Computers & Education 55 (2010) 1300–1311

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Construction and analysis of educational assessments using knowledge maps with weight appraisal of concepts

C.Y. Su*, T.I. Wang

Department of Engineering Science, National Cheng Kung University, No. 1, University Road, Tainan 701, Taiwan, ROC

ARTICLE INFO

Article history:
Received 15 March 2010
Received in revised form
21 May 2010
Accepted 28 May 2010

Keywords:
Applications in subject areas
Architectures for educational technology system
Elementary education
Human–computer interface
Improving classroom teaching

ABSTRACT

The rapid advance of information and communication technologies (ICT) has important impacts on teaching and learning, as well as on the educational assessment. Teachers may create assessments utilizing some developed assessment software or test authoring tools. However, problems could occur, such as neglecting key concepts in the curriculum or having disproportionate course topics distribution, when teachers create assessments or test items. This study proposes a novel approach, which uses knowledge map with appraisal of concept weights and other ICTs, and implements an assessment system KMAAS to help primary school teachers in Taiwan, or elsewhere, create educational assessments properly. When compiling an assessment, KMAAS ensures that teachers can include all important course concepts intended for assessing and maintain correct balance between course concepts among test items. It does so first by analyzing course material of the assessment range and displaying a concept-weight-annotated knowledge map which concretize and visualize the importance of and the relationships among concepts in the range. It then analyzes the test sheet which is being complied and displays another similar real-time updated knowledge map containing balance between course concepts among the test items. Teachers may cross-refer to these maps to help them adjust concept balances and even select appropriate test items from test banks. The system has being evaluated in both the accuracy of learning concepts extraction and the degree of user satisfaction, as measured by questionnaires given to the teachers who tested the system. The promising results confirm the feasibility of this system in helping teachers compile their educational assessments easily and precisely. Other results of the formative evaluations on techniques have being used to improve the system in order to make it more effective and efficient. The methodology and technologies KMAAS employed are all well developed and are domain independent, which makes it highly flexible to transfer to other course subject domain too.

1. Introduction

Computers and information technology have advanced rapidly in recent years, and computers are now widely used by educators to facilitate teaching and learning (Chou & Liu, 2005; Goodison, 2002). Computer technology allows for innovation in testing and assessment (Bonham, Beichner, Titus, & Martin, 2000) and significantly improves the assessment process for all stakeholders, including teachers, students and administrators (McDonald, 2002). Computerized testing has become a promising alternative to traditional paper-and-pencil measures, thanks to the rapidly decreasing cost of new computers with high processing speeds and large data storage capacities (Barak & Rafaeli, 2004). Much research has evaluated the systems or software tools designed for creating educational assessments and satisfying other test environments needs, such as assisting educators with the construction of assessments, providing data analysis tools for exercises or tests, helping to assess the performance of pupils at various levels, giving quick diagnostic feedback, allowing random item arrangement and improving the convenience and efficiency of computerized test administration (Conole & Warburton, 2005; Huang, Lin, & Cheng, 2009; Nuntiyagul, Naruedomkul, Cercone, & Wongsawang, 2008).

Educational assessment is an essential aspect of teaching. Researchers have observed that teachers in schools spend more than 35% of their time on assessment and more than 10% of their time on assessment-driven instruction (Conca, Schechter, & Castle, 2004). Studies show

* Corresponding author.
E-mail address: n9897108@mail.ncku.edu.tw (C.Y. Su).

0360-1315/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2010.05.027

that for teachers to be effective, they must possess the assessment literacy required to administer and construct proper assessments in their classroom (Mertler, 2004; Popham, 2006). Generating good assessments depends on both the subjective appropriateness of test items and on the way the assessment is constructed (Hwang, Chu, Yin, & Lin, 2008). Some research argues that in order to select appropriate test items, educators must consider multiple criteria, such as the expected time needed for answering the testing, the number of test items, the specified distribution of course concepts to be learned, and the expected degree of difficulty of the test (Chamoso & Ca'ceres, 2009).

Computer-assisted assessment tools have become increasingly indispensable to instructors in primary and secondary education (He & Tymms, 2005); In Taiwan, commercial testing systems and assessment software tools are developed by textbook publishers and made available for primary and secondary educational institutions. Most of these systems allow teachers to construct assessments by selecting test items from item banks, either manually or randomly. However, such manual or random test item selecting strategies are inefficient and are unable to meet multiple assessment requirements simultaneously (Yin, Chang, Hwang, Hwang, & Chan, 2006). When instructors create assessments in this way, they sometimes neglect key concepts in the curriculum or have disproportionate course topics distribution in test items.

