Towards Automatic Conceptual Personalization Tools

Faisal Ahmad, Sebastian de la Chica, Kirsten Butcher[2], Tamara Sumner, James H. Martin

Institute of Cognitive Science
Department of Computer Science
University of Colorado, Boulder, CO 80309

Learning Research and Development Center[2]
University of Pittsburgh, Pittsburgh, PA 15260

faisal.ahmad, sebastian.delachica, tamara.sumner, james.martin@colorado.edu
kbutcher@pitt.edu[2]

ABSTRACT

This paper describes the results of a study designed to validate the use of domain competency models to diagnose student scientific misconceptions and to generate personalized instruction plans using digital libraries. Digital library resources provided the content base for human experts to construct a domain competency model for earthquakes and plate tectonics encoded as a knowledge map. The experts then assessed student essays using comparisons against the constructed domain competency model and prepared personalized instruction plans using the competency model and digital library resources. The results from this study indicate that domain competency models generated from select digital library resources may provide the desired degree of content coverage to support both automated diagnosis and personalized instruction in the context of nationally-recognized science learning goals. These findings serve to inform the design of personalized instruction tools for digital libraries.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems; H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *User issues*; K.3.1. [**Computers and Education**]: Computer Uses in Education – *Computer-aided instruction*.

General Terms

Design, Experimentation, Human Factors

Keywords

Personalization, digital libraries, competency models, knowledge maps, student misconceptions, instructional plans

1. INTRODUCTION

Over the past two decades, cognitive research has examined the role of background knowledge, individual differences, and preferred learning styles in influencing learning outcomes [4]. A key finding is that every student brings preconceptions about how the world works to every learning situation, and that these initial understandings need to be explicitly targeted as part of the instructional process [ibid]. Simultaneously, there have been major demographic shifts taking place in learning populations, with many classrooms containing learners from diverse cultural backgrounds and prior experiences [28]. Educators increasingly need support to customize educational content and activities to meet the needs of a heterogeneous student population [11]. To address these needs, we are investigating how personalization tools capable of assessing and responding to current student conceptions and misconceptions can be embedded in educational digital libraries.

The personalization tools we are creating for digital libraries are similar in intent to prior work in adaptive learning environments, such as AutoTutor [8] and the Practical Algebra Tutor [12], but differ in two key ways. First, rather than relying on human-intensive knowledge engineering efforts to create models of student competencies within a given domain, we are examining how natural language processing techniques may be used to automatically construct learner competency models by summarizing existing digital library resources. Our hypothesis is that a carefully selected set of high-quality educational resources features the necessary breadth and depth of coverage to serve as the basis for the automatic development of pedagogically sound and age-appropriate domain competency models. Second, rather than relying on predefined content and curriculum specifically authored for a particular adaptive learning environment, we are developing personalization tools to dynamically select digital library resources that encompass a variety of instructional strategies, such as background readings, simulations, and other learning activities.

To inform the design and development of the envisioned conceptual personalization tools, and the underlying natural language processing algorithms, we have conducted a multi-part 10 month study to examine the processes used by human experts to: (1) construct domain competency models from digital library resources and (2) develop personalized instructional strategies based on the competency models. This study examined in detail how experts identify and represent key concepts that students should know by analyzing and summarizing digital library resources. We specifically asked experts to develop an age-appropriate domain competency model for high school students studying earthquakes and plate tectonics using materials from the Digital Library for Earth System Education (DLESE.org). DLESE provides access to high-quality collections of educational resources for the Earth sciences and services to help educators and learners effectively find, create, use, and share such resources. The topics chosen are recognized as important content areas in national science educational standards [2, 17]. In this article, we describe the study methodology, the results, and implications for the design of automatic conceptual personalization tools. In particular, we focus on discussing four key questions:

JCDL'07, June 18–23, 2007, Vancouver, BC, Canada.

- Can domain concepts important for the development of student competencies be reliably and consistently identified in educational digital library resources? If human experts can do so consistently, it is likely that computational approaches can be developed to automate such processes.
- Do the domain concepts embodied in a set of digital library resources provide sufficient coverage of important learning goals for the target student population?
- Is the domain competency model generated from digital library resources useful for diagnosing student misconceptions and understandings?
- How well does the domain competency model support developing personalized instructional strategies using digital library resources?

In the remainder of this paper we first review related work in the areas of adaptive learning environments, knowledge maps and digital library information extraction. We then describe the methodology and results from our study. Finally we discuss the implications for the design of automatic algorithms to construct domain competency models that support personalized instruction.

2. RELATED WORK

Developing approaches for tailoring instruction to students' current understanding, in both face-to-face classroom settings and computer-based learning environments, has been a long-term focus of learning science research [20]. For instance, conversational learning theory describes how personalization takes place in a face-to-face classroom setting. According to this theory, understanding of a topic occurs as a result of a structured and iterative conversation between an instructor and a learner within a *conversational domain* [20, 25]. Learner knowledge is constructed through an iterative process where students communicate their current understanding – through discussion, writings, or other scholarly artifacts – and the instructor develops an instructional strategy appropriate to both the learner's current understanding and the instructor's knowledge of what the learner *should* know about the topic. Together, these two forms of knowledge about the topic under study comprise the conversational domain. Pask developed this theory as part of one of the first efforts to create an adaptive learning environment in the 1970s; he provided an early illustration of how a computer-based knowledge representation of the conversational domain could be used to support personalized instruction [20].

More recently, adaptive learning environments have tried to emulate these types of student-teacher interactions by developing rich and detailed symbolic computer models to drive interactions with the student. These models typically depict the key ideas or concepts that learners should understand, how these ideas are interconnected, how these ideas change over time, and specific curriculum components to support learning of selected concepts [1, 2, 26]. Two prominent adaptive learning environments – the Practical Algebra Tutor and AutoTutor – have both reported impressive learning outcomes in controlled studies [8, 12].

