

Expanding the Plausible Solution Space for Robustness in an Intelligent Tutoring System

Hameedullah Kazi^{1,3}, Peter Haddawy¹, and Siriwan Suebnukarn²

¹ Computer Science & Information Management Program,
Asian Institute of Technology, Thailand

² School of Dentistry, Thammasat University, Thailand

³ Department of Computer Science, Isra University, Pakistan
{hameedullah.kazi,haddawy}@ait.ac.th, ssiriwan@tu.ac.th,
hkazi@isra.edu.pk

Abstract. The knowledge acquisition bottleneck is a problem pertinent to the authoring of any intelligent tutoring system. Allowing students a broad scope of reasoning and solution representation whereby a wide range of plausible student solutions are accepted by the system, places additional burden on knowledge acquisition. In this paper we present a strategy to alleviate the burden of knowledge acquisition for building a tutoring system for medical problem-based learning (PBL). The Unified Medical Language System (UMLS) is deployed as domain ontology and information structure in the ontology is exploited to make intelligent inferences and expand the domain model. Using these inferences and expanded domain model, the tutoring system is able to accept a broader range of plausible student solutions that lie beyond the scope of explicitly encoded solutions. We describe the development of a tutoring system prototype and report the evaluation of system correctness in accepting such plausible solutions. The system evaluation indicates an average accuracy of 94.59 % when compared against human domain experts, who agreed among themselves with a statistical agreement based on Pearson Correlation Coefficient of 0.48 and $p < 0.05$.

Keywords: Robustness, intelligent tutoring systems, medical problem-based learning, UMLS, knowledge acquisition bottleneck.

1 Introduction

Intelligent tutoring systems typically present a problem scenario to the students, who solve the problem and receive hints from the system to help them in reasoning towards the correct solution. Solutions presented by students are evaluated by comparing them against a particular solution accepted by the tutoring system as being correct. Thus plausible student solutions that do not match that particular solution or a small set of stored solutions recognized by the system are often rejected by the system as being incorrect. This forces students to memorize expert solutions and stifles student creativity. Students should be able to use their understanding of concepts and concept relationships and learn how to apply their knowledge to given problems. This is particularly relevant in medical PBL, where a diverse set of solutions may be acceptable instead of a single perfect solution.

Plausible solutions may differ along a number of dimensions such as comprehensiveness, level of detail and the choice of alternate and synonymous terms to describe a concept. Different solutions to a problem may also be acceptable depending on the perspective one chooses to analyze from and this may particularly be applicable to ill-defined domains such as medical education [1]. The value of being able to accept a broader set of solutions is that it supports an approach to learning that promotes free thinking and novel solutions. To the best of our knowledge, the task of accepting a range of solutions larger than the explicitly encoded scope of solutions, is yet to be addressed in intelligent tutoring systems.

In order for a tutoring system to exhibit robust human-level tutoring, it needs broad knowledge to allow students to explore a large space of solutions and work creatively. Authoring tutoring systems typically requires knowledge acquisition in the three areas of domain expert knowledge, student model and pedagogical model [2]. Domain expert knowledge is acquired to equip the system with the curriculum knowledge that is to be taught to the students; pedagogical knowledge is acquired to equip the system with teaching techniques and strategies; and knowledge of the student model is acquired to help the system assess the knowledge level of the student. Acquiring and encoding the relevant knowledge can lead to a large overhead in the development time of a tutoring system [3, 4]. Manually encoding all knowledge into the system so that it can accept the full range of plausible solutions is not feasible and places great burden on human domain experts, whose time is very costly. A natural choice to overcome this knowledge acquisition bottleneck is to make use of existing knowledge available for reuse and sharing. The use of ontologies is a viable alternative in reducing the burden of knowledge acquisition for knowledge based systems.

Ontologies have been employed in the design of various tutoring systems [5, 6, 7], which often require cumbersome encoding of the ontology. The Constraint Acquisition System [8] uses a more automated approach of encoding the ontology constraints by learning from examples using constraint based modeling. However, it still requires the initial design of the ontology to be defined manually.

