Modeling Human-Like Robustness in an Intelligent Tutoring System

Hameedullah Kazi[1], **Peter Haddawy**[1], and **Siriwan Suebnukarn**[2], Non-members

ABSTRACT

Intelligent tutoring systems are no different from other knowledge based systems in that they are often plagued by brittleness. Intelligent tutoring systems for problem solving are typically loaded with problem scenarios for which specific solutions are constructed. Solutions presented by students, are compared against these specific solutions, which often leads to a narrow scope of reasoning, where students are confined to reason towards a specific solution. Student solutions that are different from the specific solution entertained by the system are rejected as being incorrect, even though they may be acceptable or close to acceptable. This leads to brittleness in tutoring systems in evaluating student solutions and returning appropriate feedback. In this paper we discuss a few human-like attributes in the context of robustness that are desirable in knowledge based systems. We then present a model of reasoning through which a tutoring system for medical problem-based learning, can begin to exhibit human-like robust behavior in evaluating solutions in a broader context using UMLS, and respond with hints that are mindful of the partial correctness of the student solution.

Keywords: Robustness, Medical Problem-Based Learning, Intelligent Tutoring Systems, UMLS

1. INTRODUCTION

While traditional knowledge based systems often work well for narrowly defined tasks within specialized domains, they lack the meta-cognition and human-like common sense to deal with unforeseeable situations. Many systems suffer from brittleness and they are often unaware of their own limitations [1]. It is normal for a complex system to fail at some point, however what makes a system brittle is that it shows sudden failure beyond a certain point. Human beings also fail, however they are able to establish some self recovery before their failure leads to catastrophe [2]. Thus the failure humans exhibit is often soft and gradual rather than being hard and sudden.

Manuscript received on December 15, 2006; revised on March 20, 2007.

[1]The authors are with the Computer Science and Information Management Program, Asian Institute of Technology, Thailand; E-mail: hameedullah.kazi, haddawy@ait.ac.th

[2]The author is with the School of Dentistry, Thammasat University, Thailand; E-mail: ssiriwan@tu.ac.th

The need to emulate human-like behavior in intelligent systems has often led to an examination of how the human mind works. Minsky [3] describes a possible explanation of how in the event of damage to some parts of the brain, significant functionality is still maintained, by the delegation of tasks to other parts that have not suffered damage. In other words, the failure of some sub-systems leads to a task delegation to other subsystems, thereby resulting in some degree of robustness.

subsystems, thereby resulting in some degree of robustness. Sloman [4] has argued that the human mind employs a combination of rule based and heuristic methods for reasoning, where rule-based methods are characterized as systematic and logical set of laws, while heuristic methods are based on principles of association, similarity and contiguity. Some researchers have advocated the use of heuristic methods as a solution to the problem of brittleness in knowledge based systems. Accurate results may not be achievable where factual knowledge is found to be insufficient or the knowledge base is known to contain gaps, in which case heuristic methods can be employed to achieve partially correct, if not fully accurate results [5]. These heuristic methods should be able to exploit the knowledge structure of the knowledge based system to provide reasonable answers. In the next few sections we describe how the issues of gradual failure, self analysis of limitations, self recovery, task delegation and the use of multiple modes of reasoning in the context of robustness, can be applied to an intelligent tutoring system for medical problembased learning (PBL) using UMLS.

2. ROBUST OUTPUT QUALITY

A knowledge based system is designed to respond to input which has a specific format and is confined to a certain scope of knowledge. If the input happens to fall outside this scope, the output quality is expected to deteriorate. Groot, Teiji & Harmelen [6] describe how a quantitative analysis of the robustness of knowledge based systems can be achieved. They outline a few definitions of robustness, one of which is that the output quality of a knowledge based system should decrease monotonically with decrease in input quality. They mention that while this demand may be practically too strong, a system that exhibits somewhat monotonic output may be considered robust. They also argue that the rate of output quality

change in a robust knowledge based system, should be slow. A knowledge based system that is brittle, will exhibit abrupt degradation in its output quality as the input quality deteriorates beyond a certain point. However, a robust system will show a smooth degradation in its output quality as the input quality deteriorates beyond the edge of the system knowledge as shown in Figure 1.

