A Collaborative Medical Case Authoring Environment Based on UMLS

Siriwan Suebnukarn  
Thammasat University  
Faculty of Dentistry  
Pathumtani, Thailand, 12121  
ssiriwan@tu.ac.th

Phattanapon Rhienmora  
Asian Institute of Technology  
School of Engineering and Technology  
Pathumtani, Thailand, 12120  
Phattanapon.rhienmora@ait.ac.th

Peter Haddawy  
Asian Institute of Technology  
School of Engineering and Technology  
Pathumtani, Thailand, 12120  
haddawy@ait.ac.th

Abstract

In this paper, we present a novel collaborative authoring tool that was designed to allow medical teachers to formalize and visualize their knowledge for medical intelligent tutoring systems. Our goal is to increase the efficiency and effectiveness in creating the domain model – often referred to as the bottleneck in developing intelligent tutoring systems. We incorporate the Unified Medical Language System (UMLS) knowledge base to assist the authors in creating the problem solution collaboratively via the videoconferencing platform. The system consists of a share workspace gathering information visualization and tools necessary for collaborative problem-solving tasks. We found that the authoring tool can be used as a standalone program to effectively elicit the knowledge structure of the domain model. This was achieved in hours compared to months for the conventional paper-based approach.

1. Introduction

There has been increasing interest in intelligent medical tutoring systems that utilize a wide range of Artificial Intelligence techniques to model the domain knowledge that specifies what to teach, and pedagogic strategies that specify how to teach. This explicit knowledge encoding makes it possible to make inferences about a student’s mastery of topics or tasks in order to dynamically adapt the content or style of instruction [1]. The Collaborative Medical Tutor (COMET) provides an environment that emulates that of human tutored medical problem-based learning (PBL) sessions while at the same time permitting the students to participate collaboratively from disparate locations [2,3]. The system combines concepts from computer-supported collaborative learning with those from Intelligent Tutoring System (ITS). COMET contains four primary components similar to any typical ITS: domain clinical reasoning model (or domain model), student clinical reasoning model (or student model), pedagogic module, and student multi-modal interface. The system implementation is modular and the tutoring algorithms are generic so that adding a new scenario or case requires only adding the appropriate model representing how to solve a particular case (domain clinical reasoning model). Creating the domain model is not a trivial task and requires significant expert knowledge and time. The domain model for each case required about one person month to build for a 3-hour PBL tutorial session. The aim of this study was to develop an authoring tool for medical PBL tutor to assist the creation of new cases.

Researchers have been investigating ITS authoring tools almost since the beginning of ITS research in order to allow non-programmers to formalize and visualize their knowledge. Several authoring tools have been built for authoring the domain model, e.g., IDE [4], Eon [5]. They were all designed to work in a single-user mode. However, developing a complicated case particularly in medical PBL usually requires experts from different areas such as anatomist, physiologist, pathologist and specialist in the field. The necessity to collaborate to produce high-quality PBL cases lead us to the requirement of collaboration support in the authoring tool, which is the first contribution of our work.

A second issue is to support use of standardized medical terminology. In medical problem solving, there are often many ways of expressing the same idea using synonymous phrases, like “Pneumonia”, “Lung inflammation” or “Pulmonary inflammation”. Consequently, free text is not suitable to represent medical problem solutions in the domain model where student hypotheses are matched against. Terminological standardization helps to solve this problem. Furthermore, using standard terminology permits ready linking to rich sources of medical
knowledge. Our approach is to integrate the Unified Medical Language System (UMLS) [6] directly into the authoring environment in such a way that case authoring becomes a process of browsing through UMLS and selecting the relevant medical concepts. This ensures standardization of terminology and linkage of the problem solution to a rich source of medical knowledge, which can be used by the tutoring module. An additional benefit is that because authoring becomes a process of browsing and selecting, the authors may be reminded of hypotheses that they might otherwise neglect to include.

The UMLS was designed by the National Library of Medicine to integrate many authoritative biomedical source terminologies into a unified knowledge representation. The UMLS knowledge base includes: (i) a metathesaurus of terms and concepts from many different biomedical vocabularies and classifications. Each concept in the metathesaurus is assigned to one or more semantic types; (ii) a semantic network of relationships among the semantic categories to which concepts of metathesaurus are assigned; and (iii) a specialist lexicon and programs for generating the lexical variants of biomedical terms. There have been several attempts to re-use the UMLS knowledge base in a variety of contexts, e.g., in patient care [7], clinical radiology [8]. We are the first to exploit the UMLS knowledge base in a domain model authoring tool.