In the next few sections we describe how the burden of knowledge acquisition can be lightened for a medical tutoring system, through the use of the broad and widely available medical knowledge source UMLS, distributed by the U.S. National Library of Medicine [9]. We also describe a mechanism of exploiting the information structure in UMLS, through which the tutoring system can accept a range of plausible solutions larger than the explicitly encoded scope of solutions.

2 Related Work

The issue of brittleness and burden of knowledge acquisition has been addressed in the design of various intelligent tutoring systems [10, 11]. Kumar [10] discusses the use of model-based reasoning for domain modeling in the context of web-based tutoring system for helping students to learn to debug C++ programs. Arguing that rule-based systems are not flexible and are not adaptable to varying system behavioral discrepancies, he proposes model-based reasoning as an alternative.

The KASER [11] design is implemented in a tutoring system that teaches the science of crystal-laser design. They argue that production rules, when found to be in error, are corrected through explicitly encoding the revised rule back into the system.

They propose a production system, which initially requires explicit encoding of rules and can later self generate other rules as extension to the ones created earlier, easing the burden of knowledge acquisition.

The designs of medical tutoring systems built to date, have typically been based on customized knowledge bases that offer students a limited set of medical terms and concepts, to form their solution. The CIRCSIM-Tutor [12] teaches cardiovascular physiology by describing a perturbation of a cardiovascular condition, and initiating a question answer dialog with the student, to help the student in reasoning towards the correct solution. The system design lays emphasis on qualitative reasoning, but the scope of hypothesis (solution) representation is narrow, as students are confined to assigning values to a small set of variables for forming their hypothesis.

The SlideTutor [5] teaches students dermatopathology by presenting a visual slide as a problem scenario and asks students to classify the diseases. After observing the visual evidence presented in the slide, students present their hypothesis through a mouse driven menu selection, identifying features and their attributes from an ontology that has been manually encoded for the problem scenarios fed to the tutoring system. Solutions accepted by the tutoring system are also based on the ontology customized for the system. Thus students are not allowed to present alternative plausible hypotheses that may lie beyond the scope of this customized ontology.

This motivates the need to have a medical tutoring system that offers students a broad knowledge base of medical concepts and terms, such as the UMLS [9], to select their hypothesis concepts. This tutoring system should also be able to accept a wide variety of plausible hypotheses to a given problem.

3 Medical PBL and System Prototype

In a typical PBL session in the medical domain, a problem scenario is presented to a group of 6-8 students, who form their hypothesis in the form of a causal graph, where graph nodes represent hypothesis concepts and directed edges (causal links) represent cause effect relationships between respective concepts. The hypothesis graph is based on the Illness Script, where hypothesis nodes may represent enabling conditions, faults or consequences [13]. Enabling conditions are factors that trigger the onset of a medical condition, e.g., aging, smoking, etc.; faults are the bodily malfunctions that result in various signs and symptoms, e.g., pneumonia, diabetes, etc.; consequences are the signs and symptoms that occur as a result of the diseases or disorders, e.g., fatigue, coughing, etc.

There is difference of opinion on whether experts or non-experts should be used as a PBL tutor or facilitator [14, 15]. Different PBL tutors may also disagree over the extent to which a causal link is acceptable. While one tutor may find a causal link perfectly acceptable, another tutor may only be inclined to accept it with reservation. Quite often a PBL tutor may accept varying solutions that may differ in the choice of alternate terms or also in the level of detail with which the solutions are presented.

