival

Suspected Acute coronary syndrome Symptoms treatment.

All patients presenting with suspected ACS

Decision Variable: Suspected Acute coronary syndrome Symptoms

Value: All patients presenting with suspected ACS

Test Parameter:
- Sensitivity
- Specificity
- Predictive Value

Decision Variable Cost: Explicit

Action: All patients presenting with suspected Acute coronary syndrome Symptoms should be subject to ongoing surveillance and have an Electrocardiography completed within 5 minutes of arrival at the medical facility.

Action Benefit: The Electrocardiography should be assessed promptly by an appropriately qualified person.

Action Risk Harm: Oxygen and pain control should be given as required.

Evidence Quality: Details not specified

Recommendation Strength: Grade A recommendation

Directive: Reason not specified

Reference:

281
Initial investigations for Acute coronary syndrome

Initial investigations or Diagnosis suspected Acute coronary syndrome

Suspected Acute coronary syndrome

Patients presenting with a suspected Acute coronary syndrome

Sensitivity

Specificity

Predictive Value

Decision Variable Cost

Decision Variable

Patients should undergo immediate electrocardiography. Further investigations may be necessary, but should not delay treatment.

While other serious diagnoses can present similarly to ACS, eg. pulmonary embolism, aortic dissection, pericarditis once these have been excluded and ACS is considered the most likely diagnosis, further delay in treatment is unnecessary and inappropriate.
Investigations and invasive vascular access techniques should not delay reperfusion therapy.

 Patients whose condition is unstable should have early consultation with a cardiologist.
Initial investigations or Diagnosis with Electrocardiography and Chest x-ray for Acute coronary syndrome

Electrocardiography for ischaemic changes
Electrocardiography is necessary to detect ischaemic changes or arrhythmias. It should be noted that the initial Electrocardiography has a low sensitivity for Acute coronary syndrome, and a normal Electrocardiography does not rule out Acute coronary syndrome. However, the Electrocardiography is the sole test required to select patients for emergency reperfusion, that is fibrinolytic therapy or direct Percutaneous Coronary Intervention.
A chest x-ray is useful for assessing cardiac size evidence of heart failure and other abnormalities grade D recommendation, but should not delay reperfusion treatment where indicated.
Patients without ST-segment elevation on the initial Electrocardiography should be further observed and investigated to promptly identify patients suitable for an emergency reperfusion strategy based on Electrocardiography changes and or determine the best management protocol for non ST-segment-elevation acute coronary syndromes based on risk stratification.
An Extended-Knowledge Component extracted from GEM-encoded CPG (given in Table A.1) is presented in Table A.2. This part of GEM-encoded CPG produced six Ex-KCs but only one is presented in Table A.2 due to the amount of space taken by every Ex-KC file.

Table A.2 Extended-Knowledge Component with KC ID = ACS_06.xml_KC1, created from the GEM-encoded CPG shown in Table A.1

<table>
<thead>
<tr>
<th>Extended-Knowledge Component</th>
<th>ACS_06.xml_KC1 from GEM-encoded CPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS 2006</td>
<td></td>
</tr>
<tr>
<td>&lt;KC ID=&quot;ACS_06.xml_KC1&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;KC.name&gt;Acute management of chest pain.&lt;/KC.name&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;KC.MedTerm CW=&quot;0.5&quot; ID=&quot;ACS_06.xml_KC1&quot; OrgName=&quot;acute&quot; SemType=&quot;Temporal Concept&quot; Vsources=&quot;SNOMEDCT&quot; score=&quot;694&quot;&gt;acute&lt;/KC.MedTerm&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;KC.MedTerm CW=&quot;0.5&quot; ID=&quot;ACS_06.xml_KC1&quot; OrgName=&quot;Management&quot; SemType=&quot;Health Care Activity&quot; Vsources=&quot;SNOMEDCT&quot; score=&quot;861&quot;&gt;Management procedure&lt;/KC.MedTerm&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;Recomendation ID=&quot;ACS_06.xml_KC1_R1&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;R.name&gt;acute management of chest pain in going to hospital.&lt;/R.name&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;RC.MedTerm CW=&quot;0.75&quot; ID=&quot;ACS_06.xml_KC1_R1&quot; OrgName=&quot;acute&quot; SemType=&quot;Temporal Concept&quot; Vsources=&quot;SNOMEDCT&quot; score=&quot;694&quot;&gt;acute&lt;/RC.MedTerm&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;RC.MedTerm CW=&quot;0.75&quot; ID=&quot;ACS_06.xml_KC1_R1&quot; OrgName=&quot;Management&quot; SemType=&quot;Health Care Activity&quot;</td>
<td></td>
</tr>
</tbody>
</table>
Management procedure <RC.MedTerm>
OrgName="Chest Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Chest Pain</RC.MedTerm>

Hospital environment <RC.MedTerm>
OrgName="Hospital" SemType="Manufactured Object" Vsources="SNOMEDCT"
score="1000">Hospitals</RC.MedTerm>

Conditional
ID="ACS_06.xml_KC1_R1_C1" Getting to hospital <C.MedTerm>
OrgName="Hospital" SemType="Qualitative Concept" Vsources="SNOMEDCT" score="1000">Hospital environment</C.MedTerm>

DecisionVariable ID="ACS_06.xml_KC1_R1_C1_dvl"

DecisionVariable ID="ACS_06.xml_KC1_R1_C1_dv2"

DecisionVariable ID="ACS_06.xml_KC1_R1_C1_dv3"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>
<DecisionVariable ID="ACS_06.xml_KCl_Rl_Cl_dv3">
  <DV.name>acute heart failure</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Recurrent" SemType="Phenomenon or Process"
    Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="[D] Chest discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Syncope" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
</DecisionVariable>
<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dv4">
  <DV.name>acute heart failure</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Recurrent" SemType="Phenomenon or Process"
    Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="[D] Chest discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv4"
    OrgName="Syncope" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
</DecisionVariable>
<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dv5">
  <DV.name>back pain</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv5"
    OrgName="Recurrent" SemType="Phenomenon or Process"
    Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv5"
    OrgName="[D] Chest discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv5"
    OrgName="Discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv5"
    OrgName="Syncope" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv5"
    OrgName="Back Pain" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Back Pain</DV.MedTerm>
</DecisionVariable>
<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dv6">
  <DV.name>neck pain</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6"
    OrgName="Recurrent" SemType="Phenomenon or Process"
    Vsources="SNOMEDCT" score="1000">Recurrence</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6"
    OrgName="[D] Chest discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6"
    OrgName="Discomfort" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6"
    OrgName="Syncope" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6"
    OrgName="Neck Pain" SemType="Sign or Symptom"
    Vsources="SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>
</DecisionVariable>
<DecisionVariable ID="ACS_06.xml_KCl_Rl_Cl_dv6">
  <DV.name>arm pain</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6" OrgName="[D]Chest discomfort" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6" OrgName="Back Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Back Pain</DV.MedTerm>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6" OrgName="Neck Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv6" OrgName="Arm Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>
</DecisionVariable>

<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dv7">
  <DV.name>epigastric pain</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv7" OrgName="Back Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Back Pain</DV.MedTerm>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv7" OrgName="Neck Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dv7" OrgName="Arm Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>
</DecisionVariable>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="Discomfort" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="Back Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Back Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw8"
OrgName="Arm Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Recurrence" SemType="Phenomenon or Process"
Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Back Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Back Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Neck Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="[D]Chest discomfort"
SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Arm Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Acute heart failure"
SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>

<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dw9">
  <DV.name>chest tightness</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Discomfort" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Neck Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Arm Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>
</DecisionVariable>

<DecisionVariable ID="ACS_06.xml_KCl_R1_C1_dw9">
  <DV.name>chest tightness</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Discomfort" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Neck Pain" SemType="Sign or Symptom"
Vsources="MSH,SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_R1_C1_dw9"
OrgName="Arm Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>
</DecisionVariable>
Epigastric pain

Chest tightness

Dyspnoea

Recurrence

Chest discomfort

Sensory Discomfort

Syncope

Acute heart failure

Back Pain

Neck Pain

Arm Pain

Epigastric pain

Chest tightness

Dyspnea

Back Pain

Neck Pain

Arm Pain

Recurrence

Chest tightness

Syncope

Acute heart failure

Sensory Discomfort

Epigastric pain

Chest tightness

Dyspnea

Epigastric pain

Chest tightness

Recurrence

Arm Pain

Syncope

Acute heart failure

Neck Pain

Arm Pain

Epigastric pain
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvll"
OrgName="Discomfoit" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Sensory Discomfoit</DV.MedTerm>

<Vsources="MSH,SNOMEDCT" score="1000">Syncope</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Acute heart failure" SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Back Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Back Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Arm Pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Epigastric pain" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Epigastric pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Chest tightness" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Chest tightness</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Dyspnea" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Dyspnea</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Sweating" SemType="Finding"
Vsources="SNOMEDCT" score="1000">Sweating</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="nausea" SemType="Phenomenon or Process"
Vsources="SNOMEDCT" score="901">nausea</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Recurrent" SemType="Phenomenon or Process"
Vsources="SNOMEDCT" score="660">Recurrent</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Chest discomfort" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="901">Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Discomfort" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Syncope" SemType="Sign or Symptom"
Vsources="SNOMEDCT" score="1000">Syncope</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dvl1"
OrgName="Acute heart failure" SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>