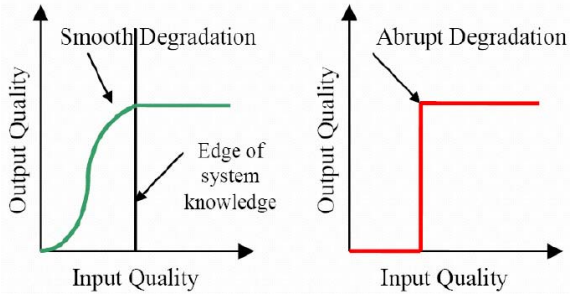

Fig.1: *Smooth vs. Abrupt Degradation*

3. REASONING SCOPE IN MEDICAL TUTORING SYSTEMS

Intelligent tutoring systems can be considered knowledge based systems whose problem solving activity is to evaluate student solutions to a posed problem and provide feedback to the students in the form of hints. The task of generating intelligent hints that are suited to the knowledge level of the student, is addressed in many tutoring systems [7-8] as part of student modeling. However the task of evaluating student solutions in a broad scope of reasoning is yet to be addressed in sufficient depth. Tutoring systems that offer some latitude in accepting differing solutions often confine students to a narrow scope of solution representation. Crowley & Medvedeva [9] accept a broad range of solutions for a given problem, but students are restricted to a local and customized ontology for choosing their solution concepts. Lulis, Michaels & Evens [10] emphasize the need for qualitative reasoning in tutoring systems and provide a mechanism through which students are able to present qualitative responses; however the response is only confined to assigning values to a small set of variables. The COMET system [11] provides an interface through which students can construct their hypothesis (solution) in the form of a directed acyclic graph. It evaluates a student hypothesis by comparing it against a specific expert solution. Nodes in the hypothesis that are not found in the expert solution are simply deflected and the system responds with the hint"¡Node¿ is beyond the scope of this problem".

The responses of such tutoring systems in unanticipated situations are quite contradictory to how a human tutor would normally respond. If the student response happens to fall outside the scope of the tutoring systems knowledge, the system responds with a premeditated hint that is often oblivious of the partial correctness of the student response. At the same time these tutoring systems are devoid of the metacognitive ability to assess their own capability in order to inform the student of the systems limitations or to attempt self recovery.

This motivates the need to have a medical tutoring system that offers students a broad scope of hypothesis representation and at the same time offers an assessment of the student hypothesis that describes the quality or degree of correctness. The tutoring system should be able to respond with certainty when the knowledge base is found to be sufficient. However when the knowledge base is not found to be sufficient, the system should be able to exploit its knowledge structure to achieve partial if not complete results. Thus the system should exhibit a gradual deterioration in quality when its knowledge limit is reached. Such a tutoring system should also have the ability to assess its own limitations and be able to inform the students about these limitations, which can help the students to reason accordingly.

4. ROBUSTNESS VIS--VIS TUTORING SYSTEMS

The proposed tutoring system is designed to cover PBL in the medical domain. A PBL session typically comprises of a group of 6-8 students, who are given a problem to solve within a period of about two hours. Based on the description of the problem scenario posed to the students, they are expected to form their solution in the form of a hypothesis graph, where graph nodes represent medical concepts and directed edges indicate the cause effect relationships between respective nodes.

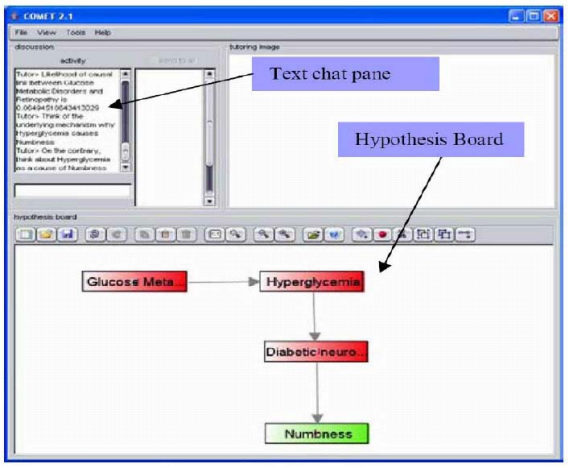

Fig.2: *System Prototype*

We have developed a system prototype using java. The problem representation in our system is the same as that in COMET [8] of a directed acyclic graph to describe a hypothesis. The hypothesis graph is based on the Illness Script [12], where nodes are en-

abling conditions, faults or consequences. The knowledge base of our system is formed by combining UMLS tables with an additional table that represents causal links between concepts. The system interface provides students with a workspace as a hypothesis board to form their hypothesis, along with a text chat pane through which the system returns feedback in the form of hints, as shown in Figure 2. For purposes of forming their hypothesis, students choose concepts from the broad and diverse UMLS Metathesaurus [16], as hypothesis nodes. For example, students are presented a problem scenario related to diabetes: A 45-year-old woman came to the clinic with following symptoms: tiredness, always thirsty, voided frequently with large amount of urine for 4-5 months. She voided approximately 10 times during the day and 4-5 times during the night. She was hungry quite often but lost 5 kgs body weight during the past 4 months. She also had numbness, leukorrhea and delayed wound healing.