The Practical Algebra Tutor supports students studying high-school algebra [23]. It uses a detailed cognitive model of desirable student competencies in algebra, represented as production rules, and a detailed model of the curriculum, represented as algebra problems that students should be able to solve, to automatically select problems to present to students. The Practical Algebra Tutor has been widely deployed in real classroom settings. To deploy this system in new settings and school districts, the researchers report that the content, i.e., the choices of problems presented to students, needed to be continually expanded and updated [ibid]. The costs and human effort associated with developing and updating the necessary instructional content is so high that the researchers have developed authoring tools to assist with these human-intensive knowledge engineering efforts [ibid].

The AutoTutor system developed by Graesser et al uses an underlying model of what students should know about qualitative physics, represented as curriculum scripts, to personalize tutorial conversations for the topic of undergraduate introductory physics [8]. These curriculum scripts, created by subject matter experts, explicitly model expected student domain competencies, ideal answers, corresponding physics problems, questions to elicit further student knowledge, and common student misconceptions. Once again, the researchers found that the intensive human effort involved in creating these scripts raises significant barriers to applying this approach to other domains; similarly to the Practical Algebra Tutor work, the researchers are also developing tools to facilitate the construction of these scripts [27].

These two examples illustrate the potential successes (positive learning gains) and the challenges (human-intensive labor needed for model and curriculum development) to be faced when developing systems to support conceptual personalization in learning interactions.

Prior research has demonstrated both the utility and production costs of using symbolic knowledge models for diagnosing student understanding and for generating personalized instruction strategies. Advances in statistical methods have prompted other researchers to explore the use of fully automatic techniques such as Latent Semantic Analysis [13] in adaptive learning environments. For instance, the Summary Street application assesses student writing by comparing the 'bag of words' derived from a student essay, represented as a vector, with those from a prepared corpus of materials about the subject domain to produce a cosine value characterizing the degree of alignment [29]. While adept at detecting the existence of discrepancies and problems in student essays, this vector-based knowledge representation does not readily support identifying the contents of learners' specific misunderstandings, nor does it support generating a specific instructional strategy.

Our approach to supporting conceptual personalization is aimed at balancing 'the best of both worlds': using statistical natural language processing approaches to automatically generate domain competency models, in the form of a semi-structured knowledge representation called a "knowledge map", and in turn, using these knowledge maps to underpin personalized instructional strategies. We believe that knowledge maps provide a rich enough representation to support our computational needs, and are also useful representations for learners to see and use. To illustrate their utility, consider a scenario where Heather, a 12th grade science student, has been assigned the task of writing an online essay on the causes of earthquakes using the envisioned personalization tools. The personalization tools have previously processed digital library resources from DLESE to construct a domain competency model depicting desired high-school level understandings of earthquakes and plate tectonics (see Figure 1.)

Heather writes that earthquakes can occur all over the world and requests feedback from the personalization tools. The tools analyze and detect critical differences between Heather's essay and nodes 1 and 2 in the domain competency model. To address this misconception, the personalization tools select age-appropriate resources from DLESE about the distribution of earthquakes and their prevalence along plate boundaries. These resources are presented to Heather along with the relevant portion of this knowledge map. The knowledge map provides Heather with an overarching conceptual guide, highlighting the core concepts Heather needs to work on and how they are related. This map also helps her understand why these resources were selected and how they can help with her specific learning needs. This personalized response prompts Heather to reflect on the inaccuracy of her current conception. Heather remembers that there are more earthquakes in California (where her grandmother lives) than in Colorado (where she lives). Heather explores this difference using a DLESE resource – a simulation illustrating the relationship between plate boundaries and earthquakes – suggested by the personalization tools.

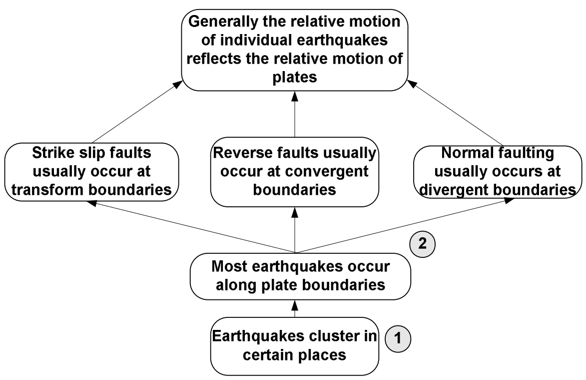

Figure 1 - Partial domain knowledge map [7]

As shown in Figure 1, knowledge maps are a specialized type of concept map. Concept maps have been shown to be reliable representations of learner understanding and flexible models to track and assess cognitive development [18]. Concept maps are hierarchical node-link diagrams that depict concepts, usually as nodes with one or two keywords, and their interrelationships, either as labeled or unlabeled links. Knowledge maps differ from concept maps by depicting knowledge using a network layout, by using richly descriptive statements in the nodes to capture robust concepts and ideas related to a domain, and by focusing on a limited number of link types. Because of these differences, knowledge maps tend to be more concise and more useful as sharable, human-readable representations than concept maps. Prior research indicates that knowledge maps are useful cognitive scaffolds, helping users lacking domain expertise – such as learners, new teachers, or educators teaching out of area – to understand the macro-level structure of an information space [9, 19].

Supporting concept map creation and their automatic analysis is an active research area in the digital library community [14, 15]. Recent research shows promise for the development of algorithms for automatically performing node and link element matching in order to assess student-produced concept maps computationally [14]. Marshall et al developed algorithms to compare student-produced maps to a 'gold-standard' expert-produced map to characterize the degree of alignment with a numerical score. We are extending this research to consider whether automatic comparisons of knowledge maps can provide learners with more specific feedback about conceptual differences, such as the interactions between Heather and the personalization tools in the above scenario.