294
<DecisionVariable ID="ACS_06.xml_KC1_R1_C1_dv13">
  <DV.name>vomiting</DV.name>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dv13" OrgName="Recurrent" SemType="Phenomenon or Process" Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dv13" OrgName="[D]Chest discomfort" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">Chest discomfort</DV.MedTerm>
  <DV.MedTerm CW="1.0" ID="ACS_06.xml_KC1_R1_C1_dv13" OrgName="Acute heart failure" SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>
</DecisionVariable>
<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="[D]Epigastric pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">[D]Epigastric pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="Chest tightness" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Chest tightness</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="DYSPNOEA" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Dyspnea</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="Diaphoresis" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">SWEATING INCREASED</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="Diaphoresis" SemType="Finding" Vsources="MSH,SNOMEDCT" score="660">Sweating</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="Nausea" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Nausea</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl3"
OrgName="Vomiting" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Vomiting</DV.MedTerm>

</DecisionVariable>

<DV.name>people with diabetes</DV.name>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Recurrent" SemType="Phenomenon or Process" Vsources="MSH,SNOMEDCT" score="660">Recurrence</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="[D]Chest discomfort" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="901">[D]Chest discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Discomfort" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Sensory Discomfort</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Syncope" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Syncope</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Acute heart failure" SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Back Pain" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Back Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Neck Pain" SemType="Sign or Symptom" Vsources="MSH,SNOMEDCT" score="1000">Neck Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Arm Pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">Arm Pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="[D]Epigastric pain" SemType="Sign or Symptom" Vsources="SNOMEDCT" score="1000">[D]Epigastric pain</DV.MedTerm>

<DV.MedTerm CW="1.0" ID="ACS_06.xml_KCl_Rl_Cl_dvl4"
OrgName="Acute heart failure" SemType="Disease or Syndrome" Vsources="SNOMEDCT" score="1000">Acute heart failure</DV.MedTerm>
People experiencing such symptoms should seek help promptly and activate emergency medical services to enable transport to the nearest appropriate health care facility for urgent assessment.
Acute heart failure

Back Pain

Neck Pain

Arm Pain

[D]Epigastric pain

Chest tightness

Dyspnea

Sweating increased

Sweating

Nausea

Vomiting

People

Diabetes Mellitus, Non-Insulin-Dependent

Diabetes Mellitus

Elderly

People

symptoms
Ideally, transport should be by ambulance. However, where ambulance response times are long, alternatives may need to be considered. Patients should be strongly discouraged from driving themselves because

<Recomendation ID="ACS_06.xml_KC1_R2"/>

<R.name>Important requirement to avoid eariy cardiac arrest</R.name>

<DecisionVariable ID="ACS_06.xml_KC1_R2_Cl_dvl">
  <DV.name>cardiac arrest</DV.name>
</DecisionVariable>

<Recomendation ID="ACS_06.xml_KC1_R2">
In the case of cardiac arrest occurring in a setting where a defibrillator is not immediately available:

**Decision Variable**

**Decision Variable ID**: ACS_06.xml_KCl_R2_C1_dv2

**Decision Variable Name**: defibrillator

**Decision Variable Value**: not immediately available

**Organ Name**: defibrillator

**Semantic Type**: Medical Device

**Sources**: MSH, SNOMED CT

Score: 1000

**Decision Variable Description**

*In the case of cardiac arrest occurring in a setting where a defibrillator is not immediately available:*

**Action**

**Action ID**: ACS_06.xml_KCl_R2_C1_A1

**Action Text**: Cardiopulmonary resuscitation should be commenced immediately

**Organ Name**: Cardiopulmonary Resuscitation

**Semantic Type**: Therapeutic or Preventive Procedure

**Sources**: MSH, SNOMED CT

Score: 1000

**Imperative**

**Imperative ID**: ACS_06.xml_KCl_R2_Im1

The most important initial requirement is access to a defibrillator to avoid early cardiac death from reversible arrhythmias. All Australian ambulances now carry defibrillators.

**Imperative Text**: Most

**Organ Name**: Defibrillators

**Semantic Type**: Medical Device

**Sources**: MSH, SNOMED CT

Score: 1000

Score: 861

**Imperative Text**: Initially

**Organ Name**: Access

**Semantic Type**: Spatial Concept

**Sources**: MSH, SNOMED CT

Score: 1000

Score: 645

**Imperative Text**: Early

**Organ Name**: Cardiopulmonary Resuscitation

**Semantic Type**: Temporal Concept

**Sources**: MSH, SNOMED CT

Score: 1000

Score: 660
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Cardiac Death" SemType="Pathologic Function" Vsources="MSH" score="901">Cardiac Death</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Reversible" SemType="Functional Concept" Vsources="SNOMEDCT" score="694">Reversible</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Arrhythmias" SemType="Pathologic Function" Vsources="SNOMEDCT, MSH" score="861">cardiac arrhythmia</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Australia" SemType="Geographic Area" Vsources="MSH, SNOMEDCT" score="589">Australia</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Ambulances" SemType="Manufactured Object" Vsources="MSH, SNOMEDCT" score="827">Ambulances</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Carries" SemType="Finding" Vsources="SNOMEDCT" score="966">Carries</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Carrying" SemType="Activity" Vsources="MSH" score="966">Carrying</IM.MedTerm>
<IM.MedTerm CW="0.50" ID="ACS_06.xml_KCl_R2_Im1" OrgName="Defibrillators" SemType="Medical Device" Vsources="MSH, SNOMEDCT" score="1000">Defibrillators</IM.MedTerm>
</imperitive>
</Recomendation>
</KC>
APPENDIX B

MEDICAL CONCEPTS FOR HEART DISEASE ONTOLOGY

This appendix shows the medical concepts used in heart diseases ontology.