Our work is based on the extension of the COMET system [16, 17] designed to cover medical PBL for various domains. In the COMET system, each problem scenario is first referred to human domain experts who provide an expert solution that is

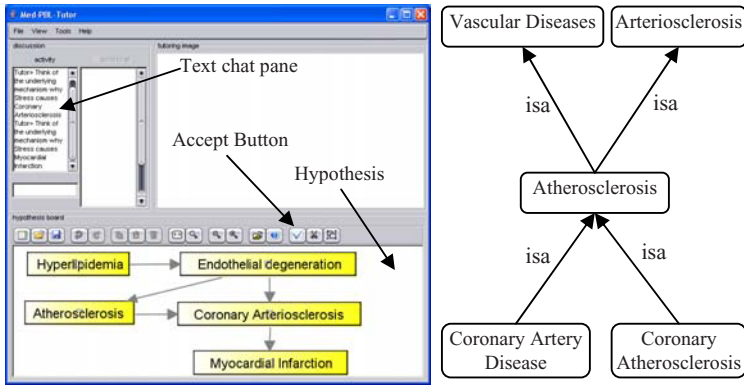


Fig. 1. (a) System Prototype Interface (b) Relationships in UMLS

eventually fed to the system. Student solutions are compared against this expert solution for evaluation. Thus a plausible student solution that does not match the expert solution is not entertained. The system allows students to form their hypothesis by choosing medical concepts from a repository manually encoded into the system.

We have developed a tutoring system for PBL in the medical domain. Problem solutions collected from experts are combined with UMLS tables to form the domain model. The pedagogical module of the system comprises of a hint generation mechanism that leverages off of the UMLS concept hierarchy and provides students a measure of partial correctness of their hypotheses [18]. Since the hint generation employs the rich domain knowledge of the UMLS in lieu of a student model, the design of our tutoring system does not include a student model.

The problem representation in our system is the same as that in COMET of a directed acyclic graph for forming the hypothesis. The student user is provided with a workspace as a hypothesis board to form the hypothesis, along with a text chat pane that returns hints from the system to guide the student in his/her clinical reasoning, as shown in Figure 1 (a). The student chooses concepts from the UMLS Metathesaurus [9] as hypothesis nodes and draws edges between nodes, using a mouse. The problem solving activity begins as the student is presented a problem scenario, such as:

“Mr. Heng-heng is a 48-year-old admitted with severe chest pain” ... “His father died of heart attack aged 55. He smokes 20 cigarettes a day. He is still in pain” ...

After studying the above problem description related to heart attack, the student hypothesizes that *endothelial degeneration* is a cause of *coronary arteriosclerosis*, which is shown to be a cause of *myocardial infarction*, as shown in Figure 1 (a).

4 System Knowledge Base and UMLS Knowledge Source

The UMLS [9] is a widely available medical knowledge source and is essentially a collation of various medical ontologies and terminologies (MeSH, SNOMED-CT, Gene Ontology, etc). The broad and diverse UMLS contains over 1 million medical

concepts covering various medical domains and about 135 semantic types, where each medical concept is assigned at least one semantic type [9].

The UMLS has been studied for use in many intelligent system applications [19, 20, 21, 22]. However, to the best of our knowledge, UMLS has not been used as the main knowledge source for inference purposes in an intelligent tutoring system. The Docs 'n Drug tutoring system [23] employs the use of medical terminologies that are a subset of UMLS, to allow students to choose concepts from these incrementally expandable terminologies. However, this system does not exploit the knowledge structure within these terminologies and make inferences for reasoning purposes.

The design of our system knowledge base comprises of UMLS tables and an additional table that is henceforth referred to as the *expert knowledge base*. The *expert knowledge base* is encoded with the help of human domain experts, and it contains the causal relationship between various medical concepts, such as:

Hyperlipidemia → *Endothelial Degeneration*
Endothelial Degeneration → *Coronary Arteriosclerosis*
Endothelial Degeneration → *Atherosclerosis*
Coronary Arteriosclerosis → *Myocardial Infarction*

Student solutions that are considered acceptable by human experts are merged into the *expert knowledge base*. The *expert knowledge base* is formed through the collation of expert solutions to various problem scenarios, along with the student solutions that are certified by the experts to be correct. The construction of an expert solution requires about 2-3 hours. Since each solution is in the form of a hypothesis graph, the collation of different solutions implies the incremental addition of the causal links in each solution, to the *expert knowledge base*. The goal is to expand the *expert knowledge base* over time. This is achieved through a convenient mechanism of knowledge acquisition, when student hypotheses are intermittently referred to a human tutor. While examining a plausible student hypothesis, the human tutor has the option to click the *Accept Button*, as shown in Figure 1 (a). This adds all links in the student hypothesis being examined, to the *expert knowledge base*.