Researchers have shown that offering overview supports by providing access to relevant information greatly help people make decisions, solve problems, and avoid disorientation problems (Edwards & Hardman, 1999; Thomas, 2009). Graphic tools or map presentations can be used to create overview supports, restructure content, and deepen or elaborate knowledge of a subject (Zumbach, 2009). The forms of maps can assist people to refer to and facilitate information searches within a domain (Chen, Wei, & Chen, 2008). Knowledge maps provide a graphical representation of the relationships among concepts. They make certain aspects of knowledge more easily understandable (Speel, Shadbolt, Vries, Dam, & O'hara, 1999), and also aid in the acquiring of explicit information or tacit knowledge, illustrating how knowledge flows throughout an organization (Grey, 1999). "Knowledge Maps" could outline where important gaps in received knowledge exist, and have been utilized in the formulation of recommendations in support of a series of related research projects and workshops. An individual knowledge map briefing sheets are meant to serve as quick snapshots of information in key areas related to the use of ICTs in education (Trucano, 2005). Some researchers, such as Pirnay-Dummer and Ifenthaler (2010), build "SMD Technology", based on the theory of mental model and graph theory, use (a) graphical representations such as concept maps or (b) natural language expressions to analyze individual processes in persons solving complex problems at single quantitative measures and standardized representations for qualitative analysis and feedback.

In the sense that the knowledge map is a knowledge representation that reveals underlying relationships of the knowledge sources, using a map metaphor for spatial display. Knowledge maps can be used as primary reference for knowledge acquisition, adjunct aids to text processing or retrieval cues, which serve as scaffolds or supports to cognitive processing for they can reduce cognitive load, enhance representation of relationships among complex constructs, provide support for students whose verbal skills are weak, and serve as important props for communicating shared knowledge (O'Donnell, Dansereau, & Hall, 2002). Hence, it can save much effort with the outline of distribution by knowledge maps assisting; teachers can recall or comprehend more specified distribution of concepts weights when they catch from a knowledge map than when they stand on from textbooks and those teachers with new in-service often benefit the most. This study integrates knowledge maps into educational assessment systems. It builds a novel educational assessment system, called the Knowledge Map Assisted Assessment System (the KMAAS), which automatically constructs real-time updated knowledge maps with weight appraisal of concepts for creating effective assessments. The KMAAS harness a number of technologies like course concept extraction, knowledge maps, natural language processing, items categorization, and items retrieval. This study extracts important "keywords" from textbooks, teaching materials, and test items and uses them as the "nodes" to represent learning concepts in knowledge maps. The "concept weights" are a set of importance values weighted within the curriculum. The terms which directly indicate course topics are considered to play a more important role in assessments (Song, Wenyin, Gu, Quan, & Hao, in press). They are given a higher weight, by multiplying their original weight with a coefficient larger than 1.

When compiling an assessment, teachers always intend to include important course concepts in the assessing and to maintain correct balance between course concepts among test items. Using most of the assessment authoring tools to date or pencils and papers, however, problems could occur, such as neglecting key concepts in the curriculum or having disproportionate course topics distribution, when teachers create assessments or test items. The KMAAS can ensure that these mistakes do not happen. Teachers can first use KMAAS to analyze course materials of the assessment range and display a concept-weight-annotated knowledge map which concretizes and visualizes importance of course concepts and the relationships among these concepts in the range. They then select test items from item banks or create they own test items in a test sheet. In the mean time, KMAAS analyzes the sheet being complied and displays another similar real-time updated knowledge map to show the balance among course concepts in the test items. Teachers can cross-refer to these maps to help them adjust concept balances and select or create appropriate test items. In KKMAS, teachers can analyze and organize their collection of test items and assemble test items they want to use by comparing the weight appraisal of course concepts and their relationships on the knowledge maps. Moreover, they can generate exercises, assessments and tests, and then use the system to help adjust the balance of test items to match the course concept weights.

Although KMAAS was experimented in this study using a science course subject with its related items or assessments, it is not necessarily that KMAAS is confined to such a limited domain. The methodology and technologies KMAAS employed are all well developed and all domain independent. To transfer KMAAS to other subject course material, test items, and test sheets, would require no further labor work than download/upload and creating files.

The rest of this paper is divided into four sections. Section 2 is the literature review, briefly addressing the relation of concepts and test items and providing an overview of knowledge maps and test item categorization. Section 3 introduces the architecture of KMAAS and its component implementations. Section 4 provides a system evaluation and summarizes levels of teacher satisfaction with the system. Finally, conclusions in Section 5 are presented and suggestions are made for future research.

2. Literature review

Course concepts, knowledge maps, and test items categorization play an important role in this study. Researchers have been investing related technologies and have been getting various achievements.