The knowledge maps we are trying to create are, in effect, a specialized type of multi-document summarization. As such, generating them will require addressing issues in information extraction and library resource summarization. Within the digital library community, there are several research efforts that inform this work. For instance, Liddy [30] uses an assortment of natural language processing techniques to extract metadata elements by analyzing educational digital library resources, including the generation of very short resource summaries to populate the brief description field in the metadata. Fox et al [10] also demonstrated the potential value of these techniques for information extraction and metadata generation. While these approaches have met with promising results, these techniques do not directly address the issue of how to identify and extract key domain concepts from digital library resources, nor how to produce summaries of more than one resource.

McKeown et al [16] directly investigated how multi-document summarization techniques could support personalized digital library interactions. This research considered how user models (in the form of patient medical records) could be used to select and re-rank the presentation of search results. Additionally, they developed algorithms for selecting and organizing important passages from retrieved documents to create a personalized summary of the search results for doctors. As part of a related but more generic research effort in multi-document summarization, Radev et al [22], have developed the MEAD toolkit to support the development of summarization applications. We are directly building upon this toolkit in our research and extending it to support the summarization of educational resources, which differ in structure and content from the news articles that the MEAD team focused on supporting.

3. METHODOLOGY

Our research efforts in the area of personalization in digital libraries aim to determine how domain competency models computationally constructed from digital library resources may support the automatic diagnosis of student misconceptions and the generation of personalized instruction plans.

We have conducted a study to elicit human expertise for constructing and utilizing a domain competency model for personalized instruction purposes and to identify design requirements for the automation of these pedagogical processes. This human-centered approach follows in the tradition of prior research efforts where data collected in human studies serves as the basis for formulating design requirements [3, 6].

To foster the construction of a scientifically accurate and pedagogically useful domain competency model, we recruited two geology experts and two instructional design experts. Our geology experts had a Master's level education in geology and at least 5 years of field experience. Our instructional designers had at least 10 years of experience in curriculum and learning materials design. These four experts collaborated with the research team on this study over a period of 10 months for a total of approximately 80 hours per expert. The study consisted of two phases: competency model construction and competency model assessment.

3.1 Competency Model Construction

The first phase of the study involved the experts creating a domain competency model. This phase included resource selection and knowledge map construction processes.

3.1.1 Resource Selection Process

The purpose of the resource selection process was to identify a suite of digital library resources suitable for constructing a comprehensive domain competency model featuring the desired level of science content accuracy and coverage.

We instructed the experts to select appropriate learning resources from DLESE using three guiding principles. (1) To encourage the selection of resources that provided adequate breadth of coverage on the topic and to avoid excessive specialization, we instructed the experts to focus on resources providing comprehensive coverage for high school age learners in earthquakes and plate tectonics. (2) Given that the National Science Education Standards highlight the importance of learners understanding the necessary scientific terminology [17], we instructed the experts to focus on resources using age-appropriate domain terminology. (3) To generate data suitable for informing the design of natural language processing tools, we instructed the experts to focus on resources consisting mainly of expository text. The experts independently selected and ranked the 10 optimal DLESE resources given the resource selection guidelines and the targeted topic and age group. To achieve adequate balance between domain and pedagogical coverage, we paired each domain expert with an instructional designer to jointly rank their respective resource choices. All four experts then collaboratively reviewed and ranked all the individually selected resources and agreed on 20 resources in a discussion facilitated by research team members. At the end of this process, the research team collected the list of the 20 resources selected by the experts.

3.1.2 Knowledge Map Construction Process

The purpose of this process was to construct a comprehensive domain competency model encoded as a knowledge map using the 20 digital library resources selected by the experts.

3.1.2.1 Individual Knowledge Map Construction

To establish the consistency of the construction process and hence the feasibility of the planned natural language processing approaches, the experts first created knowledge maps for each of the learning resources. Each expert used CmapTools [5], a knowledge modeling software, to individually create knowledge maps for 11 of the 20 resources chosen in the resource selection process. To ensure that the experts would faithfully and reliably reflect the contents of the digital library resource in their knowledge maps, we instructed the experts to use the nodes in the knowledge map to capture the concepts as presented in each digital library resource as a paragraph, a sentence, a clause or one or more words. In addition, the experts were also instructed to use the knowledge map links to reflect the relationships between the concepts as expressed in the digital library resource. To facilitate relationship tagging, we provided our experts with a preliminary list of relationship categories and examples presented in Table 1. We also encouraged our experts to introduce any relationship terms they deemed necessary to best capture the contents of the resource. Each digital library resource was mapped by at least two experts to enable the research team to analyze the degree of knowledge mapping consistency across experts. At the completion of this activity, the research team collected the knowledge maps individually created by the experts.

Table 1 - Relationship types adapted from GetSmart [15]

Relationship Type	Examples
Causes	cause-effect, cause, result, consequence
Compares	comparison, analogy, is similar to, contrasts to
Elaborates	elaboration-additional, elaboration-general-specific, elaboration-part-whole/consists-of, elaboration-process-step, elaboration-object-attribute, elaboration-set-member, example, definition
Evaluates	evaluation, interpretation, conclusion, comment
Followed by	sequence
Is a	is-a-kind-of
Supports / Is evidence for	supports, is-evidence-for

3.1.2.2 Knowledge Map Integration

To ensure the desired breadth and depth of coverage, the experts integrated their knowledge maps into a single domain competency model encoded as a knowledge map in two steps.

First, to facilitate the creation of the domain competency model, each expert used CmapTools to merge the 11 knowledge maps s/he created into a single, merged knowledge map. This process resulted in four individually merged knowledge maps (one per expert). To better inform the automation of this process, we instructed the experts to constrain their knowledge map merging activities to the actual contents of the resources using three guiding principles. (1) The merged knowledge map should be representative of the digital library resources from which it was synthesized. (2) The merged knowledge map should contain key domain concepts and relationships from the resources. (3) If the underlying resources contained conflicting concepts or propositions, such inconsistencies needed to be captured accurately.