Table B.1 Heart Disease Ontology Medical Concepts

<table>
<thead>
<tr>
<th>Medical Concepts</th>
<th>Medical Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>rheumatic fever</td>
<td>ischemic cardiomyopathy</td>
</tr>
<tr>
<td>acute rheumatic pericarditis</td>
<td>silent myocardial ischaemia</td>
</tr>
<tr>
<td>acute rheumatic endocarditis</td>
<td>subarachnoid hemorrhage</td>
</tr>
<tr>
<td>acute rheumatic myocarditis</td>
<td>cerebral infarction</td>
</tr>
<tr>
<td>other acute rheumatic heart disease</td>
<td>cerebral infarction due to embolism of precerebral arteries</td>
</tr>
<tr>
<td>acute rheumatic heart disease</td>
<td>cerebral infarction due to embolism of cerebral arteries</td>
</tr>
<tr>
<td>rheumatic chorea</td>
<td>cerebrovascular accident</td>
</tr>
<tr>
<td>rheumatic chorea with heart involvement</td>
<td>occlusion and stenosis of cerebral arteries not resulting in cerebral infar</td>
</tr>
<tr>
<td>unstable angina</td>
<td>occlusion and stenosis of middle cerebral artery</td>
</tr>
<tr>
<td>acute myocardial infarction</td>
<td>occlusion and stenosis of anterior cerebral artery</td>
</tr>
<tr>
<td>subsequent myocardial infarction of unspecified site</td>
<td>occlusion and stenosis of posterior cerebral artery</td>
</tr>
<tr>
<td>subsequent myocardial infarction</td>
<td>occlusion and stenosis of cerebellar arteries</td>
</tr>
<tr>
<td>subsequent myocardial infarction of anterior wall</td>
<td>occlusion and stenosis of multiple and bilateral cerebral arteries</td>
</tr>
<tr>
<td>subsequent myocardial infarction of inferior wall</td>
<td>other cerebrovascular disease nos</td>
</tr>
<tr>
<td>subsequent myocardial infarction of other sites</td>
<td>nonruptured cerebral aneurysm</td>
</tr>
<tr>
<td>atrial septal defect as current complication following acute myocardial inf</td>
<td>cerebral aneurysm</td>
</tr>
<tr>
<td>post-infarction ventricular septal defect</td>
<td>hypertensive encephalopathy</td>
</tr>
<tr>
<td>ventricular septal defect as current complication following acute myocardia</td>
<td>moyamoya disease</td>
</tr>
<tr>
<td>rupture of cardiac wall without hemopericardium as current complication fol</td>
<td>other specified cerebrovascular disease</td>
</tr>
<tr>
<td>post-infarction mitral papillary muscle rupture</td>
<td>cerebral amyloid angiopathy</td>
</tr>
<tr>
<td>thrombosis of atrium, auricular appendage and ventricle as current complica</td>
<td>pulmonary embolism</td>
</tr>
<tr>
<td>coronary thrombosis not resulting in myocardial infarction</td>
<td>other pulmonary heart disease</td>
</tr>
<tr>
<td>postmyocardial infarction syndrome</td>
<td>primary pulmonary hypertension</td>
</tr>
<tr>
<td>chronic ischemic heart disease nos</td>
<td>kyphoscoliotic heart disease</td>
</tr>
<tr>
<td>coronary arteriosclerosis</td>
<td>arteriovenous fistula of pulmonary vessels</td>
</tr>
<tr>
<td>old myocardial infarction</td>
<td>acute pericarditis nos</td>
</tr>
<tr>
<td>heart aneurysm</td>
<td>infectious pericarditis</td>
</tr>
<tr>
<td>coronary aneurysm</td>
<td>other pericardial disease</td>
</tr>
<tr>
<td>cardiac tamponade</td>
<td>chronic adhesive pericarditis</td>
</tr>
<tr>
<td>acute and subacute endocarditis unspecified, nos</td>
<td>chronic constrictive pericarditis</td>
</tr>
<tr>
<td>Condition</td>
<td>Subcategory</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>mitral valve insufficiency</td>
<td>pericardial effusion</td>
</tr>
<tr>
<td>mitral valve prolapse</td>
<td>cardiomyopathies</td>
</tr>
<tr>
<td>aortic valve stenosis</td>
<td>cardiomyopathy, dilated</td>
</tr>
<tr>
<td>aortic valve insufficiency</td>
<td>hypertrophic cardiomyopathy</td>
</tr>
<tr>
<td>pulmonary valve disorder</td>
<td>primary eosinophilic endomyocardial restrictive cardiomyopathy</td>
</tr>
<tr>
<td>pulmonary valve stenosis</td>
<td>endocardial fibroelastosis</td>
</tr>
<tr>
<td>pulmonary valve stenosis with insufficiency</td>
<td>cardiomyopathy, alcoholic</td>
</tr>
<tr>
<td>valvular endocarditis</td>
<td>cardiomyopathy due to drugs and other external agents</td>
</tr>
<tr>
<td>acute myocarditis, unspecified</td>
<td>first degree atrioventricular block</td>
</tr>
<tr>
<td>myocarditis due to infectious agent</td>
<td>second degree atrioventricular block</td>
</tr>
<tr>
<td>other acute myocarditis</td>
<td>left bundle-branch block</td>
</tr>
<tr>
<td>acute myocarditis, unspecified</td>
<td>complete atrioventricular block</td>
</tr>
<tr>
<td>left anterior fascicular block</td>
<td>heart failure, right-sided</td>
</tr>
<tr>
<td>left posterior fascicular block</td>
<td>left-sided heart failure</td>
</tr>
<tr>
<td>other conduction disorders nos</td>
<td>left-sided heart failure</td>
</tr>
<tr>
<td>right bundle branch block</td>
<td>myocarditis</td>
</tr>
<tr>
<td>bifascicular block</td>
<td>cardiomyopathies</td>
</tr>
<tr>
<td>trifascicular block</td>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>anomalous atrioventricular excitation nos</td>
<td>other ill-defined heart diseases</td>
</tr>
<tr>
<td>sequelae of cerebral infarction</td>
<td>heart diseases</td>
</tr>
<tr>
<td>conduction disorder of the heart</td>
<td>rheumatoid carditis</td>
</tr>
<tr>
<td>heart block</td>
<td>essential hypertension</td>
</tr>
<tr>
<td>adams-stokes syndrome</td>
<td>hypertensive disease</td>
</tr>
<tr>
<td>cardiac arrest with successful resuscitation</td>
<td>hypertensive heart disease nos</td>
</tr>
<tr>
<td>tachycardia, paroxysmal</td>
<td>hypertensive heart failure</td>
</tr>
<tr>
<td>supraventricular tachycardia</td>
<td>hypertensive nephropathy</td>
</tr>
<tr>
<td>tachycardia, paroxysmal</td>
<td>hypertensive heart and renal disease</td>
</tr>
<tr>
<td>atrial fibrillation and flutter</td>
<td>hypertensive heart and renal disease with (congestive) heart failure</td>
</tr>
<tr>
<td>atrial premature complexes</td>
<td>hypertensive heart and renal failure</td>
</tr>
<tr>
<td>sick sinus syndrome</td>
<td>hypertensive heart and renal disease with both (congestive) heart failure a</td>
</tr>
<tr>
<td>bradycardia-tachycardia syndrome</td>
<td>secondary hypertension nos</td>
</tr>
<tr>
<td>heart failure</td>
<td>hypertension, renovascular</td>
</tr>
<tr>
<td>congestive heart failure</td>
<td>hypertension secondary to endocrine disorder</td>
</tr>
<tr>
<td>postpericardiotomy syndrome</td>
<td>pericarditis, constrictive</td>
</tr>
<tr>
<td>postmastectomy lymphedema syndrome</td>
<td>hypercholesterolemia</td>
</tr>
<tr>
<td>syphilis, cardiovascular</td>
<td>aortic stenosis</td>
</tr>
<tr>
<td>coronary heart disease</td>
<td>mitral valve stenosis</td>
</tr>
<tr>
<td>myocardial infarction</td>
<td>tricuspid valve insufficiency</td>
</tr>
<tr>
<td>hypertensive disease</td>
<td>Hypertension, Pulmonary</td>
</tr>
<tr>
<td>myocardial ischemia</td>
<td>acute coronary syndrome</td>
</tr>
<tr>
<td>cardiac arrhythmia</td>
<td>left bundle branch block</td>
</tr>
<tr>
<td>tachycardia</td>
<td>diabetes mellitus, non-insulin-dependent</td>
</tr>
<tr>
<td>rheumatic heart disease</td>
<td>acute heart failure</td>
</tr>
<tr>
<td>cor pulmonale</td>
<td>tachycardia, atrioventricular nodal reentry</td>
</tr>
<tr>
<td>hypertensive heart disease nos</td>
<td>heart failure</td>
</tr>
<tr>
<td>Condition</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>heart valve disease</td>
<td>myocardial ischemia</td>
</tr>
<tr>
<td>infective endocarditis</td>
<td>anemia</td>
</tr>
<tr>
<td>congenital heart disease</td>
<td>paroxysmal atrial fibrillation</td>
</tr>
<tr>
<td>coronary heart disease</td>
<td>ischemia</td>
</tr>
<tr>
<td>atrial myxoma</td>
<td>Subclavian Steal Syndrome</td>
</tr>
<tr>
<td>hypertrophic cardiomyopathy</td>
<td>Inferior Wall Myocardial Infarction</td>
</tr>
<tr>
<td>long qt syndrome</td>
<td>Inferior Wall Myocardial Infarction</td>
</tr>
<tr>
<td>supraventricular tachycardia</td>
<td>Acute Anterior Wall Myocardial Infarction</td>
</tr>
<tr>
<td>atrial flutter</td>
<td>Anterior myocardial infarction</td>
</tr>
<tr>
<td>angina pectoris</td>
<td>chest pain on breathing</td>
</tr>
<tr>
<td>chest discomfort</td>
<td>precordial pain</td>
</tr>
<tr>
<td>palpitations</td>
<td>chest pain</td>
</tr>
<tr>
<td>cardiac arrhythmia</td>
<td>asphyxia</td>
</tr>
<tr>
<td>lightheadedness</td>
<td>pleurisy</td>
</tr>
<tr>
<td>tachycardia</td>
<td>respiratory arrest</td>
</tr>
<tr>
<td>bradycardia</td>
<td>cardiorespiratory failure</td>
</tr>
<tr>
<td>pericarditis</td>
<td>abnormal sputum nos</td>
</tr>
<tr>
<td>heart block</td>
<td>[d]abnormal sputum nos</td>
</tr>
<tr>
<td>cardiac arrest</td>
<td>weak arterial pulse</td>
</tr>
<tr>
<td>heart failure</td>
<td>rales</td>
</tr>
<tr>
<td>high output heart failure</td>
<td>bruit</td>
</tr>
<tr>
<td>congenital heart disease</td>
<td>[d]chest discomfort</td>
</tr>
<tr>
<td>congestive heart failure</td>
<td>sinus bradycardia</td>
</tr>
<tr>
<td>endocarditis</td>
<td>st segment elevation</td>
</tr>
<tr>
<td>cor pulmonale</td>
<td>st segment depression</td>
</tr>
<tr>
<td>cardiomyopathies</td>
<td>shock, cardiogenic</td>
</tr>
<tr>
<td>weak arterial pulse</td>
<td>chest tightness</td>
</tr>
<tr>
<td>epigastric pain</td>
<td>heart arrest</td>
</tr>
<tr>
<td>bruit</td>
<td>arm pain</td>
</tr>
<tr>
<td>atrial fibrillation</td>
<td>sweating increased</td>
</tr>
<tr>
<td>pain radiating to left arm</td>
<td>cardiac death</td>
</tr>
<tr>
<td>back pain</td>
<td>ecg normal</td>
</tr>
<tr>
<td>sweating</td>
<td>obstruction</td>
</tr>
<tr>
<td>cold sweat</td>
<td>cardiac arrhythmia</td>
</tr>
<tr>
<td>nausea</td>
<td>hypertrophy, left ventricular</td>
</tr>
<tr>
<td>vomiting</td>
<td>diastolic dysfunction</td>
</tr>
<tr>
<td>lightheadedness</td>
<td>non-smoker</td>
</tr>
<tr>
<td>syncope</td>
<td>Sinus rhythm</td>
</tr>
<tr>
<td>shortness of breath</td>
<td>jugular venous pressure</td>
</tr>
<tr>
<td>fatigue</td>
<td>heart murmur</td>
</tr>
<tr>
<td>anxiety</td>
<td>Premature ventricular contractions</td>
</tr>
<tr>
<td>dyspnea</td>
<td>ventricular premature complexes</td>
</tr>
<tr>
<td>increased sweating</td>
<td>wheezing</td>
</tr>
<tr>
<td>panic</td>
<td>infarction</td>
</tr>
<tr>
<td>tachycardia</td>
<td>P mitrale (finding)</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>bradycardia</td>
<td>P-mitrale</td>
</tr>
<tr>
<td>palpitations</td>
<td>Sinus bradycardia</td>
</tr>
<tr>
<td>innocent murmurs</td>
<td>Radiofemoral delay</td>
</tr>
<tr>
<td>precordial friction</td>
<td>Carotid bruit</td>
</tr>
<tr>
<td>gangrene</td>
<td>hypokinesia</td>
</tr>
<tr>
<td>atherosclerosis</td>
<td>ST segment</td>
</tr>
<tr>
<td>diabetes mellitus</td>
<td>Stress</td>
</tr>
<tr>
<td>peripheral vascular diseases</td>
<td>Ejection Murmurs</td>
</tr>
<tr>
<td>gas gangrene</td>
<td>Ventricular Dysfunction, Left</td>
</tr>
<tr>
<td>elevated blood-pressure reading without diagnosis of hypertension</td>
<td>Stress</td>
</tr>
<tr>
<td>hypotension</td>
<td>Ejection murmur</td>
</tr>
<tr>
<td>maternal hypotension syndrome unspecified</td>
<td>Soft systolic murmur</td>
</tr>
<tr>
<td>dyspnea</td>
<td>Tachycardia, Ventricular</td>
</tr>
<tr>
<td>orthopnea</td>
<td>Normal blood pressure</td>
</tr>
<tr>
<td>shortness of breath</td>
<td>Dyspnea on exertion</td>
</tr>
<tr>
<td>stridor</td>
<td>Premature Cardiac Complex</td>
</tr>
<tr>
<td>respiration intermittent</td>
<td>Ventricular bigeminy</td>
</tr>
<tr>
<td>stethoscopes</td>
<td>systemic arterial pressure</td>
</tr>
<tr>
<td>electrocardiography</td>
<td>renal arteriography</td>
</tr>
<tr>
<td>exercise stress ecg test</td>
<td>doppler</td>
</tr>
<tr>
<td>stress test</td>
<td>atrial pressure</td>
</tr>
<tr>
<td>radionuclide imaging</td>
<td>troponin measurement</td>
</tr>
<tr>
<td>coronary angiography</td>
<td>blood tests</td>
</tr>
<tr>
<td>electrocardiogram</td>
<td>blood analysis (lab test)</td>
</tr>
<tr>
<td>echocardiography</td>
<td>Pulse taking</td>
</tr>
<tr>
<td>auscultation</td>
<td>cardiac markers</td>
</tr>
<tr>
<td>blood analysis (lab test)</td>
<td>Left ventricular size</td>
</tr>
<tr>
<td>diagnosis</td>
<td>Exercise stress test</td>
</tr>
<tr>
<td>radiography, thoracic</td>
<td>Coronary angiography</td>
</tr>
<tr>
<td>angiogram</td>
<td>Echocardiography, Stress</td>
</tr>
<tr>
<td>Holter Electrocardiography</td>
<td>coronary artery bypass surgery</td>
</tr>
<tr>
<td>coronary angioplasty</td>
<td>convulsive therapy</td>
</tr>
<tr>
<td>reperfusion therapy</td>
<td>antibiotic prophylaxis</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>ablation</td>
</tr>
<tr>
<td>fibrinolysis</td>
<td>radiofrequency ablation</td>
</tr>
<tr>
<td>cardiopulmonary resuscitation</td>
<td>oxygen therapy care</td>
</tr>
<tr>
<td>defibrillation</td>
<td>annuloplasty</td>
</tr>
<tr>
<td>defibrillators</td>
<td>Modified Bruce protocol</td>
</tr>
<tr>
<td>thrombolytic therapy</td>
<td>repair of mitral valve</td>
</tr>
<tr>
<td>percutaneous coronary intervention</td>
<td>anti-thrombins</td>
</tr>
<tr>
<td>mogadon</td>
<td>aspirin</td>
</tr>
<tr>
<td>nexium</td>
<td>angiotensin-converting enzyme inhibitors</td>
</tr>
<tr>
<td>tramal</td>
<td>adrenergic beta-antagonists</td>
</tr>
<tr>
<td>cephalixin</td>
<td>nitrates</td>
</tr>
<tr>
<td>norvasc</td>
<td>calcium channel blockers</td>
</tr>
<tr>
<td>cardizem</td>
<td>diuretics</td>
</tr>
<tr>
<td>metoprolol</td>
<td>coronary artery bypass surgery</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>amiodarone</td>
<td>coronary angioplasty</td>
</tr>
<tr>
<td>warfarin</td>
<td>isoproterenol</td>
</tr>
<tr>
<td>zorubicin</td>
<td>hydroxymethylglutaryl-coa reductase inhibitors</td>
</tr>
<tr>
<td>lipitor</td>
<td>atorvastatin</td>
</tr>
<tr>
<td>pfizer brand of atorvastatin calcium</td>
<td>lovastatin</td>
</tr>
<tr>
<td>atenolol</td>
<td>pravastatin</td>
</tr>
<tr>
<td>adrenergic beta-antagonists</td>
<td>omega-3 acid ethyl esters (usp)</td>
</tr>
<tr>
<td>deptran</td>
<td>niacin</td>
</tr>
<tr>
<td>arimidex</td>
<td>gemfibrozil</td>
</tr>
<tr>
<td>tambocor</td>
<td>fenofibrate</td>
</tr>
<tr>
<td>verapamil</td>
<td>probucol</td>
</tr>
<tr>
<td>electrolytes</td>
<td>ezetimibe</td>
</tr>
<tr>
<td>lanoxin</td>
<td>ticlopidine</td>
</tr>
<tr>
<td>Cartia</td>
<td>clopidogrel</td>
</tr>
<tr>
<td>digoxin</td>
<td>reperfusion therapy</td>
</tr>
<tr>
<td>oxygen</td>
<td>fibrinolysis</td>
</tr>
<tr>
<td>Water</td>
<td>fibrinolysis</td>
</tr>
<tr>
<td>Coversyl</td>
<td>adjuvant</td>
</tr>
<tr>
<td>Renitec</td>
<td>3-carbamoyl-2,2,5,5-tetramethyl-3-pyrroline-1-yloxyl</td>
</tr>
<tr>
<td>Toprol-XL</td>
<td>imdur</td>
</tr>
<tr>
<td>Rocaltrol</td>
<td>betaloc-astra</td>
</tr>
<tr>
<td>Pariet</td>
<td>zoloft</td>
</tr>
<tr>
<td>Akineton</td>
<td>creatinine</td>
</tr>
<tr>
<td>Slow-K</td>
<td>minipress</td>
</tr>
<tr>
<td>Furosemide</td>
<td>aldomet</td>
</tr>
<tr>
<td>Temazepam</td>
<td>avapro</td>
</tr>
<tr>
<td>fish oils</td>
<td>lasix</td>
</tr>
<tr>
<td>Coversyl</td>
<td>mobic</td>
</tr>
<tr>
<td>Folic Acid</td>
<td>ogen</td>
</tr>
</tbody>
</table>
This appendix provides semantic types used in Semantic Type Filter and for query type categories. It also presents the “associated-with relation” used for finding Semantic Relations score among medical concepts.