2.1. Course concepts and assessment items

Some researchers argue that the notion of *concept* can be thought of as an atomic unit of knowledge (Chang, Chiu, Lin, & Heh, 2003). Integrated knowledge is created by forming connections among concepts (Hsu, Kuo, Chang, & Heh, 2002). Instruction in schools is intended to teach concepts (Novak, 1979). Each course unit is made up of a certain number of concepts, and each concept has its own size and its own special importance in the course material. In 2006, Agarwal, Edwards, & Perez-Quinone (2006) proposed that when selecting test items, some general principles should be followed. Each item should relate to a specific course concept so that a correct answer implies knowledge of that concept. Also, test items should cater to varied knowledge levels. A number of studies have been conducted related to course concepts and assessment questions. For example, Smith, Wood, and Knight (2008) designed an assessment tool recently to assess central course concepts in genetics and diagnose misconceptions or preconceptions. The created assessments were taken by more than 600 students at three institutions. This tool can be used by genetics instructors to measure students' learning gains as well as to identify strengths and weaknesses in teaching approaches for specific concepts. Huang, Chen, Luo, Chen, and Chuang (2008) proposed to incorporate the intelligent diagnosis and assessment tool into an open software e-learning platform developed for programming language courses. This study applies the corresponding relation of the concepts and questions to explore the degree of student familiarization with the concepts, according to student's assessment records. Their experiments were conducted in two introductory-undergraduate programming courses to examine the effectiveness of the proposed diagnosis and assessment tools. Also, Hwang, Tseng, and Hwang (2008) proposes a novel approach for constructing concept–effect relationships, diagnosing learning problems, and providing personalized learning suggestions for Science and Mathematics courses, based on students' current test and historical assessment records. It has been shown that a clear and strong relationship exists between course concepts and test items. Thus, by determining concept weights for course materials and understanding the relationship between these course concepts and test items in advance can provide important information for supporting teachers when they construct educational assessments. It can also help teachers to know which concepts need further focus or remedial instruction.

2.2. Knowledge map

Many knowledge-representation methodologies have been devised to date for knowledge-based applications. These techniques provide a bridge between human knowledge and machine knowledge (Kim, Suh, & Hwang, 2003). The term graphic knowledge organizer, or knowledge representation tool, is commonly used to describe two-dimensional visual knowledge representations, which show relationships among concepts or processes by means of spatial positions, connecting lines, and intersecting figures (Ives & Hoy, 2003). Knowledge mapping is a method for representing knowledge, which expresses the knowledge structure of a particular theme in the form of a visualization or figure. It can also serve as a common framework or context for the purposes of knowledge sharing, decision-making and problem-solving (Rouse, Thomas, & Boff, 1998). Speel et al. (1999) define knowledge mapping as the process, methods and tools for analyzing knowledge areas in order to discover features or meaning and to visualize these in a comprehensive, transparent form, such that the complexes or implied-relevant features are clearly highlighted. Knowledge maps are created by transferring certain aspects of knowledge into a graphical format that is easy to understand. According to Grey (1999), the knowledge map is an aid in understanding the relationship between explicit information and tacit knowledge, illustrating how knowledge flows throughout the domain. A knowledge map portrays the sources, flows, constraints and terminations of knowledge within the domain. Prusak and Davenport (1998) note that developing a knowledge map involves locating important knowledge in the domain and then publishing some sort of list or picture that shows where to find it. Knowledge maps typically point to people as well as to documents and databases. They help people to understand the relationships between knowledge stores and knowledge dynamics (Grey, 1999). O'Donnell et al. (2002) report that knowledge maps can be used as primary sources for knowledge acquisition, aids for text processing, communication tools for organizing ideas, or retrieval cues. Knowledge maps serve as scaffolds or supports to cognitive processing because they can reduce cognitive load, enhance representation of relationships among complex constructs, provide multiple retrieval paths for accessing knowledge and serve as important props for communicating shared knowledge.

Many definitions of knowledge maps found in the descriptions of knowledge management or in essays are similar, but less specific. For instance, the terms mind maps, concept maps, reasoning maps, topic maps, or semantic networks and even ontology are commonly used in different fields to describe a common set of ideas, or concepts, connected by labeled links. Some researchers (Canas et al., 2003) assume that knowledge maps are synonymous with concept maps, (Novak & Gowin, 1984) which represent externalizing structures and ideas as node-link assemblies (Zumbach, 2009). Other scholars argue that knowledge maps can display complex relationships more easily than concept maps (Crampes et al., 2006). Ontology, on the other hand, stress more on the hierarchical relationship of concepts in a domain, but with the non-consanguineous relationship among concepts implicitly specified in the slots. Knowledge maps, however, stress explicitly more on the general relationships among domain concepts. Knowledge maps can be either strategic or tactical, depending upon the need and intent (Thomas, 2009). Knowledge maps have six key advantages. They can help in the following ways (Thomas, 2009):

- To find key sources, opportunities and constraints to knowledge creation and flows.
- To encourage re-use and prevent re-invention, saving search time and acquisition costs.
- To highlight islands of expertise and suggest ways to build bridges to increase knowledge sharing and exchange.
- To discover effective and emergent communities of practice where informal learning is happening.
- To reduce the burden on experts by helping staff to find critical solutions & information quickly.
- To improve user response, decision making and problem solving by providing access to applicable information, internal and external experts.

The variety of the knowledge map linking systems, which specify the properties of the relationship among concepts, allow for the use of knowledge map in a variety of content areas such as statistics, human physiology, science, math and psychology (Lambiotte, Dansereau, Cross, & Reynolds, 1989; McCagg & Dansereau, 1991). It is also widely used in the education field (Kuo, Lien, Chang, & Heh, 2004; O'Donnell et al., 2002; Shaw, 2010). Initial evaluations on the usage of knowledge maps as navigational guides in web-based environments are promising (Hall, Balestra, & Davis, 2000). Such guides are designed to reflect various knowledge prototypes (Lambiotte et al., 1989). McCagg and Dansereau (1991) point out three main categories of links that can be used in knowledge maps: dynamic links that denote a changing relationship between the linked ideas (e.g., a cause and effect relationship); static links that describe structural relationships between ideas, and elaborative links that extend information. With such richness of link categories, knowledge maps can be used in diverse applications, showing the various nodes and links on the screens in accordance with the purpose of the applications, and can be transferred, either directly or by analogy, from one domain to another.