Second, we conducted a one-day collaborative workshop where the experts integrated the four individually merged knowledge maps into a single knowledge map. During this workshop, the experts also validated the contents of the final domain competency model for accuracy, completeness and fidelity to the original resources. At the workshop, the research team provided access to large printouts of the four individually merged knowledge maps as well as to the electronic versions of the individually merged knowledge maps and the emerging domain competency model. During the workshop, research team members facilitated the discussion and construction of the domain knowledge map using CmapTools to capture the final outcome. At the completion of the workshop, the experts had produced a draft of the domain competency model encoded as a knowledge map. This draft was circulated via email for a final round of individual

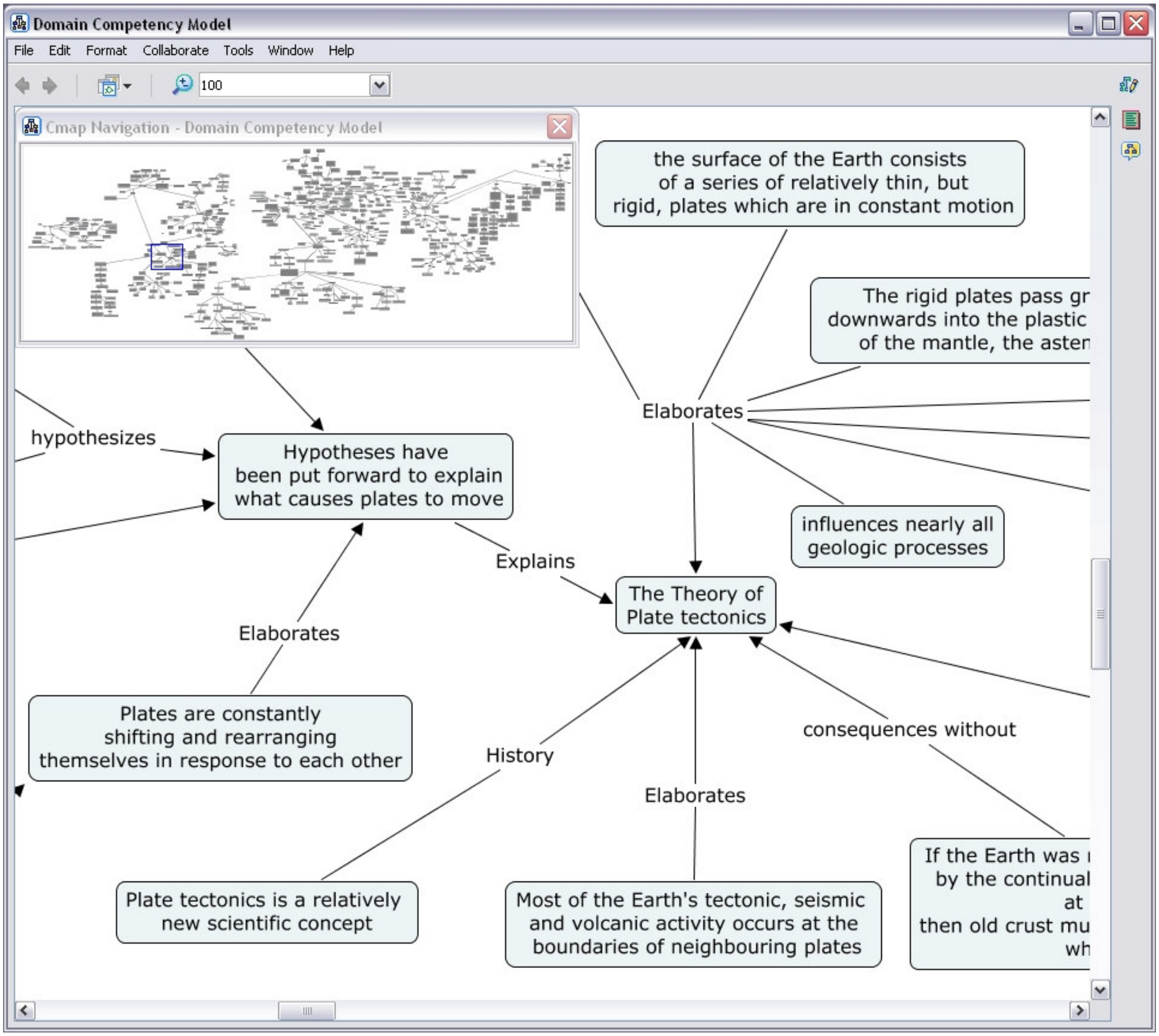

Figure 2 - Domain competency model for earthquakes and plate tectonics

reviews and to reach final consensus on its contents. The research team collected the final version of the domain knowledge map at the completion of this offline review. Figure 2 shows a portion of the final domain competency model, including an inset showing the full extent of the model.

3.2 Competency Model Assessment

The second phase of the study involved the experts assessing aspects of the domain competency model crucial to our guiding research questions. This phase included construction process reliability, content coverage, and pedagogical utility assessments.

3.2.1 Reliability

The purpose of the reliability assessment was to establish whether the experts represented the same content from the original digital library resources in their knowledge maps. Establishing consistency in the knowledge mapping process would provide an indication of the feasibility of the proposed automated domain competency model construction approach.

Two members of the research team served as annotators and independently created a hierarchical outline of the contents of four randomly selected digital library resources. The annotators then aligned their outlines to the concepts used by the experts in the corresponding knowledge maps. Table 2 shows the hierarchical outline created by one of the annotators for a DLESE resource and its alignment to the knowledge map for that same resource. All the hierarchical content outlines created by the annotators depicted the resources using two or more levels. We leveraged this commonality to drive our reliability analysis. To this end, we created clusters consisting of knowledge map nodes aligned to the first and second levels in the outline. In addition, we rolled up nodes aligned to third and lower levels of the outline to the clusters at the second level of the outline. For example, Table 2 shows the outline levels below *4.3 Convergent* aligned to the relevant nodes from the knowledge map as a single cluster. We collected the resource outlines and their alignments to the knowledge map nodes created by the research team annotators.