2. COMET authoring tool description

The overall goals of the COMET authoring tool are to: i) decrease the effort (time, cost, other resources) for creating the domain models; ii) decrease the skill threshold for building intelligent tutors by allowing non-programmers to take part in the design process; iii) help the authors articulate or organize their domain knowledge; and iv) support good design principles. Consequently, our interface incorporates UMLS for knowledge management and knowledge re-use, as well as the graph visualization for knowledge visualization with the tools essential for collaborative work via the videoconference.

2.1. Knowledge management

As in many other ITSs, COMET predefines some of the main components including the tutoring module, student model and student interface, and requires from the author to construct only the domain model for each scenario. Essentially, authors who are medical PBL experts, are asked to create the problem solution. In this process, medical problem-solving consists in the real-time construction of a problem model in the form of a network comprising hypothesis nodes characterized by their ontological levels – enabling conditions, faults and consequences – that the authors can enumerate further from a scenario. The semantic relations are cause-effect relationships among hypotheses. Every hypothesis node has a unique apply node as one of its parents. The apply node represents the application of a medical concept to a goal in order to derive the hypothesis. During this process, the problem is progressively transformed into a solution model by assigning particular goals and medical concepts to each hypothesis. The resulting network which is a shared mental model of the authors becomes the structure of the Bayesian network (BN) domain model. The details of the BN domain model are explained in [9].

The COMET authoring tool provides online communication among authors from different locations via text chat and video conferencing. Connection to medical terminology in the Unified Medical Language System (UMLS) is included so that the effort required to bind medical concepts to nodes in the model is reduced.

Figure 1. User interface of the COMET authoring tool.

Figure 1 illustrates the main interface of the tool. The drawing pane dominates most of the interface and allows interaction and visualization of the created model. Basic graphical tools are provided in the toolbar e.g. copy, paste, undo, redo and zoom. The experts with the write privilege can edit the model simultaneously while those with read privilege can only see the changes in real-time and discuss with others via the chat pane.

A hypothesis node can be created from the context menu and a link can be drawn to represent the cause-effect relationship between two nodes. Colors are used
to indicate different kinds of nodes. Goal, concept and apply nodes are shown in gold, pink and blue respectively. Hypothesis nodes have three sub-categories, enabling condition, fault and consequence. They are in red, yellow and green respectively as shown in Figure 1. Double clicking each node brings up its property window as shown in Figure 2. The Search tab allows searching for a medical concept in UMLS. Returned concept results including their explanation can be selected and added to a node name. Note that the input for searching can be a phrase or a sentence containing many medical terms. The Metamap Transfer (MMTx) program provided by the UMLS is called to convert the input string to a list of medical concepts to be searched.

Below the drawing pane is the chat pane. The last pane located on the right is the video conference interface. This allows experts with web camera and microphone to communicate via voice and video in addition to text. The video and audio conference is implemented over the Java Media Framework (JMF) using the unicast Real-Time Protocol (RTP). Changes made to model data are sent to all authors synchronously using TCP/IP socket. Text data in chat pane is also transmitted this way.

2.2. Knowledge re-use

There have been several attempts to reuse the UMLS knowledge base in a variety of contexts, e.g., patient care and clinical laboratory [7]. The COMET authoring tool is built to take advantage of the reusability of concept relations in the UMLS semantic networks. We have experimented with the UMLS knowledge base for providing classes, concepts, terms and interconcept links for creating the medical PBL domain model. Browsing through the UMLS semantic networks helps remind the author the relationships among the concepts which can be added to the domain model. Figure 3 shows part of pneumonia model built using the UMLS semantic network of pneumonia concept, its parent and child.

![Figure 3. Part of pneumonia model built using the UMLS semantic network of pneumonia concept and its parent (Lung consolidation) and child (Alveolitis).](image)

2.3. Knowledge visualization

One effective way to help authors comprehend large amounts of complexly interconnected knowledge is with visualization tools. The system provides key features such as copying, pasting, zooming, undo/redo, event-handling, and drag-and-drop support. These functions are implemented using JGraph—an open source Java graph visualization library [10]. JGraph is compliant to Java Swing API. Therefore, it is compatible with Swing features such as serialization, and data-transfer.

The medical case for our application has its specific layout. The goal and concept nodes, which are parents of apply nodes, are placed on the left most. The apply nodes, which can have multiple parents, are placed next to the right. The hypothesis nodes are placed on the right most and in hierarchical layout to illustrate the causal relationship among them. The authors can draw a case easily by following this layout; however, the authoring tool also provides the automatic layout feature using JGraph [10] to help creating a better visualization of the layout such as moving nodes to the
straight alignment and minimizing the crossing of edges.