2.3. Test item categorization

With rapid advancement of information and the Internet technologies, the use of computerized assessments is growing popular in a quick pace. Teachers can now create assessments or exercises using electronic form test items, collected from many sources such as peer teachers, students, schools, textbooks or educational publishing companies. Collected test items are usually put into different item resources without being categorized by topics, or by concepts. When the volume of resources becomes very large, the number of test item combinations increases exponentially, and the item selection process for teachers becomes more time consuming and tedious (El-Alfy & Abdel-Aal, 2008). Moreover, when test items are periodically collected, at each exam or from year to year, test items become grouped by the time rather than by their context (Nuntiyagul et al., 2008). However, teachers always want to retrieve test items according to the categories, not to time, to be in accordance with the curriculum or assessment range. Test items have to be organized and categorized, so that they can be retrieved by subject content, before they can be used effectively (Nuntiyagul et al., 2008). Some information categorization techniques are better at managing test items systematically and automatically, allowing items to be efficiently reused and retrieved (Li & Roth, 2002; Zhang & Lee, 2003). With the recent successes in e-learning and advances in the machine learning techniques, such as information extraction (IE) and natural language processing, efficient automatic test item categorization has become available (Song et al., in press).

Natural language processing technique can be used effectively in language-based domains, such as dialogue systems, speech recognition, informational retrieval and knowledge representation (Alshawi, 1992). As Huang proposes, text categorization involves looking for patterns in natural language text and may be defined as the process of analyzing text to extract information (Hwang, Chu, et al., 2008; Hwang, Tseng, et al., 2008). Some studies have cited keyword extraction as a way to classify text; keywords are usually used for representing key topics included in a research article and are useful for researchers to locate documents of interest (Barki, Rivard, & Talbot, 1993). Luhn (1999) suggests that the occurrence frequency of a word is a useful measurement of word significance. Early classification method often adopts statistics calculations on extracted keywords, then follows up with some of the text categorization techniques that focus on machine learning to construct a classifier and use it to categorize new documents, such as Term Frequency-Inverse Document Frequency (TF-IDF), vector space model (VSM). These techniques are to explore the importance of keywords in documents. For example, the design of the TF-IDF weight means the importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus (Salton & McGill, 1983). The vector space model is that various information retrieval objects are modeled as elements of a vector space. Specifically, terms, documents, queries, concepts, and so on are all vectors in the vector space. The VSM is used to handle multiclass categorization by calculating and comparing the distances between terms. The VSM can also be used to acquire similarities through applications, such as training or making lexical share characteristics with tasks such as information extraction, knowledge extraction from texts and text mining (Basili & Pazienza, 1997; Knight, 1999). Due to the rapid growth of electronic testing items which are mainly text-based shapes, this study makes use of a text categorization technique, which is able to handle specific characteristics of textual data, to help teachers categorize test items automatically and more efficiently.

3. System overview

This study proposes a novel approach, which uses knowledge map with appraisal of concept weights and other ICTs, and implements an assessment system KMAAS to help primary school teachers in Taiwan, or elsewhere, create educational assessments properly. Besides knowledge maps, KMAAS combines natural language processing and machine learning technologies for concept knowledge extraction and test item classification. This section details the system's architecture and its implementation.

3.1. System architecture

The architecture of KMAAS is depicted in Fig. 1, which has six major functional modules, including the item and course material preprocessing, the concept extraction, the concept relation computation, the concept weights computation, the item retrieval processing, and the knowledge map generation modules. The details of each module are described as follows.

(1) Item and course material preprocessing module: This module is responsible for performing sentence segmentation and determining morphological features of test items and course materials. In general, specific text preprocessing is required before automated concept extraction and item categorization can be performed. Chinese sentences are characterized by the absence of delimiters to mark word boundaries like those found in English sentences. Chinese sentence segmentation is thus essential for most language processing systems that utilize techniques originally developed for the English language (Wang, Chang, & Li, 2008). This study employees the CKIP technology (Chen & Ma, 2002; Ma & Chen, 2003) to perform sentence segmentation and text tagging. This technology automatically locates various Chinese keywords in sentences of test items or course materials and tags them with appropriate morphological feature tags; the module selects among them important concept terms and filters out superfluous keywords so that key concept candidate terms can be identified and sent to the next module.
(2) Learning concept extraction module: This module allows teachers to browse and select course concept vocabularies from the concepts repository produced in the preprocessing stage. Teachers can add or modify concept vocabulary, producing a list of important keywords that are then defined as the vocabularies of the course learning concepts. The module uses one of the training datasets to construct a concept extraction model, which is then employed to extract concept vocabularies accurately from test items and course materials.