A student hypothesizes that hyperglycemia is a cause of diabetic neuropathy which is shown to be a cause of numbness, in Figure 2.

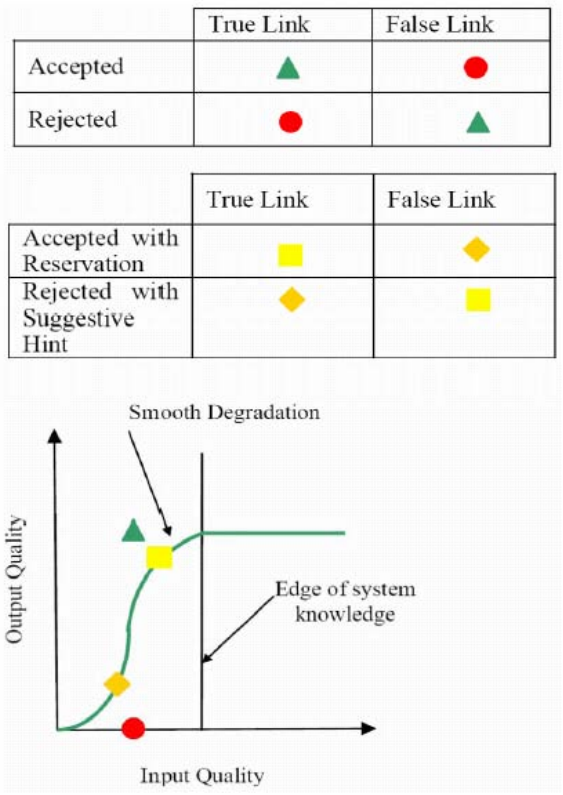

Fig.3: *Desired Degradation Curve*

Each hypothesis causal link drawn by the students needs to be evaluated against the knowledge base to determine whether the link drawn by the student should be accepted or rejected along with a hint to provide feedback. The output quality of an intelligent tutoring system is essentially comprised of two main components: evaluating student hypothesis and returning intelligent feedback as in the form of hints.

A causal link that is considered by a human tutor to be correct is henceforth referred to as a true link, whereas a causal link considered by a human tutor to be incorrect, is referred to as a false link. For all links that lie beyond the edge of the system knowledge, the output quality will be high if a true link is accepted or a false link is rejected by the system, as shown in Figure 3. However, if a false link is accepted without reservation or a true link is rejected without suggestive feedback that recognizes the partial correctness of the link, the output quality will be very low, as shown in Figure 3. Thus the output quality, without reservation or suggestive feedback in the hints, will be marked by fluctuating highs and lows. A system which produces fluctuating output quality as a result of deteriorating input quality is less predictable [6] and is considered less robust.

Therefore, for all hypothesis links that lie beyond the edge of the system knowledge, accepted links need to be supported with hints that show some form of reservation and suggest improvement to the causal link. Likewise rejected links need to be supported with suggestive hints that acknowledge partial correctness of the link or the closeness of the link to a true causal link, to result in somewhat smooth degradation as shown in Figure 3. The exact gradient of the curve shown in Figure 3, will be dictated by the nature of hints, as they vary from one situation to another.

5. THREE TIER MODEL FOR ROBUSTNESS

Robustness in our system is made possible through the use of a broad and widely available medical knowledge source such as the UMLS. The system design towards maintaining human-like robustness comprises of a three tier model, as shown in Figure 4. The tiers are successively applied in order of necessity. The first tier is a rule-based expert knowledge base, while the second tier is a heuristic method of computing semantic distance using knowledge structure within UMLS, whereas the third tier is based on a probabilistic Bayesian model.