<table>
<thead>
<tr>
<th>Semantic Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired Abnormality</td>
</tr>
<tr>
<td>Anatomical Abnormality</td>
</tr>
<tr>
<td>Chemical Viewed Functionally</td>
</tr>
<tr>
<td>Congenital Abnormality</td>
</tr>
<tr>
<td>Diagnostic Procedure</td>
</tr>
<tr>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>Finding</td>
</tr>
<tr>
<td>Hazardous or Poisonous Substance</td>
</tr>
<tr>
<td>Health Care Activity</td>
</tr>
<tr>
<td>Immunologic Factor</td>
</tr>
<tr>
<td>Indicator, Reagent, or Diagnostic Aid</td>
</tr>
<tr>
<td>Inorganic Chemical</td>
</tr>
<tr>
<td>Laboratory or Test Result</td>
</tr>
<tr>
<td>Laboratory Procedure</td>
</tr>
<tr>
<td>Medical Device</td>
</tr>
<tr>
<td>Organic Chemical</td>
</tr>
<tr>
<td>Organism Function</td>
</tr>
<tr>
<td>Pathologic Function</td>
</tr>
<tr>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>Physiologic Function</td>
</tr>
<tr>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>Steroid</td>
</tr>
<tr>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>Diagnosis</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Diagnostic Procedure</td>
</tr>
<tr>
<td>Pathologic Function</td>
</tr>
<tr>
<td>Health Care Activity</td>
</tr>
<tr>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>Finding</td>
</tr>
<tr>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>Medical Device</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Semantic Relations

`associated_with`

Definition

has a significant or salient relationship to.