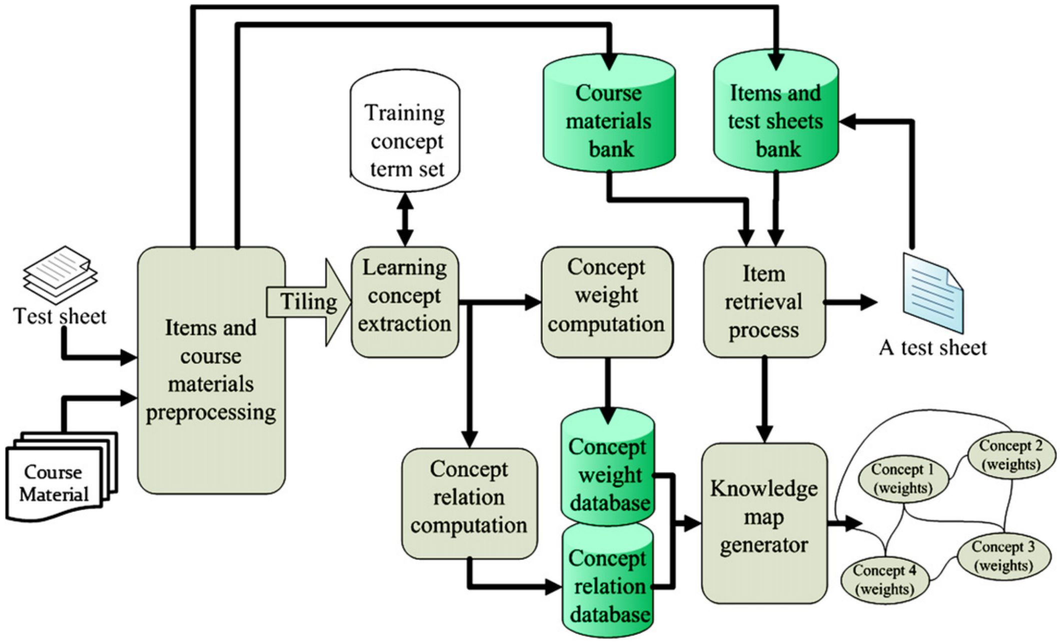

Fig. 1. System architecture.

(3) Concept relation computation module: This module calculates the relationship strength between the concepts extracted from test items and course materials, so that teachers can clearly see the distribution of concepts and the interdependent among these concepts. Formula (1) refers to TF-IDF and calculates the relation between two concepts.

$$RS(Concept1,\ Concept2) = Frequency/AVG(Distance) \quad (1)$$

where RS (Concept1, Concept2) is the degree of correlation between concept1 and concept2. The Frequency is the number of times that concept1 and concept2 appeared in the same sentence in the course materials. The Distance between two concepts is measured by the distance between their representing keywords as they appear in the same sentence and AVG (Distance) is the average distance between two keywords, i.e. summation of the distances divided by the number of times the two keywords appear in the same sentence (Johnson & Wichern, 2002). The higher the RS value is, the stronger the relationship is. A concept relation database is managed by this module.

(4) Concept weight computation module: This module computes the weight of each concept in the concept repository to represent their relative importance in the course material. An algorithm called RSISF (Relation Strength Inverse Sentence Frequency) is proposed for the concept weight conversion. In short, it converses the summation of the relationship strengths a concept has to other concepts into the weight of the concept. Formula (2) realizes RSISF and works as follows.

$$Concept\ Weight = \sum_{i=0}^{n} RS \times \left(\log\frac{N}{n}\right) \quad (2)$$

- RS: the relationship strength between two concepts stored in the concept relation database, as calculated by Formula (1).
- N: the total number of sentences in the course materials.
- n: the total number of sentences in which the concept appears together with others in the course materials.

The higher the concept weight is, the more important the learning concept is in the course material. The results of the computations are stored in a concept weight database. In KMAAS, teachers can manually select different versions of course materials on a same topic and combine them for computations to increase the precision of concept weights.

(5) Item retrieval process module: This module helps teachers with the selection of appropriate test items from the item bank and assists them in compiling a test sheet. All collected items follow the concept classification and are stored in the item bank. The weights and relationships of the concepts in the specified course range for assessment are displayed in their strength order for teachers' reference. When a teacher chooses a concept, the associated test items are retrieved and displayed in the order of their importance, which is based on the criteria such as the number of course concepts an item has, concept strength summation the item has, and how many times this test item has been used.
(6) Knowledge map generation module: This module generates graphical knowledge maps that display relationships and concept nodes using Graphviz technology (Low, 2004). Concepts identified by the learning concept extraction module from the course materials of the specified assessment range are used to generate an annotated knowledge map, with concept weights and link strengths both coming from the concept relation database and the concept weight database. The map concretizes and visualizes the importance of these concepts and their relationships' strengths. This module also analyzes the test sheet which is being currently complied by a teacher with

test items selected from the item bank or created on site and displays another similar real-time updated knowledge map to show the balance among course concepts in the test items. This map depicts the overall structure of all the relevant concepts in the test sheet to assist the teacher in developing a comprehensive understanding of the concepts' distribution among test items and in making decisions about item selections and creations.