Table 2 - Resource outline created by research team annotator

Annotator Resource Outline	Knowledge Map Nodes
1. Earthquakes and volcanoes are concentrated on plate boundaries	8, 9
2. What are the driving forces for plate movement?	4, 5, 6
3. Geologic activities away from plate boundaries	1, 2, 3
4. Three major ways that plates interact along boundaries	7
4.1 Divergent	10, 13
4.2 Transform	12, 15
4.3 Convergent	11, 14
4.3.1 Oceanic crust meets continental crust	16, 19
4.3.2 Oceanic crust meets oceanic crust	17, 20
4.3.3 Continental crust meets continental crust	18, 21

3.2.2 Coverage

The purpose of the coverage assessment was to determine whether the domain competency model provided the desired breadth and depth of content coverage for personalized instruction purposes.

To establish alignment with widely accepted educational standards, one of our domain experts completed a learning goals alignment activity. We asked the domain expert to rate how well the generated domain competency model provided coverage for nationally-recognized learning goals from the *Benchmarks for Scientific Literacy* [2], published by the American Association for the Advancement of Science (AAAS). Using a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree, the domain expert rated whether the relevant *AAAS Benchmarks* were represented in the domain competency model. To obtain a consistent characterization of the rest of the topics covered in the domain competency model, one of our instructional experts completed a competency model topic grouping activity. We asked the instructional expert to identify and label groups of related concepts in the domain competency model. We collected the benchmark alignment to the domain competency model created by the domain expert and the topical groups identified by the instructional designer in the domain competency model.

3.2.3 Pedagogical Utility

The purpose of the pedagogical utility assessment was to gauge the adequacy and usefulness of the domain competency model for identifying student misconceptions and for generating personalized instruction plans. This assessment also aimed to provide an indication of the feasibility of the envisioned automated personalization tools.

To capture typical student conceptions, we asked 23 freshman students from the University of Colorado psychology subject pool to write essays on earthquakes and plate tectonics. These essays were processed by two of our experts to create knowledge map representations of their contents. To generate data suitable to guide the design of computational tools, we instructed the experts to only use comparisons between the domain competency model and the student essay knowledge maps. This guideline was intended to preclude the experts from using knowledge external to the domain competency model to identify misconceptions or to generate the appropriate instruction plan. Each expert completed formative assessments for six student knowledge maps, and each student knowledge map was assessed by at least two experts to generate enough data to establish the consistency of the assessment process and the potential for successful automation. The experts reported the results of their assessments using a student assessment template provided by the research team. These assessment reports included information about student concepts, competency model concepts, misconception description, misconception type and instructional plan. Experts provided this information for each identified misconception. The *student concepts* contained the nodes from the student essay knowledge map that constituted a misconception. The *competency model concepts* contained the nodes from the domain competency model illustrating the correct scientific conception. The *misconception description* and *misconception type* described the misconception and provided a label for the misconception. As a reference, the research team provided descriptions of seven misconception types, including definitions and examples (see Table 3).

Table 3: Misconception types

Misconception Category	Definition
Extraneous-Differentiation	A concept split into more then one concept or used in incompatible contexts due to inadequate knowledge about the characteristics of the concept
Form–Misconception	Ascription of incorrect form to objects and materials
Fragmented–Knowledge	Disconnected concepts that should have been connected
Inadequate–Differentiation	Two related concepts are confused due to inadequate knowledge about the unique characteristics of the concepts
Incomplete–Understanding	Conceptual understanding that is lacking some aspects or attributes of concepts
Overgeneralization	Concepts, theories or principles are applied to situations or cases where they cannot be used
Spatial–Misunderstanding	Incorrect geospatial location of objects or processes

The *instructional plan* described the strategy considered suitable for transforming the student misconception into a scientifically acceptable understanding, including relevant digital library resources. We conducted think-aloud sessions with each one of the experts completing an assessment and filling out an assessment report. These sessions were audio and video taped for subsequent analysis. In addition, we also collected all the assessment reports created by the experts.

4. RESULTS

Analysis of the data collected during our study has yielded important results related to the construction and assessment of a domain competency model from digital library resources.

4.1 Competency Model Construction

This study has produced a domain competency model encoded as a knowledge map on earthquakes and plate tectonics for high school age learners. This knowledge map provides a rich characterization of key domain concepts for learning about this topic. At a fine level of analysis, the experts represented the digital library resources by summarizing their contents using 8,582 words, approximately 5% of the resources textual content length. This knowledge map includes 564 nodes containing key domain concepts, and 578 relationships representing 105 unique relationship types. Experts' use of relationship types approximates a Zipf-like distribution with the top 10 relationship types accounting for 64% of all relationships in the domain competency model. Figure 3 depicts this distribution.

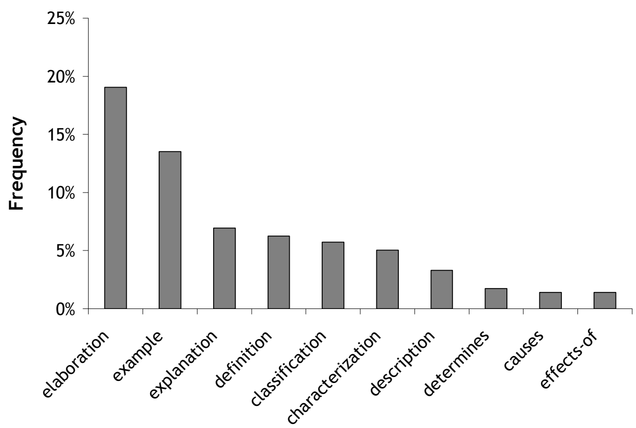

Figure 3 – Domain competency model: Top relationship types

This distribution suggests that it may be possible to focus our algorithm design efforts on a small set of relationship types to address most relationship tagging strategies typically used by human experts during knowledge mapping.