5 System Reasoning and Hypothesis Evaluation

Each hypothesis causal link drawn by the student is evaluated by the system. The system refers to its knowledge base to check whether the link is acceptable. If the link is found to be acceptable, the system allows the directed edge (causal link) to be drawn; otherwise the system disallows the edge to be drawn and returns an appropriate hint as feedback to the student. The acceptability of the student hypothesis link is evaluated by comparing it against the *expert knowledge base*. If there is a match, the link under evaluation is considered acceptable; however if the link is not found in the *expert knowledge base*, the system makes use of a heuristic method to see if the link is close to acceptable.

This heuristic method makes use of the information structure within UMLS to make inferences. We use the relationships *alike* and *child-parent* for our purpose of heuristic based evaluation of the student solutions. A hierarchy based on *parent-child* relationships between UMLS concepts is shown in Figure 1 (b). While evaluating a

causal link, we consider the *alike* and *parent-child* relationships of each of the parent node and the child node in the causal link. All concepts that are found to have *alike* relationship or *parent* relationship with the parent node are considered equivalent to the parent node, while all concepts that have *alike* or *parent* relationship with the child node are considered equivalent to the child node. For a causal link under evaluation, let set *A* contain all concepts considered equivalent to the parent node and let set *B* contain all concepts considered equivalent to the child node. If the system finds a causal link in the *expert knowledge base* whose parent node is a concept from set *A* and whose child node is a concept from set *B*, then the link under evaluation will be considered acceptable.

Our experiments with expert ratings of causal links, revealed that links with concepts belonging to the four terminologies of Medical Subject Headings (MSH), UMLS Metathesaurus (MTH), Systematized Nomenclature of Medicine-Clinical Terms (SNOMEDCT) and National Cancer Institute (NCI), were found to be more acceptable. Thus in order to reduce the number of erroneous links, the system only accepts causal links whose concepts belong to any of these four terminologies.

The tutoring system starts by evaluating each causal link in the hypothesis separately, and then evaluates the graph as a whole. A hypothesis is considered valid if it contains a chain of reasoning, where nodes representing enabling conditions lead to other nodes in succession that eventually lead to the symptoms.

5.1 Example

Consider the problem scenario related to heart attack described above. Two different solutions shown in Figure 2, considered acceptable by human experts are merged into the *expert knowledge base*. Thus the student hypothesis shown in Figure 3 (a) is accepted by the system, since this hypothesis is justified through the merging of the expert solutions shown in Figure 2.

For illustration of the heuristic method of accepting inferred links, consider the causal link from an expert solution, shown in Figure 3 (b). The expert knowledge base contains this causal link leading from *hypoinsulinism* to *glucose metabolism disorder*. For this link, two lists are generated. The first list L1 comprises of concepts that have *alike* or *parent* relationship with *hypoinsulinism*, so that $L1 = \{hypoinsulinism, diseases\ of\ endocrine\ pancreas\}$. The second list L2 comprises of concepts that have *alike* or *parent* relationship with *glucose metabolism disorder*, so that $L2 = \{glucose\ metabolism\ disorders, metabolic\ diseases, disorder\ of\ carbohydrate\ metabolism\}$.

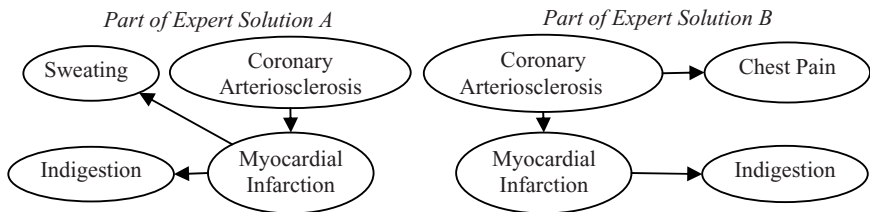


Fig. 2. Expert Solutions merged into the *expert knowledge base*

The space of plausible solutions is expanded, as the system accepts all hypothesis links that are formed through $L1 \rightarrow L2$, such as:

- hypoinsulinism* \rightarrow *glucose metabolism disorder*
- hypoinsulinism* \rightarrow *metabolic diseases*
- hypoinsulinism* \rightarrow *disorder of carbohydrate metabolism*
- diseases of endocrine pancreas* \rightarrow *glucose metabolism disorder*
- diseases of endocrine pancreas* \rightarrow *metabolic diseases*
- diseases of endocrine pancreas* \rightarrow *disorder of carbohydrate metabolism*

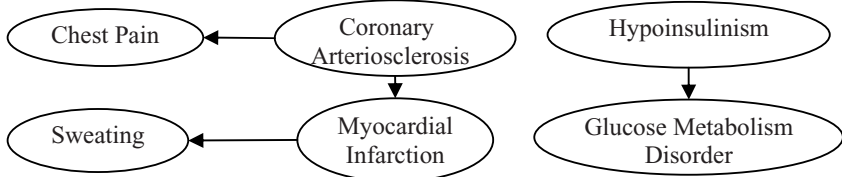


Fig. 3. (a) Student solution accepted by the system (b) Expert Causal Link

6 Results

We conducted system evaluations of the acceptability of causal links beyond those explicitly encoded into the system. We randomly selected 14 causal links from expert solutions to three problem scenarios related to diseases and disorders such as heart attack, diabetes and pneumonia. For each node in the causal link we generated a list of concepts that were found to have *alike* and *parent* relationship with the nodes in the causal link. Thus we generated a pair of lists of concepts for each causal link. We then formed links between each pair of concepts found in the respective lists.

From an initial number of 14 links, the system generated a total of 228 links based on the *alike* and *parent* relationships. These system generated causal links were then presented to a total of 10 medical experts from Thammasat University Medical School, who had at least 5 years of experience in conducting PBL sessions. The experts were asked to rate the acceptability of each link on a scale of 1-5, where 1 implied unacceptable, 2 implied not quite acceptable, 3 implied not sure, 4 implied close to acceptable and 5 implied acceptable. The ratings were so chosen to accommodate the difference of opinion often found among PBL tutors as mentioned in section 3. Figure 4 shows part of the expert evaluation form based on the links generated from the causal link shown in Figure 3 (b).

These 228 links were short listed by medical experts to comprise only of links that could conceivably be formed by medical students according to expert judgment. This resulted in 213 links, which were further short listed to comprise of only those links whose concepts belonged to any of the above mentioned four terminologies: MSH, SNOMEDCT, MTH and NCI. Thus a total of 111 system generated causal links were considered for evaluation. Based on the ratings 1-5 assigned by the medical experts, we computed the average score for each causal link. The overall mean of the ratings came out to be 4.11 with a standard deviation mean of 0.69.

Diabetes case		Acceptability
Causal Links		
Hypoinsulinism-->Glucose Metabolism Disorders		5 4 3 2 1
Hypoinsulinism-->Metabolic Diseases		5 4 3 2 1
Hypoinsulinism-->Disorder of carbohydrate metabolism		5 4 3 2 1
Disorder of endocrine pancreas-->Glucose Metabolism Disorders		5 4 3 2 1
Disorder of endocrine pancreas-->Metabolic Diseases		5 4 3 2 1
Disorder of endocrine pancreas-->Disorder of carbohydrate metabolism		5 4 3 2 1