The system makes it a matter of priority to first employ the rule-based tier which contains sure knowledge for reasoning purposes. If the first tier fails to deliver, the system employs the heuristic mechanism in the second tier. If the second tier fails too, the system uses the robust but not so accurate, third tier of probabilistic Bayesian model. Thus the system applies a step wise fallback approach of employing multiple modes of reasoning that are designed to provide self recovery and smooth degradation in output quality with deteriorating input quality.

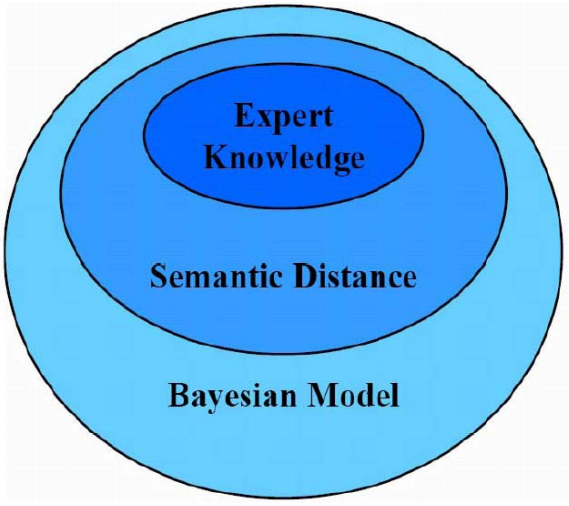

Fig.4: *Three Tiers of Robust Reasoning*

5.1 Rule-Based Expert Knowledge Base

This knowledge base is in the form of a database table that comprises of sure knowledge which contains causal links such as:

Diabetes → Hyperglycemia
Hypoinsulinism → Hyperglycemia
Glucose Metabolism Disorder → Hyperglycemia
Hyperglycemia → Diabetic Neuropathy
Diabetic Neuropathy → Numbness

This knowledge base is formed through the collation of causal links found in expert solutions to various problems, and the causal links found in student solutions that are certified by the domain experts to be correct.

While evaluating a causal link between two concepts in the student hypothesis, the system first attempts to find the respective link in this knowledge base, as an attempt to use rule based certain knowledge. If the system finds the hypothesis link in this knowledge base, the link is accepted, knowing that this comes from part of the systems rule-based certain knowledge. Additionally, the system also checks to see if an indirect link between the two concepts is found or if there is a reverse link that exists between the respective concepts. However, if the link is not found in this knowledge base, the system resorts to the heuristic method in an attempt to achieve a partial if not completely accurate assessment of the link under evaluation.

5.2 Heuristic Measure of Partial Correctness Using Semantic Distance

In this mode of reasoning, the system exploits the knowledge structure within UMLS to evaluate partial correctness of the causal link under evaluation, thereby attaining some degree of robustness. The node, from which the causal edge in the student hypothesis is emanating, is henceforth referred to as the source node, whereas the node, to which the causal edge is leading to, is referred to as the target node. The system checks if either the target node or source node is found in any of the acceptable solutions to the given problem. If the target node is found, the system measures the semantic distance between the source node and each of the nodes that are known to cause the target node. Thus the system measures the closeness of the source node to nodes that are known to cause the target node, thereby obtaining a measure of partial correctness of the hypothesis link under evaluation.

The semantic distance is measured by employing a modified version of the method described by Al-Mubaid & Nguyen [13]. Parent-child relationships from the UMLS Metathesaurus are used to construct the parental hierarchy of both nodes between which semantic distance is to be measured. An appropriate hint indicating the partial correctness or the closeness of the link to a plausible one is returned to the students.

If however, the target node is not found in the acceptable solutions, the system checks if the source node is found, in which case the comparison is made between the target node and each of the nodes that are known to be caused by the source node.

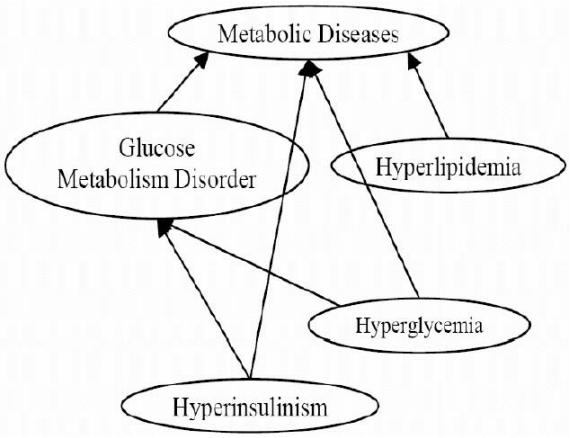

Fig.5: *Parental Trees of Two Concepts*

The semantic distance is only computable if the parental trees of both concepts, between which distance is to be measured, are actually connected. For illustration, a portion of the connected parental trees of concepts, hyperlipidemia and glucose metabolism disorder, is shown in Figure 5. If however, the parental trees from both concepts happen to be disjoint, semantic distance is not computable. In this situation, the system resorts to the method of estimating likelihood of the source node causing the target node through the Bayesian model.