3. Evaluation

3.1. Effectiveness of the authoring tool

In order to evaluate the effectiveness of the authoring tool, we measured the quality of the created problem solution. The problem solution is in the form of network consisting of nodes (hypotheses) linked together by semantic relations. We compared the network built using COMET authoring tool, with the network of a benchmark solution for the same domain. The benchmark solution was built from scratch using a paper-based problem solving approach.

Six medical PBL tutors participated in the study: 2 physiology majors, 2 anatomy majors, 2 pathology majors. Each tutor had at least five years experience in conducting the PBL course at Thammasat university medical school. The participants were divided into two groups with the same proportion of expertise. Each group was asked to create a problem solution for three different scenarios – Pneumonia, Heart attack and Diabetes – which provide a reasonable diversity of the case. The first group used COMET authoring tool via the school LAN to elicit the problem solution on given domains collaboratively, and the resulting semantic networks of the problem solutions were compared with the benchmark group. The benchmark group created the problem solution collaboratively using the conventional paper-based problem solving approach. Medical Textbooks, journals and online resources were used in both groups.

All valid medical hypotheses and the links created were counted. The validity of hypotheses and links were verified by an expert in the area of Pneumonia, Heart attack and Diabetes. In order to represent a global property of the networks that could account for the interdependence of the elements in the representation, we used a numeric function called Integration [11]. An integration mark (I) is defined by taking the ratio of the number of links (L) within the network to the number of nodes (N). We can generalize that formula for the whole representation: 

$$ I = \frac{L}{N} $$

Figure 4 shows resulting semantic networks for Pneumonia scenario created by COMET authoring tool and benchmark group respectively. The numbers of hypotheses, links and value of integration of all scenarios are presented in Table 1. With the support from COMET authoring tool, there was a greater in the number of nodes, the number of links, and the value of integration demonstrated by the COMET authoring tool group in all scenarios. The results show that from the same scenario, the authors who work with the COMET authoring tool built solution networks that were both richer and more integrated.

Figure 4. Resulting semantic networks of Pneumonia case
is the use of colors to represent hypothesis categories. This helps the users to quickly perceive and comprehend the evolving model, particularly in a multi-author environment. This feature could explain the fact that there are more ideas related to the enabling condition for the Upper respiratory tract infection in the COMET authoring tool group (Fig. 5, Box A and C).

### Table 1. Numbers of hypotheses and links as well as value of integration, by cohorts and scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Cohort</th>
<th>Number of hypotheses</th>
<th>Number of links</th>
<th>Value of integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td>Benchmark</td>
<td>26</td>
<td>28</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>COMET</td>
<td>35</td>
<td>43</td>
<td>1.23</td>
</tr>
<tr>
<td>Heart attack</td>
<td>Benchmark</td>
<td>22</td>
<td>23</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>COMET</td>
<td>36</td>
<td>49</td>
<td>1.36</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Benchmark</td>
<td>26</td>
<td>29</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>COMET</td>
<td>34</td>
<td>49</td>
<td>1.44</td>
</tr>
<tr>
<td>All</td>
<td>Benchmark</td>
<td>74</td>
<td>80</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>COMET</td>
<td>105</td>
<td>141</td>
<td>1.34</td>
</tr>
</tbody>
</table>

### 3.2. Usability of the authoring tool

COMET authoring tool has been tested and improved to fulfill the authors’ requirements during the system design and implementation. Initial training in the use of the tool requires between 1 and 2 h. No author has found the overall creating the problem solution incompatible with his or her approach. Authors took between 6 and 11 h (to author a 3 h problem analysis session on pneumonia, heart attack and diabetes) – a ratio of around 3 h per hour of tutoring. For the most part authors wanted the functionality that COMET provides and they could use it to create problem solutions that reflect their own pedagogic preferences.

### 4. Conclusions and future work

We have exploited Computer-supported collaborative work environments and re-used UMLS in the development of COMET authoring tool. This has allowed us to provide flexible mechanisms and interfaces to allow authors to collaboratively building the domain knowledge of the medical PBL cases. The evaluation showed that the authors who worked with the COMET authoring tool built solution networks that were both richer and more integrated than the network of a benchmark solution built from scratch using the conventional paper-based approach for the same domain. This was achieved in hours compared to months for the conventional paper-based approach.

Although our primary goal was to build the authoring tool for COMET, we believe that the tool will also be useful for case authoring in medical PBL in general. In our future work, we plan to develop reasoning algorithms from the UMLS semantic networks in order to automatically convert them into the medical ITS domain models.

### 5. References