Fig. 4. Part of Expert Evaluation Form showing inferred links

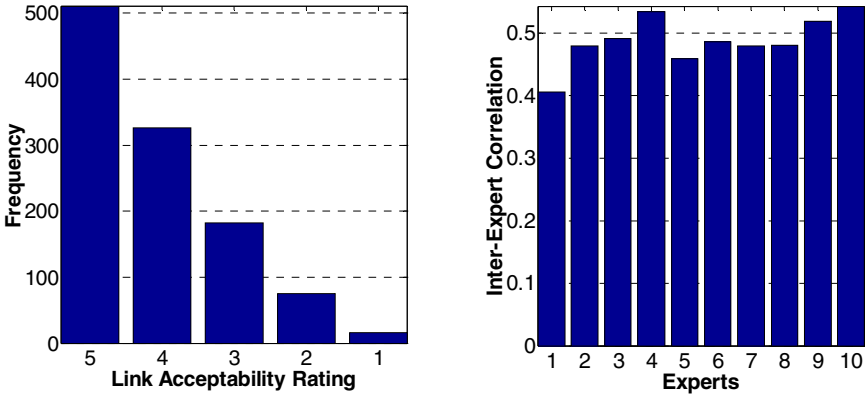


Fig. 5. (a) Distribution of Acceptability Ratings (b) Inter-Expert Correlation

To get a measure of acceptability we collapsed the rating scale to divide the links into acceptable or unacceptable. All links that had an average rating above 3 were considered acceptable, while the rest were considered unacceptable. Based on this criterion, 105 links were found acceptable, whereas 6 links were found unacceptable, leading to an overall accuracy of 94.59 %. The experts were found to agree with each other with a good degree of statistical agreement (Pearson Correlation Coefficient = 0.48, $p < 0.05$). Figure 5 (a) shows the frequency distribution of the 1110 acceptability ratings assigned by 10 experts for 111 samples, whereas Figure 5 (b) shows the correlation values of individual experts with rest of the experts.

7 Discussion

The average score of 4.11 out of 5.0 and the overall accuracy of 94.59 % together with good statistical agreement among expert ratings, indicate that the system is mostly correct in accepting inferred links. As can be observed from the correlation values in Figure 5 (b), the experts were not in perfect agreement with one another over the acceptability of the causal links. Furthermore, the expert evaluations also revealed that in some cases, links inferred by the system scored higher than the corresponding original links created by human experts. For example, the inferred links: *heredity* → *vascular diseases* and *heredity* → *arteriosclerosis* received mean scores

of 4.5 and 4.2 respectively, whereas the original expert link *heredity* → *atherosclerosis* received a mean score of 3.3. This evidence further points to the inherent variation involved in the evaluation of PBL hypothesis links and reinforces the need to have a tutoring system that evaluates student solutions in a broad context.

8 Conclusion

In this paper we have described how to reduce the burden of knowledge acquisition in an intelligent tutoring system. We have described how a broad and easily available medical knowledge source such as UMLS can be deployed as the domain ontology for a tutoring system for medical PBL. We have presented a strategy of making the tutoring system more robust by broadening the scope of solutions that are accepted by the tutoring system as being correct. The inference mechanism for expanding the solution space can also be applied to other domains, where the problem representation is in the form of causal graphs.

We intend to evaluate the effectiveness of the system's tutoring hints vis-à-vis the two major components of hint generation in our system: the measure of partial correctness and the leveraging off of the UMLS concept hierarchy. Finally we intend to conduct evaluations of learning outcomes by assessing the clinical reasoning gains acquired by student users as a result of using this medical tutoring system.

Acknowledgments. We thank the medical experts at Thammasat University Medical School for generously donating their time and effort for the system evaluations.