5.3 Bayesian Model of Causal Links

Work done in extracting causal relationships between medical concepts in UMLS [14-15] inspires us to use the Bayesian Network shown in Figure 6. This Bayesian network is used to determine the likelihood of a causal relation between nodes representing concepts A and B. Causal Relation is a Boolean node, where a true value indicates causal relation between nodes A and B, while a false value indicates the lack of a causal link between the respective nodes. Semantic Type A is the semantic type of concept A as defined in UMLS, and Semantic Type B is the semantic type of concept B. Each concept in the UMLS Metathesaurus is categorized under at least one semantic type from a list of 135 semantic types in the UMLS semantic network [16]. Co-Occurrence Frequency gives the frequency with which the two concepts are known to have co-occurred in medline citations, and is extracted from the UMLS table mrcoc. Co-Relation Radius is the radius distance within which concept A is known to be related to concept B. Co- Relation Radius is assigned a value of zero if the concepts are found to be directly related in the UMLS Metathesaurus, one if there is one intermediate node between A and B, and two if the relation radius is greater than one or if the concepts are not related at all.

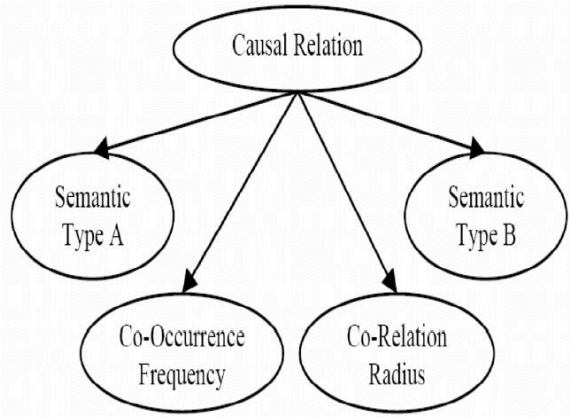

Fig.6: *Bayesian Network for Causal Relationship*

In order to estimate the likelihood of the causal link between two concepts A and B, the semantic types of both concepts, their co-occurrence frequency, and their relation radius is fed to the Bayesian network as evidence. The updated belief for true value of Causal Relation is examined to get the probability of causal relation between A and B. Based on the retrieved probability value, appropriate hints are returned to the student.

6. PEDAGOGICAL STRATEGY BASED ON STEPWISE FALLBACK

While evaluating hypothesis links, only those links that are found in the expert knowledge base are accepted without any kind of feedback, explanation, or reservation. Links, for which the semantic distance is found to be below a certain threshold, are accepted with reservation. All other links are rejected, and appropriate feedback is returned based on the reasoning tier that was applied.

For purposes of illustration, we present a few examples of how the three tiers are applied in a stepwise fallback fashion while evaluating hypothesis links and how the tutor responds with appropriate hints. Consider the problem scenario described earlier of a patient with diabetic symptoms. While solving the case, the student draws causal links between various concepts and receives corresponding feedback from the tutoring system.

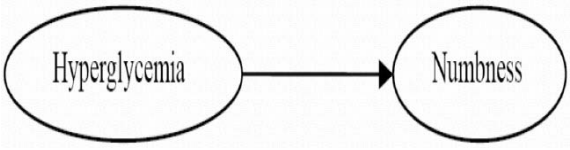

Fig.7: *Student Hypothesis Link*

For the hypothesis link in Figure 7, the system detects an indirect link, rejects this link and responds with the hint: "Think of the underlying mechanism why hyperglycemia causes numbness". However if the student tries to draw a link from numbness to hyperglycemia, the system detects a reverse link and responds with the hint: "On the contrary, think of hyperglycemia as a cause of numbness".