4.2 Competency Model Assessment

4.2.1 Reliability

To establish the reliability of the expert knowledge mapping process, we conducted an inter-rater reliability analysis. Examining how clusters of nodes in the knowledge maps created by the experts aligned to conceptual outlines of the original resources created by the research team annotators, we computed Kappa to assess if the experts had represented similar content from the resources in their knowledge maps. The average Kappa score was 0.74, indicating that human experts in this study reliably represented very similar concepts from the digital library resources for representation in their knowledge maps. This high level of agreement across experts indicates that the selected digital library resources do contain important domain concepts that can be targeted for automated summarization.

4.2.2 Coverage

To assess the domain coverage provided by the domain competency model, one of our domain experts completed a learning goal alignment activity to determine whether the relevant *AAAS Benchmarks* were covered by the domain competency model. Using a Likert scale, the expert *agreed* that three of the four relevant benchmarks were covered and *strongly agreed* that the fourth relevant benchmark was covered by the domain competency model. Furthermore, our domain expert aligned 82 out of 564 nodes from the domain competency model to the relevant *AAAS Benchmarks*. This alignment shows that 15% of the domain competency model suffices to provide adequate coverage of the relevant *AAAS Benchmarks*. This analysis of the content coverage provided by the domain competency model supports our hypothesis that a carefully selected suite of digital library resources may serve as the basis for the automatic construction of a pedagogically sound and age-appropriate domain competency model. The fact that the experts were able to address the relevant *AAAS Benchmarks* with a small percentage of the nodes in the domain competency model is not surprising, as the benchmarks concentrate on common core concepts and do not represent an exhaustive listing of all the learning goals that belong in the K-12 curriculum [2]. Preliminary analysis of the topical groups identified by the instructional expert in the competency model topic grouping activity indicates that the remainder of the nodes in the model provides very detailed elaborations of the learning goals, e.g., earthquake hazards and earthquake prediction. Therefore, the domain competency model provides not just adequate coverage of relevant science learning goals, but rather detailed pedagogical content useful for learning technology design and implementation.

4.2.3 Pedagogical Utility

To determine the usefulness of the domain competency model for educational assessment purposes, we analyzed the data collected during the pedagogical utility portion of the study.

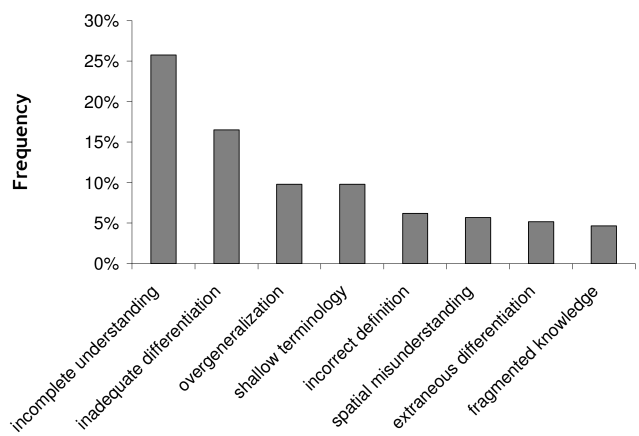

Figure 4 - Top misconception types

For the 12 student knowledge maps created from essays, 24 expert formative assessments were completed and a total of 128 misconceptions were identified. The 128 identified misconceptions were classified into 23 misconception types, out of which 8 misconception types accounted for 83% of the reported misconceptions (see Figure 4 for details). Although the experts were only provided 7 misconception types at the onset of the study, the richness of the domain competency model allowed them to expand the misconception types more than threefold. This expansion shows that a rich variety of misconceptions can be identified using the domain competency model to accommodate the pedagogical realities observed in the student essays. Similarly, a total of 26 relationship types were used by the experts to identify student misconceptions, out of which 8 relationship types participated in identifying 82% of student misconceptions (see Figure 5 for details). This result follows a Zipf-like pattern very similar to the one observed in the reported misconception types.

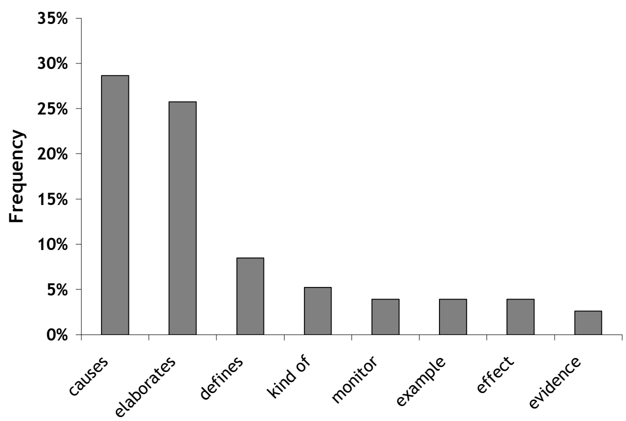

Figure 5 - Relationship types used for assessment

These results illustrate the conceptual completeness and expressiveness of the domain competency model and have three important implications: (1) they show that a majority of misconceptions across students belong to a small set of misconception types, (2) they provide a focus for developing computational components to automate the misconception identification process, and (3) they suggest that by focusing on a small set of relationship types we may be able to automate the identification of most student misconceptions.

Moreover, the inter-rater reliability (Kappa = 0.59) of the misconception identification activity showed that experts can reliably identify student misconceptions by comparing student essay knowledge maps to the domain competency model. This result is slightly lower than the Kappa value observed during competency model construction due to the more open-ended nature of the misconception identification activity; however, this level of agreement still indicates promise for the successful development of computational components to support this activity using the domain competency model.

To further analyze the pedagogical utility of the domain competency model we have defined two probabilistic measures of coverage based on the assessment reports. First, the *student concept coverage* represents the probability that a concept in the student essay knowledge map will be present in the domain competency model. Second, the *misconception identification coverage* represents the probability that a misconception identified in the student essay knowledge map will correspond to a portion of the domain competency model illustrating the correct scientific conception. The adequacy and utility of the competency model was further confirmed by its high student concept coverage (98%) and misconception identification coverage (98%).