3.2. System implementation

The operational Knowledge Map Assisted Assessment System, the KMAAS, can help teachers select test items that effectively assess teaching objectives of a course and avoid inappropriate distribution of test items and concepts. Cross-referring the real-time generated knowledge maps, teachers can analyze and adjust, according to the course concept weights and relation strengths, the configuration proportion of test items in a test sheet.

KMAAS is now being evaluated using the fifth- and sixth-grade "natural science and life technology" curriculum materials in Taiwan and the item bank is initially fed with over 830 related test items. Teachers register for accounts and passwords and the system provided space for teachers to manage and store their created test items and test sheets. Teachers could perform operations using five major interfaces, including item input interface, course materials input interface, data review interface, item selection guide interface, and relevant records interface. Teachers may upload a test sheet with the test items put into the item bank and, in the mean time, receive a concepts distribution analysis knowledge map. Alternatively, teachers could create a new test sheet, using any test items from the item bank, which are in accordance with their testing need and teaching objectives. Either way, a real-time updated knowledge map will always be display for their reference.

Fig. 2 shows the item input interface. In this interface, teachers could add, delete and review electronic test sheets and upload test items in them to the item bank. Teachers hit the "Add a new test sheet" button to add a new text-based test sheet file, fill in the relevant information about the test sheet, and save the file. The system then conduct an analysis on the sheet by, such as CKIP processing, morphological features checking, and automatic concept classifying, on the test item sentences. A knowledge map is displayed afterwards showing the result of the analysis. An example is shown as the lower right corner in Fig. 1.

Fig. 3 shows the course materials input interface, through which teachers add course curriculum materials into KMAAS to facilitate concept extraction from these course materials. The system uses CKIP technique to mark the morphological features of the course materials, removes redundant contents, and extracts candidate concepts from these materials. The data review interface can then be used to help improve the accuracy of automatic concept extractions. Although the system could extract concept vocabularies fully automatically, teachers might want to see how effective the system is in extracting appropriate concept vocabularies from textbooks or from any course materials. When inappropriate concept vocabularies do appear, teachers can manually filter out irrelevant or superfluous ones. This practice ensures the extracted concept keywords are more accurate with each use of the system later. The result is then put into the concept extraction training model for further processing.

Fig. 4 shows the item selection guide interface, through which teachers choose one or more electronic test sheets for final analysis. Teachers could choose students' grade level and then select desired electronic test items with their assessment aims in mind. They can use this interface to see the overview of concept weights and concept relationships for an archived or just compiled test sheet by generating a corresponding knowledge map. The ranking of concept weights and their relationships corresponds to their importance in the course

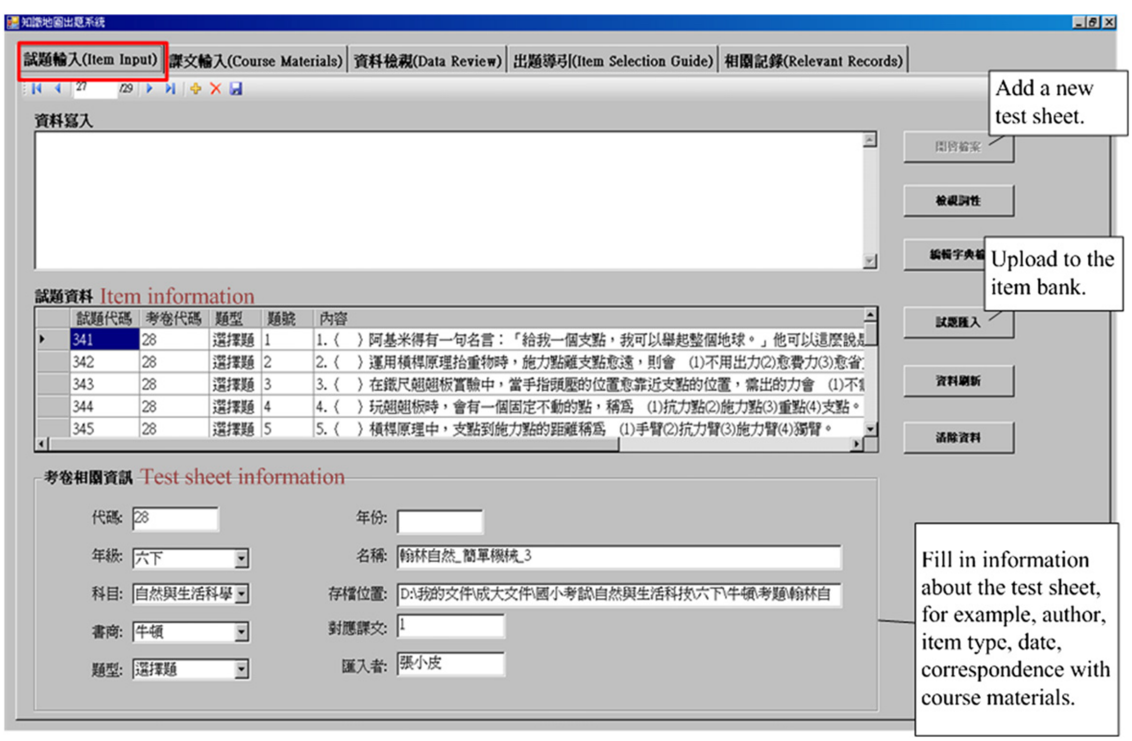

Fig. 2. The items input interface.