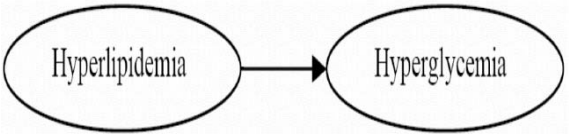

Fig.8: *Student Hypothesis Link*

For the hypothesis link in Figure 8, the system does not find a corresponding link in the knowledge base, so it checks the semantic distance, rejects the link, and responds with the hint:"Hyperlipidemia is fairly close to a known cause of hyperglycemia. Instead of hyperlipidemia, think more specifically about other metabolic diseases".

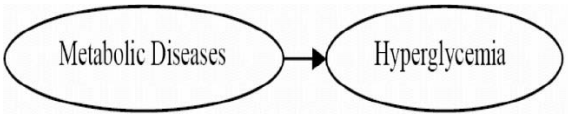

Fig.9: *Student Hypothesis Link*

For the hypothesis link in Figure 9, the system does not find a corresponding link in the knowledge base, so it checks the semantic distance, accepts the link with reservation and responds with the hint:"Metabolic diseases is very close to a known cause of hyperglycemia. Metabolic diseases may be

acceptable. However, think more specifically about kinds of metabolic diseases".

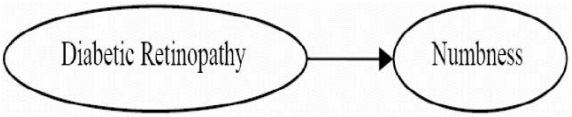

Fig.10: *Student Hypothesis Link*

For the hypothesis link in Figure 10, the system does not find the link in the knowledge base, and semantic distance is not computable. The system rejects the link and responds with the hint: "Diabetic retinopathy is not known to be a cause of numbness. Likelihood of causal relation between diabetic retinopathy and numbness is very low".

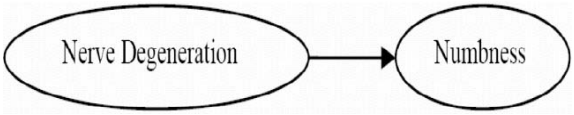

Fig.11: *Student Hypothesis Link*

For the hypothesis link in Figure 11, the system does not find the link in the knowledge base, and semantic distance is not computable, so the system rejects the link. The Bayesian probability is quite high, so the system responds with the hint: "There may be a causal relation between nerve degeneration and numbness".

As shown above, the hints inform the student about the closeness of the hypothesis link to a plausible link. If this information is not available, the system provides information about the likelihood of the causal link. At the same time, the language of the hints generated by the system, informs the student of the tutors reasoning limitations, which is likely to lead to improved reflective thinking and hence better learning.

7. INITIAL EVALUATION

The initial evaluation of our system was based on the agreement ratings of a collection of 15 causal links along with their respective hints, which were presented to an experienced human medical tutor at Thammasat University. The causal links comprised of five links each from three cases, for which we have already collected human expert solutions. The three cases are based on disorders such as diabetes, heart attack and pneumonia. On an agreement scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), the human tutor was asked to rate various hints for each causal link. The average score of hints based on our measure of partial correctness and causal likelihood was 4.13, whereas the average score of the hints without the partial correctness and causal likelihood feedback was 2.13.

8. CONCLUSIONS

In this paper we have described a multi tier approach in an intelligent tutoring system towards exhibiting human-like robust behavior in evaluating student hypotheses and responding in the form of hints. We have also discussed how the notion of gradual and smooth degradation in the output quality as a result of deteriorating input quality, applies to intelligent tutoring systems. Our approach towards incorporating robustness is innovative in employing a combination of rule-based, heuristic and probabilistic approaches applied successively in order of necessity, incorporating the notions of self recovery and task delegation. We have presented illustrative examples of how such human-like gradually deteriorating output quality can be observed in the responses of a medical tutoring system for PBL.

The initial assessment of our approach and feedback from human domain experts seems to indicate that the proposed methods can be effective in helping medical students acquire clinical reasoning skills. We have started to collect samples of student hypotheses for three different problem scenarios covering diseases and disorders such as diabetes, heart attack and pneumonia. We intend to conduct sub- system evaluations of the method of computing semantic distance and the method of estimating likelihood of a causal link between two concepts using the Bayesian model. Finally, we plan to measure the effectiveness of our generated hints compared with human tutors and perform quantitative evaluations of the pedagogical strategy incorporated in our system.

References

[1] McCarthy, J. (1984). Some expert systems need common sense. In H. Pagels (ed.), Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer. Annals of the New York Academy of Sciences, Vol. 426, 129-137.