Verbal analysis of the think aloud protocols collected during the student essay knowledge map assessment revealed that the experts predominantly used two instructional strategies: increasing student knowledge level and providing comprehensive instruction. These two strategies with typical instructional plan items are listed in Table 4.

Experts selected digital library resources to support their proposed instruction plan that (1) provided comprehensive conceptual coverage, (2) contained key domain concepts, or (3) featured particular types of content (e.g., audio, data, learning materials, tools, visuals). The comprehensive nature of a resource and its coverage of key domain concepts may be directly evaluated against the domain competency model. The misconception type can be used to select digital library resources containing suitable types of content. For instance, spatial misunderstandings may best be addressed through the selection and presentation of digital library resources containing scientifically accurate visuals and diagrams. These analyses of expert assessment of student work reveal that the domain competency model can be used to inform the choice of instructional strategy as well as digital library resource selection to provide personalized instruction.

Table 4: Typical expert instructional strategies

Instructional Strategy	Instructional Plan Items
Increase student knowledge level	• Displaying relevant sections of the domain competency model • Providing digital library resources that explain the relevant scientific processes • Using digital library resources to elaborate current student understanding • Providing digital library resources that help align student knowledge to national science education standards
Provide comprehensive instruction	• Addressing all aspects of the misconception • Providing multiple complementary digital library resources

5. DISCUSSION

The results from our study have provided promising insights into our four guiding research questions.

First, our study results suggest that domain concepts and their relationships may be reliably and consistently identified from a select set of digital library resources. Our reliability results indicate the process of representing digital library resources as knowledge maps is consistent across human experts. This finding bodes well for the potential success of the planned multi-document summarization techniques for automatically constructing a domain competency model. A common difference between expert knowledge maps is the level of detail used to represent a particular concept in a single digital library resource. However, the experts showed a preference for using more detailed representations of key domain concepts over succinct summarizations during the collaborative construction of the final version of the domain competency model. The implication of this finding for the computational construction of the domain competency model is that we need to reconcile the notions of sentence and word compression that typically guide automatic summarization techniques with those of educational relevance and completeness. In other words, we need to infuse the notions of pedagogical relevance and completeness into existing multi-document summarization tools, such as MEAD, perhaps by modifying the computed relevance of the sentences in the digital library resources based on such pedagogical dimensions. In addition, our analysis of the relationships used by experts for knowledge mapping shows that experts often use alternative terminology for the same relationship type. For instance, within the top 10 relationship types used by experts to construct the domain competency model in our study, the terms 'cause' and 'effects of' appear as separate entries. For the purposes of automated relationship tagging, such similar relationships emerge as prime candidates for factoring into a single relationship. Such relationship factoring should assist in improving terminology consistency in the automatically generated domain competency model. Having a reduced and consistent relationship vocabulary should enhance the ability to perform misconception diagnosis automatically as well as the usability of the domain competency model for direct presentation to learners.

Second, our study results also indicate that the domain concepts contained in a select set of digital library resources provide sufficient coverage of important learning goals on earthquakes and plate tectonics for high school age learners. Given the promising results obtained in our learning goals alignment and competency model topic grouping activities, the domain competency model also emerges as a valid candidate as the 'gold-standard' to evaluate the performance of the planned computational model construction algorithms.

Third, the results from the pedagogical utility assessment part of the study illustrate that the domain competency model is useful for the purposes of diagnosing student misconceptions. During the assessment of student work, the model enabled our experts to identify most misconceptions on earthquakes and plate tectonics using a rich set of misconception types. Because the misconception types used by the experts appear to be domain-independent, we believe the computational components developed for identifying these misconception types may also work across other domains. In addition, the misconceptions identified from the student essay knowledge maps provide the basis for an emerging taxonomy of typical student misconceptions on earthquakes and plate tectonics. Such organized collections of typical student understandings are noticeably lacking in Earth science education research as compared to other topics areas such as physics or mathematics [21, 24].

Fourth, the verbal protocol analysis results from the pedagogical utility assessment part of the study illustrate that the domain competency model supports the development of personalized instruction strategies using digital library resources. We have found that experts use a limited set of instructional strategies to address the most frequent misconception types. This small set of strategies also forms the basis for selecting and presenting digital library resources. In addition, since the experts selected digital library resources for high school age learners, the textual content of the nodes in the knowledge map contains material suitable for direct presentation to the learner. The domain competency model may then serve not only as a model for computationally diagnosing student misconceptions and generating instruction plans, but also as a learning resource itself.

Although the results of our study show promise for the automation of the personalization tools envisioned for digital libraries, we acknowledge limitations inherent in our approach. Due to our focus on such a specific topic and age group, our current findings may or may not generalize to other topics or to learning resources designed for different audiences. Even if we are successful at automating the construction of the domain competency model for this topic area, evaluation of such algorithms using different content materials may prove the need for additional studies in other domains. However, our findings do provide new and promising insights into the pedagogical usefulness and computational feasibility of personalized instruction tools underpinned by educational digital libraries.

6. CONCLUSIONS AND FUTURE WORK

Our study of human experts constructing and utilizing a domain competency model indicates that such models may indeed be created from a select set of digital library resources and used to diagnose student misconceptions and to generate personalized instruction plans. In this study, human experts perform these tasks in consistent and predictable ways, thus suggesting that such processes may also be amenable to computational automation. Our evaluation of the contents of the domain competency model indicates that digital library resources may contain the pedagogical breadth and depth of content to construct a comprehensive representation of detailed learning goals for a particular topic. Moreover, our analysis of the strategies used by human experts to construct a domain competency model and to develop customized instruction plans has provided key design requirements for the envisioned personalization tools.