Semantic Types

- Anatomical Abnormality associated_with Finding
- Finding associated_with Pathologic Function
- Pathologic Function associated_with Finding
- Pathologic Function associated_with Pathologic Function
- Acquired Abnormality associated_with Diagnostic Procedure
- Acquired Abnormality associated_with Health Care Activity
- Acquired Abnormality associated_with Laboratory Procedure
- Acquired Abnormality associated_with Laboratory or Test Result
- Acquired Abnormality associated_with Sign or Symptom
- Acquired Abnormality associated_with Therapeutic or Preventive Procedure
- Anatomical Abnormality associated_with Diagnostic Procedure
- Anatomical Abnormality associated_with Health Care Activity
- Anatomical Abnormality associated_with Laboratory Procedure
- Anatomical Abnormality associated_with Laboratory or Test Result
- Anatomical Abnormality associated_with Sign or Symptom
- Anatomical Abnormality associated_with Therapeutic or Preventive Procedure
- Congenital Abnormality associated_with Diagnostic Procedure
- Congenital Abnormality associated_with Finding
- Congenital Abnormality associated_with Health Care Activity
- Congenital Abnormality associated_with Laboratory Procedure
- Congenital Abnormality associated_with Laboratory or Test Result
- Congenital Abnormality associated_with Pathologic Function
- Congenital Abnormality associated_with Sign or Symptom
- Congenital Abnormality associated_with Therapeutic or Preventive Procedure
- Diagnostic Procedure associated_with Disease or Syndrome
- Diagnostic Procedure associated_with Pathologic Function
Disease or Syndrome associated_with Diagnostic Procedure
Disease or Syndrome associated_with Disease or Syndrome
Disease or Syndrome associated_with Finding
Disease or Syndrome associated_with Health Care Activity
Disease or Syndrome associated_with Laboratory Procedure
Disease or Syndrome associated_with Pathologic Function
Disease or Syndrome associated_with Sign or Symptom
Disease or Syndrome associated_with Therapeutic or Preventive Procedure
Finding associated_with Acquired Abnormality
Finding associated_with Congenital Abnormality
Finding associated_with Disease or Syndrome
Health Care Activity associated_with Acquired Abnormality
Health Care Activity associated_with Anatomical Abnormality
Health Care Activity associated_with Congenital Abnormality
Health Care Activity associated_with Disease or Syndrome
Health Care Activity associated_with Pathologic Function
Laboratory Procedure associated_with Acquired Abnormality
Laboratory Procedure associated_with Anatomical Abnormality
Laboratory Procedure associated_with Congenital Abnormality
Laboratory Procedure associated_with Cell or Molecular Dysfunction
Laboratory Procedure associated_with Disease or Syndrome
Laboratory Procedure associated_with Experimental Model of Disease
Laboratory Procedure associated_with Injury or Poisoning
Laboratory Procedure associated_with Mental or Behavioral Dysfunction
Laboratory Procedure associated_with Neoplastic Process
Laboratory Procedure associated_with Pathologic Function
Laboratory or Test Result associated_with Acquired Abnormality
Laboratory or Test Result associated_with Anatomical Abnormality
Laboratory or Test Result associated_with Congenital Abnormality
Laboratory or Test Result associated_with Disease or Syndrome
Laboratory or Test Result associated_with Pathologic Function
Pathologic Function associated_with Diagnostic Procedure
Pathologic Function associated_with Disease or Syndrome
Pathologic Function associated_with Experimental Model of Disease
Pathologic Function associated_with Health Care Activity

311
Pathologic Function associated with Laboratory Procedure
Pathologic Function associated with Laboratory or Test Result
Pathologic Function associated with Sign or Symptom
Pathologic Function associated with Therapeutic or Preventive Procedure
Sign or Symptom associated with Acquired Abnormality
Sign or Symptom associated with Anatomical Abnormality
Sign or Symptom associated with Congenital Abnormality
Sign or Symptom associated with Disease or Syndrome
Sign or Symptom associated with Pathologic Function
Therapeutic or Preventive Procedure associated with Acquired Abnormality
Therapeutic or Preventive Procedure associated with Anatomical Abnormality
Therapeutic or Preventive Procedure associated with Congenital Abnormality
Therapeutic or Preventive Procedure associated with Disease or Syndrome
Therapeutic or Preventive Procedure associated with Pathologic Function
This appendix provides all extracted terms of medical response letter used in a working example of AMIPF processing in section 6.4. It shows the semantic types used for general classification of medical concepts. It presents the evaluation results discussed in section 8.4, for each letter’s general classification and cardiology classification.

Table E.1 Medical concepts extracted as potential candidates during response letter processing discussed in section 6.4