[2] Nielsen, P., Beard, J., Kiessel, J., & Beisaw, J. (2002). Robustness in Modeling Behavior Overview. Proceedings of 11th CGF-BR Conference, May, 2002.

[3] Minsky, M. (1986). The Society of Mind. New York: Simon & Schuster.

[4] Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22

[5] Paritosh, P. K. (2006). The Heuristic Reasoning Manifesto. Proceedings of QR06, Hanover, New Hampshire, July 10-12, 2006.

[6] Groot, P., Teije, A. T., & Harmelen, F. V. (2005). A Quantitative Analysis of the Robustness of Knowledge- Based Systems Through Degradation Studies. Knowledge and Information Systems, 7 (2), 224-245.

[7] Kabassi, K., Virvou, M., & Tsihrintzis, G. A. (2006). Requirements Capture for a Personalized

Medical Tutor. Proceedings of International Special Topic Conference on Information Technology in Biomedicine, Ioannina, Greece, October 26-28, 2006.

[8] Suebnukarn, S., & Haddawy, P. (2006). Modeling individual and collaborative problem-solving in medical problem-based learning. User Modeling and User Adapted Interaction, 16 (3), 211-248.

[9] Crowley, R., & Medvedeva, O. (2006). An Intelligent Tutoring System for Visual Classification Problem Solving. Artificial Intelligence in Medicine, 2006, 36 (1), 85-117.

[10] Lulis, E., Michael, J., & Evens, M. (2004). Using Qualitative Reasoning in the Classroom and in Electronic Teaching Systems. Proceedings of the 18th International Workshop on Qualitative Reasoning, Northwestern University, Evanston, IL August.

[11] Suebnukarn, S., & Haddawy, P. (2006). A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system. Artificial Intelligence in Medicine, 2006, 38, 5-24.

[12] Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt and M. L. De Volder (Eds.) Tutorials in problem-based learning: A new direction in teaching the health professions. The Netherlands: Van Gorcum.

[13] Al-Mubaid, H., & Nguyen, H. A. (2006). A Cluster Based Approach for Semantic Similarity in the Biomedical Domain. Proceedings of the 28th IEEE EMBS Annual International Conference, New York, USA, Aug 30-Sept. 3, 2006.

[14] Burgun, A., & Bodenreider, O. (2001). Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. MedInfo, 2001, 10(Pt 1), 171-175.

[15] Mendonca, E. A., & Cimino, J. J. (2000). Automated Knowledge Extraction from MEDLINE Citations. Proceedings of AMIA 2000 Fall Symposium, 575-579.

[16] U.S. National Library of Medicine, (2007). 2007AA Introduction. Retrieved April 19, 2007 from the World Wide Web: http://www.nlm.nih.gov/research/umls/umlsdoc_intro.html

Hameedullah Kazi is a doctoral student in the department of Computer Science and Information Management at Asian Institute of Technology, Thailand. He obtained his MS in Electrical Engineering from Purdue University, USA in 2000. He is interested in pursuing research in the areas of Intelligent Tutoring Systems, Medical Informatics, and Knowledge Representation & Reasoning.

Peter Haddawy received his PhD in Computer Science in 1991 from the University of Illinois- Urbana. He was Associate Professor in the Department of EE&CS at the University of Wisconsin-Milwaukee and Director of the Decision Systems and Artificial Intelligence Lab there through 2002. He is presently Professor in the CSIM Program, Vice President for Academic Affairs at the Asian Institute of Technology in Bangkok, Thailand and Senior Scientist at the University of Wisconsin-Milwaukee. He has been an Avery Brundage Scholar and a Fulbright Fellow. His current research interests include decision-theoretic problem solving, probabilistic reasoning, preference elicitation, and applications in ecommerce and medicine.

Siriwan Suebnukarn, DDS, PhD is Assistant Professor at Thammasat University Dental School, and adjunct faculty of Computer Science and Information Management at Asian Institute of Technology. She received her PhD in Computer Science and Information Management in 2005 from School of Engineering and Technology, Asian Institute of Technology. She obtained her DDS from Prince of Songkhla University in 1992, her Graduate Diploma in Clinical Science (Endodontics) from Chulalongkorn University in 1995. Her interests lie in the areas of intelligent medical training systems, medical and dental informatics, clinical skill training, and problem-based learning.