As a result of this study, we have begun the design of computational components to generate domain competency models by extending state-of-the-art multi-document summarization tools to identify concepts and to tag relationships between concepts. We are also designing knowledge map comparison components based on natural language processing techniques and personalized instructional plan generation tools. We hope the promising results reported in this paper will yield equally beneficial advancements on the computational front and further our progress towards supporting automatic conceptual personalization in digital libraries.

7. ACKNOWLEDGMENTS

This work is funded in part by the National Science Foundation through NSF Award IIS/ALT 0537194.

8. REFERENCES

[1] AAAS. *Atlas of Science Literacy*. Project 2061, American Association for the Advancement of Science, and the National Science Teachers Association, Washington, DC, 2001.

[2] AAAS. *Benchmarks for Science Literacy*. Project 2061, American Association for the Advancement of Science, Oxford University Press, New York, 1993.

[3] Adams, A. and Blandford, A. Digital libraries' support for the user's 'information journey'. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, (Denver, Colorado, 2005), 160-169.

[4] Bransford, J.D., Brown, A.L. and Cocking, R.R. (eds.). *How People Learn: Brain, Mind, Experience, and School*. National Academy Press, 2000.

[5] Cañas, A.J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M. and Carvajal, R. CmapTools: A Knowledge Modeling and Sharing Environment. In *Proceedings of the 1st International Conference on Concept Mapping*, (Pamplona, Spain, 2004), 125-133.

[6] Cunningham, S.J., Knowles, C. and Reeves, N. An ethnographic study of technical support workers: why we didn't build a tech support digital library. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, (Roanoke, Virginia, 2001), 189-198.

[7] Digital Library for Earth System Education. DLESE Teaching Boxes - Evidence for Plate Tectonics. Retrieved from http://www.teachingboxes.org/jsp/teachingboxes/plateTectonics/overview/concepts.jsp on 1/8/2007.

[8] Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H.H., Ventura, M., Olney, A. and Louwerse, M.M. AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *36*, *2* (2004), 180-192.

[9] Hall, R.H., Hall, M.A. and Saling, C.B. The effects of graphical postorganization strategies on learning from knowledge maps. *Journal of Experimental Education*, *67*, *2* (1999), 101-112.

[10] Han, H., Giles, C.L., Manavoglu, E., Zha, H., Zhang, Z. and Fox, E.A. Automatic document metadata extraction using support vector machines. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, (Houston, Texas, 2003), 37-48.

[11] Jonassen, D.H. and Grabowski, B.L. *Handbook of Individual Differences, Learning and Instruction*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.

[12] Koedinger, K.R., Anderson, J.R., Hadley, W.H. and Mark, M.A. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8* (1997), 30-43.

[13] Landauer, T.K., Foltz, P.W. and Laham, D. An introduction to latent semantic analysis. *Discourse Processes*, *25* (1998), 259-284.

[14] Marshall, B. and Madhusudan, T. Element matching in concept maps. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, (Tucson, Arizona, 2004), 186-187.

[15] Marshall, B., Zhang, Y., Chen, H., Lally, A., Shen, R., Fox, E. and Cassel, L. Convergence of knowledge management and E-learning: the GetSmart experience. In *Proceedings of the 3rd ACM/IEEE Joint Conference on Digital Libraries*, (Houston, Texas, 2003), 135-146.

[16] McKeown, K.R., Elhadad, N. and Hatzivassiloglou, V. Leveraging a common representation for personalized search and summarization in a medical digital library. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, (Houston, Texas, 2003), 159-170.

[17] National Research Council. *National Science Education Standards*. National Academy Press, Washington, DC, 1996.

[18] Novak, J. and Musonda, D. A twelve-year longitudinal study of science concept learning. *American Educational Research Journal*, *28* (1991), 117-153.

[19] O'Donnell, A.M., Dansereau, D.F. and Hall, R.H. Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, *14*, *1* (2002), 71-86.

[20] Pask, G. *Conversation, Cognition and Learning: A Cybernetic Theory and Methodology*. Elsevier, Amsterdam, The Netherlands, 1975.

[21] Pfundt, H. and Duit, R. *Bibliography. Students' alternative frameworks and science education*. Institute for Science Education at the University of Kiel, Kiel, Germany, 1991.

[22] Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Celebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Toper, M., Winkel, A. and Zhang, Z. MEAD - a platform for multidocument multilingual text summarization. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, (Lisbon, Portugal, 2004).

[23] Ritter, S., Anderson, J., Cytrynowicz, M. and Medvedeva, O. Authoring Content in the PAT Algebra Tutor. *Journal of Interactive Media in Education*, *98*, *10* (1998).

[24] Schoenfeld, A. Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In Grouws, D. ed. *Handbook of research on mathematics teaching and learning*, McMillan Publishing Company, New York, NY, 1992, 334-370.

[25] Scott, B. Conversation theory: A constructivist, dialogical approach to educational technology. *Cybernetics & Human Knowing*, *8*, *4* (2001), 25-46.

[26] Sumner, T., Ahmad, F., Bhushan, S., Gu, Q., Molina, F., Willard, S., Wright, M., Davis, L. and Janee, G. Linking Learning Goals and Educational Resources through Interactive Concept Map Visualizations. *International Journal on Digital Libraries*, *5*, *1* (2005), 18-24.

[27] Susarla, S.C., Adcock, A., Van Eck, R., Moreno, K. and Graesser, A. Development and evaluation of a lesson authoring tool for AutoTutor. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, (Sydney, Australia, 2003), 378-387.

[28] U.S. Department of Education - National Center for Education Statistics. The Condition of Education 2002, U.S. Government Printing Office, Washington, D.C., 2002.

[29] Wade-Stein, D. and Kintsch, E. Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*, *3* (2004), 333-362.

[30] Yilmazel, O., Finneran, C.M. and Liddy, E.D. Metaextract: an NLP system to automatically assign metadata. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, (Tucson, Arizona, 2004), 241-242.