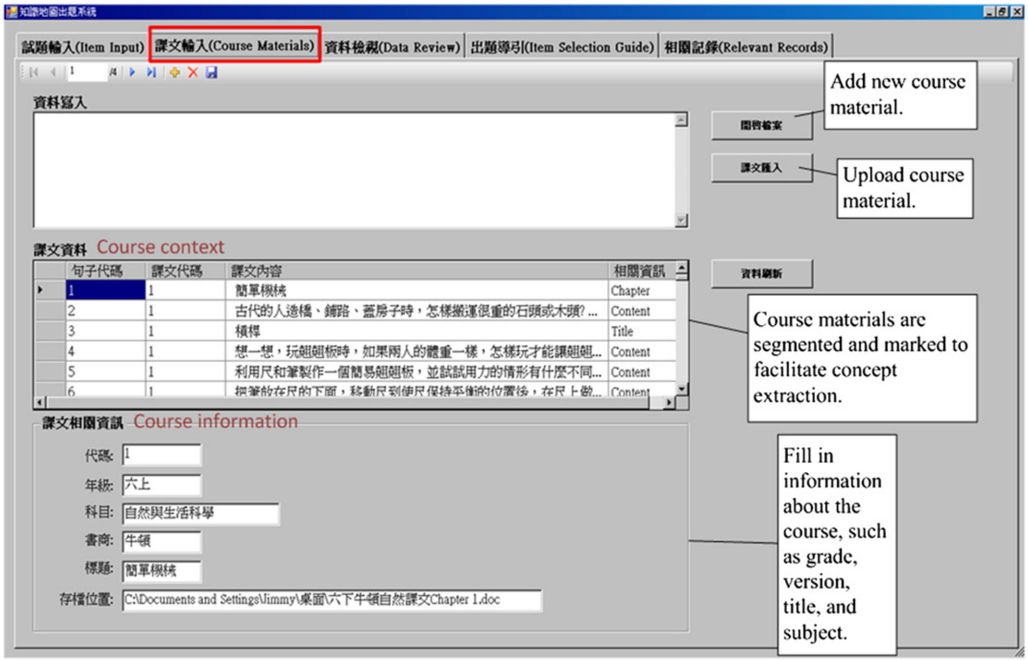

Fig. 3. The course materials input interface.

material. Fig. 5 shows the appraisal of concept weights and their relationships of a knowledge map example on a primary school course on levers. It is created from the combination of three related test sheets after the systematic analysis and shows the whole picture of the core concepts distributions, with regard to the concept trends in terms of distributive test items about the specific subject domain "levers".When teachers click on a concept in the concept list column of the item selection guide interface, all the test items related to this concept would be displayed in the order of their importance, which is based on the criteria such as the number of course concepts an test item has, the concept strength summation the test item has, and how many times this test item has been used before, as shown in an example in Fig. 6. Teachers could select test items from this list in accordance with their assessment objectives. After test items selection is done, teachers could export these items into a test sheet text file, which consists of the selected items and the proportion of each concept, as shown in Fig. 7. They could also transfer the file into an assessment portfolio and save it in the test sheet bank for sharing with other teachers.

4. Evaluation

To make KMAAS more convenient for teachers, it is designed to operate on a personal computer or a notebook, available in most educational settings, and does not require any special CPU configuration. The evaluations conducted are to investigate, in addition to its

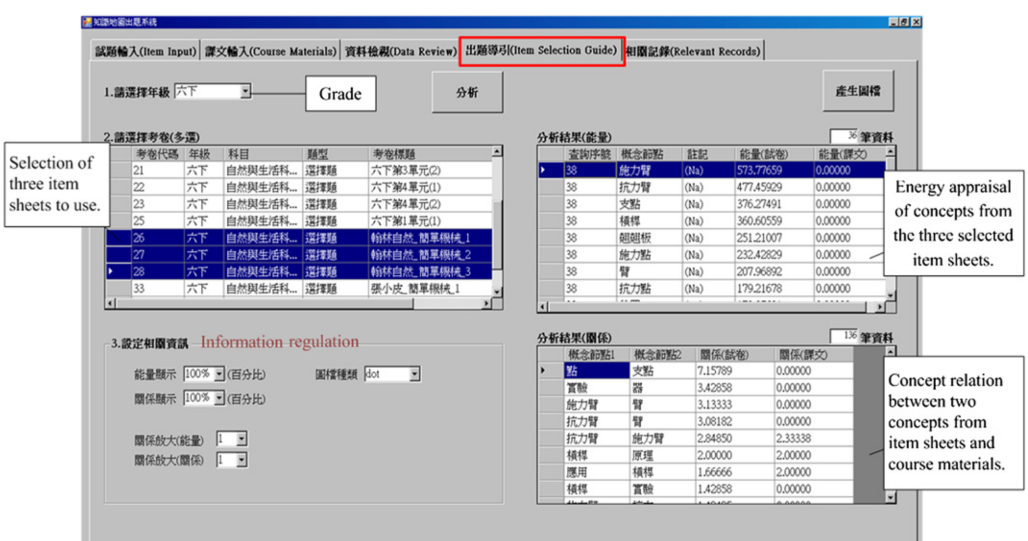

Fig. 4. The item guide interface.