References

1. Pople Jr., H.E.: Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics. In: Szolovits, P. (ed.) *Artificial Intelligence in Medicine*, ch.5. Westview Press, Boulder (1982)
2. Murray, T.: Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. *International Journal of Artificial Intelligence in Education* 10, 98–129 (1999)
3. Anderson, J.R., Corbett, A., Koedinger, K., Pelletier, R.: Cognitive tutors: Lessons learned. *Journal of the Learning Sciences* 4(2), 167–207 (1996)
4. Mitrovic, A.: Experiences in implementing constraint-based modelling in SQL-Tutor. In: 4th International Conference on Intelligent Tutoring Systems, pp. 414–423 (1998)
5. Crowley, R., Medvedeva, O.: An Intelligent Tutoring System for Visual Classification Problem Solving. *Artificial Intelligence in Medicine* 36(1), 85–117 (2006)
6. Day, M.Y., Lu, C., Yang, J.D., Chiou, G., Ong, C.S., Hsu, W.: Designing an Ontology-Based Intelligent Tutoring Agent with Instant Messaging. In: Fifth IEEE International Conference on Advanced Learning Technologies, pp. 318–320 (2005)
7. Lee, C.H., Seu, J.H., Evens, M.W.: Building an ontology for CIRCSIM tutor. In: 13th Midwest AI and Cognitive Science Society Conference, MAICSS, pp. 161–168 (2002)
8. Suraweera, P., Mitrovic, A., Martin, B.: A Knowledge Acquisition System for Constraint Based Intelligent Tutoring Systems. In: Conference on Artificial Intelligence in Education (2005)
9. U.S. National Library of Medicine, <http://www.nlm.nih.gov/research/umls/>

10. Kumar, A.: Model-Based Reasoning for Domain Modeling in a Web-Based Intelligent Tutoring System to Help Students to Learn to Debug C++ programs. In: 6th International Conference on Intelligent Tutoring Systems (2002)
11. Rubin, S.H., Rush Jr., R.J., Smith, M.H., Murthy, S.N.J., Trajkovic, L.: A Soft Expert System for the Creative Exploration of First Principles of Crystal-Laser Design. In: IEEE International Conference on Systems, Man, and Cybernetics, SMC 2002, Hammamet, Tunisia (October 2002)
12. Mills, B., Evens, M., Freedman, R.: Implementing directed lines of reasoning in an intelligent tutoring system using the atlas planning environment. In: International Conference on Information Technology, pp. 729–733 (2004)
13. Feltovich, P.J., Barrows, H.S.: Issues of generality in medical problem solving. In: Schmidt, H.G., De Volder, M.L. (eds.) *Tutorials in problem-based learning: A new direction in teaching the health professions*. Van Gorcum, The Netherlands (1984)
14. Hay, P.J., Katsikitis, M.: The ‘expert’ in problem based and case-based learning: necessary or not? *Medical Education* 35, 22–26 (2001)
15. Albanese, M.A.: Treading tactfully on tutor turf: Does PBL tutor content expertise make a difference? *Medical Education* 38, 916–920 (2004)
16. Suebnukarn, S., Haddawy, P.: Modeling individual and collaborative problem-solving in medical problem-based learning. *User Modeling and User Adapted Interaction* 16(3), 211–248 (2006)
17. Kazi, H., Haddawy, P., Suebnukarn, S.: Enriching Solution Space for Robustness in an Intelligent Tutoring System. In: 15th International Conference on Computers in Education, Hiroshima, Japan (2007)
18. Kazi, H., Haddawy, P., Suebnukarn, S.: Towards Human-Like Robustness in an Intelligent Tutoring System. In: 8th International Conference on Cognitive Modeling, Ann Arbor, Michigan, USA (2007)
19. Achour, S.L., Dojat, M., Rieux, C., Bierling, P., Lepage, E.: A UMLS-Based Knowledge Acquisition Tool for Rule-Based Clinical Decision Support Systems Development. *Journal of the American Medical Informatics Association* 8(4), 351–360 (2001)
20. Burgun, A., Bodenreider, O.: Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. *MedInfo* 10(Pt 1), 171–175 (2001)
21. Mendonca, E.A., Cimino, J.J.: Automated Knowledge Extraction from MEDLINE Citations. In: *AMIA 2000 Fall Symposium*, pp. 575–579 (2000)
22. Srinivasan, S., Rindfleisch, T.C., Hole, W.T., Aronson, A.R., Mork, J.G.: Finding UMLS Metathesaurus concepts in MEDLINE. In: *AMIA Symposium, 2002*, pp. 727–731 (2002)
23. Martens, A., Bernauer, J., Illmann, T., Seitz, A.: Docs ‘n Drugs – The Virtual Polyclinic: An Intelligent Tutoring System for Web-Based and Case-Oriented Training in Medicine. In: *AMIA Fall Symposium* (2001)