<table>
<thead>
<tr>
<th>Standardized Medical concepts</th>
<th>Original Medical Concepts</th>
<th>Semantic Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>ask protein, human</td>
<td>ASK</td>
<td>Biologically Active Substance</td>
</tr>
<tr>
<td>seen</td>
<td>Seen</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>vision</td>
<td>seeing</td>
<td>Organism Function</td>
</tr>
<tr>
<td>magnetic resonance spectroscopy</td>
<td>MRS</td>
<td>Diagnostic Procedure</td>
</tr>
<tr>
<td>mineralocorticoid receptor</td>
<td>MR</td>
<td>Receptor</td>
</tr>
<tr>
<td>recent</td>
<td>Recently</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>entire right foot</td>
<td>Right foot</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>structure of right foot</td>
<td>Right foot</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>fracture of phalanx of foot</td>
<td>Fractured toe</td>
<td>Injury or Poisoning</td>
</tr>
<tr>
<td>immobile</td>
<td>Immobile</td>
<td>Finding</td>
</tr>
<tr>
<td>time</td>
<td>Time</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>with time</td>
<td>Time</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>swollen ankle</td>
<td>ANKLE OEDEMA</td>
<td>Finding</td>
</tr>
<tr>
<td>edema</td>
<td>Edema</td>
<td>Pathologic Function</td>
</tr>
<tr>
<td>day</td>
<td>day</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>deterioration of status</td>
<td>Worse</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>worse</td>
<td>Worse</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>day</td>
<td>day</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>dyspnea</td>
<td>Dyspnea</td>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>acquired</td>
<td>Acquired</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>increased</td>
<td>Increased</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>increase</td>
<td>Increase</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>increasing</td>
<td>Increasing</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>obtained</td>
<td>Obtained</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>significant</td>
<td>Significant</td>
<td>Idea or Concept</td>
</tr>
<tr>
<td>quantity</td>
<td>Amount</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>body weight</td>
<td>WEIGHT</td>
<td>Organism Attribute</td>
</tr>
<tr>
<td>weighing patient</td>
<td>Weight</td>
<td>Diagnostic Procedure</td>
</tr>
<tr>
<td>weight</td>
<td>Weight</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>treats</td>
<td>Treats</td>
<td>Classification</td>
</tr>
<tr>
<td>arimidex</td>
<td>Arimidex</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>breast carcinoma</td>
<td>Breast Cancer</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>malignant neoplasm of breast</td>
<td>Breast Cancer</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>magnetic resonance spectroscopy</td>
<td>MRS</td>
<td>Diagnostic Procedure</td>
</tr>
<tr>
<td>mineralocorticoid receptor</td>
<td>MR</td>
<td>Receptor</td>
</tr>
<tr>
<td>historical aspects</td>
<td>history</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>history of present illness</td>
<td>History, NOS</td>
<td>Organism Attribute</td>
</tr>
<tr>
<td>medical history</td>
<td>History</td>
<td>Finding</td>
</tr>
<tr>
<td>recording of previous events</td>
<td>History</td>
<td>Occupation or Discipline</td>
</tr>
<tr>
<td>myocardial ischemia</td>
<td>Ischaemic heart disease</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>rheumatic fever</td>
<td>Rheumatic Fever</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>aging</td>
<td>Aging</td>
<td>Organism Function</td>
</tr>
<tr>
<td>elderly</td>
<td>Aged</td>
<td>Age Group</td>
</tr>
<tr>
<td>eight</td>
<td>Eight</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>breast carcinoma</td>
<td>Breast Carcinoma</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>year</td>
<td>year</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>limited</td>
<td>Limited</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>excision</td>
<td>Resection</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>radio communications</td>
<td>Radio</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>therapeutic aspects</td>
<td>therapy</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>radio communications</td>
<td>Radio</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>therapeutic procedure</td>
<td>Therapy</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>radio</td>
<td>Radio</td>
<td>Manufactured Object</td>
</tr>
<tr>
<td>therapeutic aspects</td>
<td>therapy</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>radio</td>
<td>Radio</td>
<td>Manufactured Object</td>
</tr>
<tr>
<td>therapeutic procedure</td>
<td>Therapy</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>long-term</td>
<td>Long-term</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>endocrine therapy</td>
<td>therapy, hormonal</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>risk factors</td>
<td>risk factors</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>heart diseases</td>
<td>Heart disease</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>including</td>
<td>Includes</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>hypertensive disease</td>
<td>Hypertension</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>hypercholesterolemia</td>
<td>hypercholesterolemia</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>pharmaceutical preparations</td>
<td>MEDICATIONS</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>current</td>
<td>Currently</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>including</td>
<td>Includes</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>atacand</td>
<td>Atacand</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>animal mange</td>
<td>Mane</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>atacand</td>
<td>Atacand</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>morning</td>
<td>mane</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>night time</td>
<td>nocte</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>pfizer brand of atorvastatin calcium</td>
<td>Lipitor</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>animal mange</td>
<td>Mane</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>pfizer brand of atorvastatin calcium</td>
<td>Lipitor</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>morning</td>
<td>mane</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>animal mane</td>
<td>Mane</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>Term</td>
<td>Category</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>morning</td>
<td>Temporal Concept</td>
<td>mane</td>
</tr>
<tr>
<td>lasix</td>
<td>Pharmacologic Substance</td>
<td>Lasix</td>
</tr>
<tr>
<td>recent</td>
<td>Temporal Concept</td>
<td>Recently</td>
</tr>
<tr>
<td>increased</td>
<td>Quantitative Concept</td>
<td>Increased</td>
</tr>
<tr>
<td>arimidex</td>
<td>Pharmacologic Substance</td>
<td>Arimidex</td>
</tr>
<tr>
<td>nexium</td>
<td>Pharmacologic Substance</td>
<td>Nexium</td>
</tr>
<tr>
<td>animal mane</td>
<td>Body Part, Organ, or Organ Component</td>
<td>Mane</td>
</tr>
<tr>
<td>nexium</td>
<td>Pharmacologic Substance</td>
<td>Nexium</td>
</tr>
<tr>
<td>morning</td>
<td>Temporal Concept</td>
<td>mane</td>
</tr>
<tr>
<td>alodorm</td>
<td>Pharmacologic Substance</td>
<td>Alodorm</td>
</tr>
<tr>
<td>physical examination</td>
<td>Health Care Activity</td>
<td>Examination</td>
</tr>
<tr>
<td>weighing patient</td>
<td>Diagnostic Procedure</td>
<td>Weighing</td>
</tr>
<tr>
<td>physiologic pulse</td>
<td>Physiologic Function</td>
<td>Pulse</td>
</tr>
<tr>
<td>pulse taking</td>
<td>Diagnostic Procedure</td>
<td>Pulse</td>
</tr>
<tr>
<td>regular</td>
<td>Qualitative Concept</td>
<td>Regular</td>
</tr>
<tr>
<td>minute of time</td>
<td>Temporal Concept</td>
<td>minute</td>
</tr>
<tr>
<td>minute</td>
<td>Quantitative Concept</td>
<td>Minute</td>
</tr>
<tr>
<td>blood pressure</td>
<td>Organism Function</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>systemic arterial pressure</td>
<td>Finding</td>
<td>Blood pressure</td>
</tr>
<tr>
<td>high</td>
<td>Qualitative Concept</td>
<td>Elevated</td>
</tr>
<tr>
<td>very</td>
<td>Qualitative Concept</td>
<td>Very</td>
</tr>
<tr>
<td>minimal</td>
<td>Qualitative Concept</td>
<td>Mild</td>
</tr>
<tr>
<td>swollen ankle</td>
<td>Finding</td>
<td>ANKLE OEDEMA</td>
</tr>
<tr>
<td>limited</td>
<td>Functional Concept</td>
<td>Limited</td>
</tr>
<tr>
<td>ankle</td>
<td>Body Location or Region</td>
<td>Ankles</td>
</tr>
<tr>
<td>feet</td>
<td>Quantitative Concept</td>
<td>feet</td>
</tr>
<tr>
<td>pes</td>
<td>Body Part, Organ, or Organ Component</td>
<td>Feet</td>
</tr>
<tr>
<td>auscultation</td>
<td>Diagnostic Procedure</td>
<td>Auscultation</td>
</tr>
<tr>
<td>grade ii</td>
<td>Intellectual Product</td>
<td>Grade</td>
</tr>
<tr>
<td>heart murmur</td>
<td>Sign or Symptom</td>
<td>Murmur</td>
</tr>
<tr>
<td>grade</td>
<td>Classification</td>
<td>Grade</td>
</tr>
<tr>
<td>ejection murmur</td>
<td>Finding</td>
<td>Ejection murmur</td>
</tr>
<tr>
<td>histopathologic grade</td>
<td>Classification</td>
<td>Grade</td>
</tr>
<tr>
<td>ejection murmur</td>
<td>Finding</td>
<td>Ejection murmur</td>
</tr>
<tr>
<td>base</td>
<td>Chemical Viewed Functionally</td>
<td>Base</td>
</tr>
<tr>
<td>chest</td>
<td>Body Location or Region</td>
<td>Chest</td>
</tr>
<tr>
<td>translucent</td>
<td>Qualitative Concept</td>
<td>Clear</td>
</tr>
<tr>
<td>electrocardiogram</td>
<td>Diagnostic Procedure</td>
<td>ECG</td>
</tr>
<tr>
<td>[d]sinus bradycardia</td>
<td>Pathologic Function</td>
<td>[D]Sinus bradycardia</td>
</tr>
<tr>
<td>unspecified</td>
<td>Qualitative Concept</td>
<td>Non-specific</td>
</tr>
<tr>
<td>casp4 protein, human</td>
<td>Enzyme</td>
<td>TX</td>
</tr>
<tr>
<td>unspecified</td>
<td>Therapeutic or Preventive Procedure</td>
<td>Non-specific</td>
</tr>
<tr>
<td>echocardiography</td>
<td>Diagnostic Procedure</td>
<td>Echocardiography</td>
</tr>
<tr>
<td>normal</td>
<td>Qualitative Concept</td>
<td>Normal</td>
</tr>
<tr>
<td>entire left ventricle</td>
<td>Body Part, Organ, or Organ Component</td>
<td>Left ventricle</td>
</tr>
<tr>
<td>normal</td>
<td>Qualitative Concept</td>
<td>Normal</td>
</tr>
<tr>
<td>left ventricular structure</td>
<td>Body Part, Organ, or Organ Component</td>
<td>Left Ventricle</td>
</tr>
<tr>
<td>Term</td>
<td>Value</td>
<td>Type</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>entire right ventricle</td>
<td>Right ventricle</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>size</td>
<td>size</td>
<td>Spatial Concept</td>
</tr>
<tr>
<td>right ventricular structure</td>
<td>Right Ventricle</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>size</td>
<td>size</td>
<td>Spatial Concept</td>
</tr>
<tr>
<td>function</td>
<td>Function</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>physiological aspects</td>
<td>function</td>
<td>Physiologic Function</td>
</tr>
<tr>
<td>minimal</td>
<td>Mild</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>aortic valve stenosis</td>
<td>Aortic Stenosis</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>minimal</td>
<td>Mild</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>mitral valve stenosis</td>
<td>Mitral Stenosis</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>minimal</td>
<td>Mild</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>tricuspid valve insufficiency</td>
<td>Tricuspid Regurgitation</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>minimal</td>
<td>Mild</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>hypertension, pulmonary</td>
<td>Hypertension, Pulmonary</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>increased</td>
<td>Increased</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>lasix</td>
<td>Lasix</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>recent</td>
<td>Recently</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>improved</td>
<td>Improved</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>dyspnea</td>
<td>Dyspnea</td>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>swollen ankle</td>
<td>ANKLE OEDEMA</td>
<td>Finding</td>
</tr>
<tr>
<td>reduced</td>
<td>Reduced</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>animal mane</td>
<td>Mane</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>morning</td>
<td>mane</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>vision</td>
<td>seeing</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>potassium</td>
<td>Potassium</td>
<td>Element, Ion, or Isotope</td>
</tr>
<tr>
<td>overall publication type</td>
<td>Overall</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>condition</td>
<td>Heart</td>
<td>Body Part, Organ, or Organ Component</td>
</tr>
<tr>
<td>condition</td>
<td>Condition</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>minimal</td>
<td>Mild</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>probable diagnosis</td>
<td>Likely</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>edema</td>
<td>Edema</td>
<td>Pathologic Function</td>
</tr>
<tr>
<td>more</td>
<td>More</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>probable diagnosis</td>
<td>Likely</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>related personal status</td>
<td>Related</td>
<td>Finding</td>
</tr>
<tr>
<td>recent</td>
<td>Recent</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>immobile</td>
<td>Immobility</td>
<td>Finding</td>
</tr>
<tr>
<td>weight gain</td>
<td>Weight Gain</td>
<td>Finding</td>
</tr>
<tr>
<td>antibiotic prophylaxis</td>
<td>Antibiotic Prophylaxis</td>
<td>Therapeutic or Preventive Procedure</td>
</tr>
<tr>
<td>using</td>
<td>Use</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>utilization qualifier</td>
<td>use</td>
<td>Functional Concept</td>
</tr>
<tr>
<td>regular exercise</td>
<td>Regular exercise</td>
<td>Daily or Recreational Activity</td>
</tr>
<tr>
<td>able</td>
<td>Able</td>
<td>Finding</td>
</tr>
<tr>
<td>[d] sinus bradycardia</td>
<td>[D]Sinus bradycardia</td>
<td>Pathologic Function</td>
</tr>
<tr>
<td>stage</td>
<td>Stage</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>tumor stage</td>
<td>Stage</td>
<td>Clinical Attribute</td>
</tr>
<tr>
<td>asymmetric</td>
<td>Asymptomatic</td>
<td>Finding</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>episode, nos</td>
<td>Episodes</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>presyncope</td>
<td>Presyncope</td>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>syncope</td>
<td>Syncope</td>
<td>Sign or Symptom</td>
</tr>
<tr>
<td>future</td>
<td>Future</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>problem</td>
<td>Problem</td>
<td>Finding</td>
</tr>
<tr>
<td>organized</td>
<td>Organised</td>
<td>Qualitative Concept</td>
</tr>
<tr>
<td>review [publication type]</td>
<td>Review</td>
<td>Intellectual Product</td>
</tr>
<tr>
<td>six</td>
<td>Six</td>
<td>Quantitative Concept</td>
</tr>
<tr>
<td>time</td>
<td>Time</td>
<td>Temporal Concept</td>
</tr>
<tr>
<td>with time</td>
<td>Time</td>
<td>Temporal Concept</td>
</tr>
</tbody>
</table>

Table E.2 Semantic Types used for General Classification strategy

<table>
<thead>
<tr>
<th>Diseases or Syndrome</th>
<th>Types</th>
<th>Finding</th>
<th>Diagnostic Procedure</th>
<th>Therapies</th>
<th>Medications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease or Syndrome</td>
<td></td>
<td>Finding</td>
<td>Diagnostic Procedure</td>
<td>Therapeutic or Preventive Procedure</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td>Neoplastic Process</td>
<td></td>
<td>Sign or Symptom</td>
<td>Health Care Activity</td>
<td>Physiologic Function</td>
<td>Organic Chemical</td>
</tr>
<tr>
<td>Pathologic Function</td>
<td></td>
<td>Laboratory or Test Result</td>
<td>Organism Function</td>
<td>Steroid</td>
<td></td>
</tr>
<tr>
<td>Medical Device</td>
<td></td>
<td>Laboratory Procedure</td>
<td></td>
<td>Inorganic Chemical</td>
<td></td>
</tr>
</tbody>
</table>

Table E.3 Results for General Classification for each letter where TP stands for true positive, FP for false positive, FN for false negative, Re for Recall, Pr for Precision, and Fm for F-measure

<table>
<thead>
<tr>
<th>Letter1</th>
<th>Letter2</th>
<th>Letter3</th>
<th>Letter4</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>FP</td>
<td>FN</td>
<td>Re</td>
</tr>
<tr>
<td>G.Sym</td>
<td>10</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>G.Dis</td>
<td>12</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>G.Med</td>
<td>7</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>G.Dig</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G.The</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>0.88</td>
<td>0.93</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Letter5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
</tr>
<tr>
<td>G.Sym</td>
</tr>
<tr>
<td>G.Dis</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Letter7</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter8</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter9</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter10</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter11</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter12</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter13</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter14</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter15</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Letter16</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
</tr>
</tbody>
</table>

Table E.4 Results for Cardiology Classification for each letter where TP stands for true positive, FP for false positive, FN for false negative, Re for Recall, Pr for Precision, and Fm for F-measure.
<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>C.Dis</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>C.Dig</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**AVERAGE** 1.00 1.00 1.00

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**AVERAGE** 1.00 1.00 1.00

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0.63</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**AVERAGE** 0.90 1.00 0.95

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0.50</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**AVERAGE** 0.88 1.00 0.94

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.89</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**AVERAGE** 0.95 1.00 0.97

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**AVERAGE** 0.94 1.00 0.97

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**AVERAGE** 0.96 1.00 0.98

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**AVERAGE** 0.92 1.00 0.94

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0.67</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**AVERAGE** 0.93 0.97 0.95

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Re</th>
<th>Pr</th>
<th>Fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Sym</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dis</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Med</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>0.80</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.Dig</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>C.The</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**AVERAGE** 0.95 1.00 0.97
APPENDIX F

CLINICAL KNOWLEDGE ASSISTANCE SYSTEM

In this appendix, we present the the screen shots of “Clinical Knowledge Assistance” (CKA) system, which covers different functionalities of the CKA. Figure F.3 shows the indexing process of EX-KC with its ID ACS_06_KC1.xml. In CKA, indexing can also be performed automatically as new Ex-KCs are created. Figure F.4 shows the process of generating query from Ex-KCs to retrive and link online medical literature. This process can also be done automatically without manual intervention once a new Ex-KC or Ex-KCs is/are created. Figure F.5 shows the generated query for Ex-KC Id (shown in Figure F.4) and coresponidng retrieved medical articles from Pubmed that are linked to this Ex-KC. Figure F.7 shows the information-view of processed letter. Figure F.8 Information-view of letter presents general classification and cardiology classification of medical concepts found in the letter. Figure F.9 information-view shows potential query terms and interface to formulate customized query. For example, a query having two medical concepts has been keyed in that retrived the computerized CPGs segments and related medical articles shown in Figure F.10. Clinking on to any CPG segment opens its detailed view with its related articles and other segments. Figure F.11 shows an interface of submitting user query without processing a letter. Once a query is submitted, CPGs segments are retrived and displayed as shown in Figure F.10. Furthermore, clinking on any retrived segment opens a detailed view. Figure F.12 shows a detailed view of a retrived CPG segment, for a query shown in Figure F.11, after clicking on a link provided on a retrived results page.
Welcome to Clinical Knowledge Assistance Administration

- Create Extended Knowledge Components
- Start Indexing Ex-KCs
- Generate Ex-KC Query
- Process Medical Referral Letter
- Get Knowledge Assistance Without Processing Letter

**Figure F.1** Main Interface for CKA administration and navigation to its different functions

**Figure F.2** Showing GEM-encoded CPG to create Ex-KCs
Indexing Extended-Knowledge Components

A directory can be selected to index all Ex-KCs in it

Ex-KC file name that is being indexed.

Figure F.3 Showing the indexing process of EX-KC

Generating Queries to Link Extended-Knowledge Components with online Medical Literature

Queries can be generated for multiple newly created Ex-KCs

Query to be generated for Ex-KC to retrieve and link medical literature form Pubmed

Figure F.4 Showing the process of generating query from Ex-KCs
Retrieved and Linked Medical Article to Extended-Knowledge Component

Generated Query for Ex-KC "ACS. 06: Ex-KC" is:

reperfusion therapy, st segment elevation, myocardial infarction, percutaneous coronary intervention. Coronary artery bypass graft

Its query type is: Therapy

Following are Retrieved and Linked Medical articles:

- Abciximab in patients with acute ST-segment-elevation myocardial infarction undergoing primary percutaneous coronary intervention after thrombolysis: a randomized double-blind trial
- Trends in acute reperfusion therapy for ST-segment elevation myocardial infarction from 1999 to 2006: we are getting better but we have got a long way to go
- Distal embolic protection during percutaneous coronary intervention in patients with acute coronary syndromes: the RUBY study
- Clinical characteristics and management of patients with ST-segment elevation myocardial infarction in China: survey of >510 cases
- Gender differences in acute non-ST-segment elevation myocardial infarction
- Early administration of rescue intravenous abciximab vs abciximab-alone in patients with acute myocardial infarction referred for percutaneous coronary intervention: a randomized controlled trial. Related Articles: Free article at journal site
- Mortality benefit of immediate recanalization of acute ST-segment elevation myocardial infarction in patients with contraindications to thrombolytic therapy: a propensity analysis

Figure F.5 Showing the generated query for Ex-KC coresponding retrieved medical articles from Pubmed

Process Referral or Response Letter

Path of the letter that need to be processed

Figure F.6 Showing the interface to select the letter that will be processed by the CKA
Thank you for writing to us, ..., when I saw today the 16th November, 2007. She recently injured her right foot with frostbite and has been relatively sedentary for a while and during this time has developed right calf ulcers bilaterally. The ulcers are there most of the day, and does not appear to become worse during the day. Also however, she has noticed dyspnoea going up stairs, but she has also gained a significant amount of weight after being treated with ilinizide for a breast cancer. Her ... had no history of ischaemic heart disease but she did have rheumatic fever aged eight.

She had a breast excision in January this year with a limited resection being followed by radiotherapy and now long-term hormonal therapy. Risk factors for heart disease include hypertension and hypercholesterolaemia.

Her medications currently include Alendronate, aspirin, Asthma lotion, Atorvastatin, Lasix, Losartan, Metformin, Nexium, Panutol, and Prolific.

### Figure F.7: Showing the information-view of processed letter

#### General Classification of Medical Concepts Found in the Letter

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Diseases</th>
<th>Therapies</th>
<th>Medications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>Hypertensive Cardiomyopathy</td>
<td>Thrombolytic Therapy</td>
<td>Amlodipine, Losartan</td>
</tr>
<tr>
<td>High Fasting</td>
<td>Acute Myocardial Infarction</td>
<td>Penicillin</td>
<td>Atorvastatin, Lipitor</td>
</tr>
<tr>
<td>Glucose</td>
<td>Acute Myocardial Infarction</td>
<td>Amiodarone</td>
<td>Atorvastatin, Lipitor</td>
</tr>
<tr>
<td>Glucose</td>
<td>Acute Myocardial Infarction</td>
<td>Amiodarone</td>
<td>Atorvastatin, Lipitor</td>
</tr>
<tr>
<td>Glucose</td>
<td>Acute Myocardial Infarction</td>
<td>Amiodarone</td>
<td>Atorvastatin, Lipitor</td>
</tr>
</tbody>
</table>

### Figure F.8: Information-view showing general classification and cardiology classification
Figure F.9 Information-view showing potential query terms and interface to formulate customized query

Figure F.10 Showing the retrieved CPGs segments for the query presented in Figure F.9
Accessing Clinical Evidences without Processing Letters

Figure F.11 Showing an interface of submitting user query without processing a letter

Knowledge Segments in Detail

Reperfusion therapy Procedures

Recommendation: Reperfusion therapy Procedures for ST segment elevation myocardial infarction

Recommendation Type = Imperative:

Reperfusion may be obtained with fibrinolytic therapy or Percutaneous Coronary Intervention. A combination of fibrinolysis and Percutaneous Coronary Intervention may also be used facilitated or recent Percutaneous Coronary Intervention. Coronary artery bypass graft surgery may occasionally be more appropriate particularly in patients who have suitable anatomy and are not candidates.

Related Knowledge Segments

Choice of reperfusion therapy procedures
CT segment elevation myocardial infarction

Related Evidence-based Medical Literature

Abnormal in patients with acute ST segment elevation myocardial infarction undergoing primary percutaneous coronary intervention after cluebaiger.

Figure F.12 Showing detailed view of a retrieved CPG segment
In this Appendix, we have provided some of the letters.

Thank you for asking me to see Mrs. ---, a very pleasant lady who is a keen golfer. Gradually over more than a year she has been noticing that she is slower with her regular walks and golf, mostly fatigue, but some dyspnoea. She has had no chest discomfort and no palpitations or syncope recently. However she is rightly concerned because she has a strong family history of cardiac problems and her mother before she had serious cardiac problems complained of fatigue. Mrs. --- stopped smoking 40 years ago and has hypertension. Her pathology you provided shows a total cholesterol of 5.2, and HDL cholesterol 1.4. Interestingly her blood tests also show that her TSH is gradually rising and although it is still in the normal range, she may well develop hypothyroidism in the future. Her medications include Avapro HCT 150/12.5mg mane and Deptran 20mg nocte. She also uses Glucosamine, Fish oil and vitamin B for her knee arthritis.

On examination the pulse was regular at 70 beats per minute, blood pressure 130/80. There were no signs of heart failure, the peripheral pulses were normal without radiofemoral delay and there was no sign of abdominal aortic aneurysm. The heart sounds were normal and the chest was clear. ECG at rest showed sinus rhythm and was within normal limits.

A stress echo was performed today. The resting echo showed normal left ventricular size and function and normal right ventricular function as well. There was thickening and slight calcification of the non-coronary aortic leaflet with no haemodynamic abnormality.

She was able to exercise on the Bruce Protocol for four and a half minutes without chest discomfort, but she did become fatigued. She reached 100% of her predicted maximum heart rate. The heart rate and blood pressure responses were normal. The stress ECG was normal and the echo afterwards showed no evidence of segmental wall motion abnormality and as a result was a normal study. She did however have several episodes of supraventricular tachycardia during the resting phase. This was asymptomatic.

I have organised a Holter monitor for her to wear when she is playing golf and walking to see if she does have episodes of SVT which maybe contributing to her symptoms. I will review her then.

With kind regards,

Yours sincerely.

---

Figure G.1 Medical letter 1
Thank you for your note concerning Mrs —, whom I saw today the 18th April, 2008. She was recently admitted to St. George Hospital with a history of chest discomfort, which she described as a rather sharp discomfort radiating to the left arm and associated with dyspnoea. In hospital there were no ECG changes of ischaemia and the Troponins were negative. A CTPA was performed to exclude pulmonary embolism which was negative and an exercise Sestamibi scan was performed looking for ischaemia, which suggested there was a small area of inferoapical ischaemia and she was only able to exercise for three minutes on the treadmill. It maybe that this was a false positive, but given her history today of gradually decreasing effort tolerance with dyspnoea and chest discomfort, along with also chest discomfort at rest and given her risk factors, it seems likely she would have some coronary disease.

In terms of risk factors, she stopped smoking 20 years ago, has diabetes and hypertension.

On examination she weighed 156kg. The pulse was regular at 60 beats per minute and the blood pressure was 140/80. There were no signs of heart failure and the heart sounds were normal.

The ECG showed sinus rhythm and was normal.

Echocardiography was performed largely to exclude pericarditis. This showed normal left ventricular size and function when adjusted for her body surface area. There was slight thickening of the mitral valve with trivial mitral regurgitation with really no significant valve disease. The arch of the aorta appeared to be normal, with no aortic dilatation of the ascending aorta as well and the interatrial septum was normal. There was no pericardial effusion. Her diastolic study suggested normal diastolic function.

On balance, I feel we should continue with medical therapy for ischaemic heart disease and I have started Lipitor 40mg nocte today, increased the Imdur to 60mg mane, continued with Betaloc 25mg b.d. as well as Diaformin 500 b.d. Zoloft 150mg mane and Aspirin 150mg mane. She does take Celebrex and I have suggested she should try and limit the use of this medication if possible.

I have organised a review in three month’s time and have asked her to let me know if she has any problems in the meantime. I have instructed her in the use of sublingual nitrolingual spray.

With kind regards,
Thank you for asking me to see – who came along with her husband whom I know well. Her main problem is severe hypertension. This initially started about four years ago, but in the last twelve months has become much more difficult to manage. A renal angiogram twelve months ago she tells me shows occlusion of the renal artery on one side, but the other renal artery was said to be normal. I understand her electrolytes and creatinine have been stable and upon calling Douglas Hanley Moir her creatinine in August last year was 310, so this is something we need to keep in mind. From a cardiac point of view she has no specific symptoms with no angina or dyspnoea, but her exercise tolerance is limited by knee pain. She has been intolerant of Minipress resulting in presyncopal episodes, but there has been no evidence of arrhythmias. She has no history of ischaemic heart disease, rheumatic fever or pulmonary emboli in the past.

Her medications currently include Aldomet 250mg t.d.s., Actonel 35mg weekly, Avapro 50mg x half b.d., Betaloc 50mg b.d., Lasix 20mg mane, Mobic, Mogadon, Nexium, Ogen, Tramal and sometimes Keflex for urinary tract infection.

On examination the pulse was regular at 60 beats per minute, the blood pressure was 180/90 in the right arm and 170/85 in the left arm. Peripheral pulses were present without radio femoral delay and there were no carotid bruits. There were no signs of heart failure and on auscultation the heart sounds were normal.

ECG showed sinus rhythm with left ventricular hypertrophy.

M – was very worried about the potential for the hypertension to affect her heart, so echocardiography was performed today. She has moderate left ventricular hypertrophy with normal systolic function. There is some thickening of the right coronary leaflet of the aortic valve with mild aortic regurgitation and mild mitral regurgitation. There was evidence on doppler of diastolic dysfunction and elevated left atrial pressure.

When managing her blood pressure we need to be careful of her renal function, but today I have cautiously increased the Avapro to 150mg b.d. asking her to keep an eye on her blood pressure at home and report to you and her renal physician. If this is ineffective I would be inclined to add Norvasc to the regime.

I have suggested she have a check of her electrolytes and creatinine next week as well. In terms of the aortic regurgitation, the valve is slightly abnormal and I have advised her to have antibiotic prophylaxis for this.

With kind regards.

Figure G.3 Medical letter 3
FINAL LIBRARY COPY