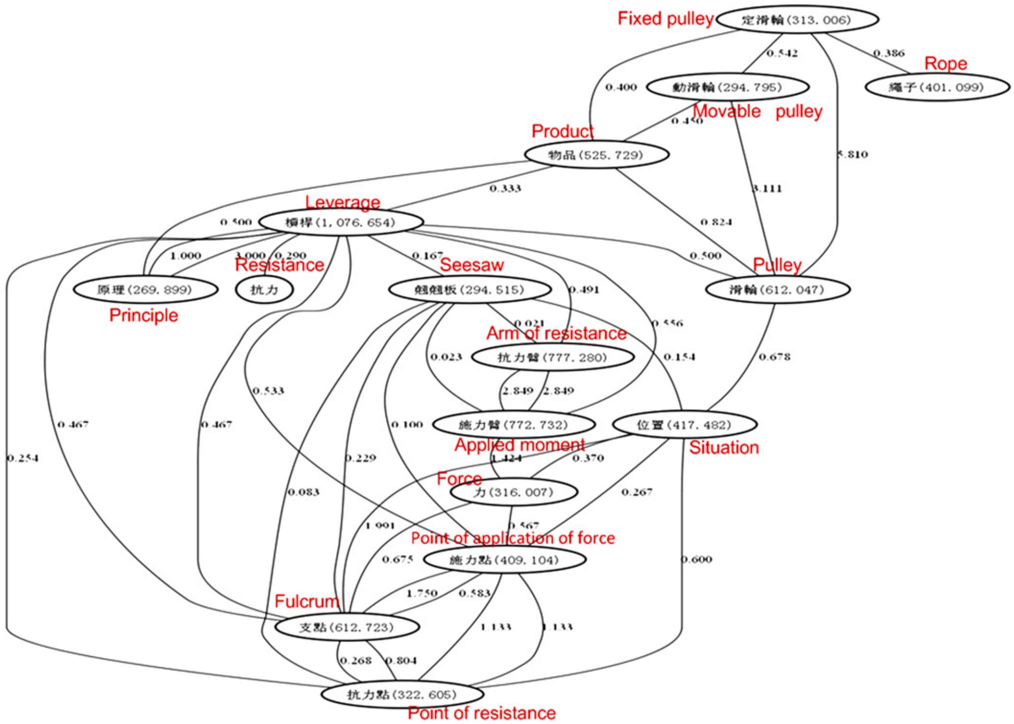

Fig. 5. The knowledge map presents the result of the integration of three electronic item sheets related to course content on the topic of leverage.

usability and appealing, how to improve KMAAS in order to make assessment compilations more effective and more efficient. The method of formative evaluation is successful in that it has resulted in an improved quality of the system. The method, typically involving a small group of users or participants in the project, has been extremely useful in producing modules which are suitable for their target audience, are easy to use, and are instructive. It is also relatively easy for developers to modify the modules according to the recommendations of the evaluators (Phelps & Reynolds, 1999). Using such a method, six elementary school teachers who actually teach a physics course on "levers and forces" in Taiwan were invited to use and evaluate KMAAS. System performance was evaluated in terms of user satisfaction and extraction accuracy. The teachers reported on the usefulness and usability of the system in a Technology Acceptance Model (TAM) questionnaire by providing written feedbacks about their satisfaction in using the system. These evaluations are discussed in the follows. For the extraction accuracy evaluation, standard performance measures, precision and recall, were examined by comparing the concepts and their rankings as extracted and obtained by the system to those obtained by participating teachers, with special focus on whether the concept importance ranking provided by the system are similar to that of by the teachers. The precision and recall measures are defined as that shown in Formula (3).

$$\begin{aligned} &\textit{Precision}\ (p)\ =\ B/(B+C); \\ &\textit{Recall}\ (r)\ =\ B/(A+B) \end{aligned} \tag{3}$$

where A is the concept selected by teachers, C is the concept selected by KMAAS, B is the intersection of A and B.

Fifteen most important concepts of the "levers and forces" course automatically extracted by the system were compared to the 15–20 concepts considered important by teachers. Eleven concepts were common to both lists, and each concept in the list generated by KMAAS was picked an average of 4.27 times from a total of 6 teacher lists. The average precision is 0.733, with the highest being 0.865 and the lowest being 0.61. This result demonstrates that the comparison between automatic system extraction and teachers selection reaches up to 73% consistency.

Numerous researches have shown that ICT technology can enhance and transform teaching and learning in educational environments. Recently, increasing attention has also been paid to understanding what drives teacher's usage of technology. In addition, the pivotal role that teachers play in ensuring effective integration of technology in education has been acknowledged (Teo, Lee, Chai, & Wong, 2009). For this reason, first of the evaluations of this study is set to examine teachers' satisfaction with the KMAAS. The Technology Acceptance Model (TAM) is a robust and widely-used framework for assessing users' satisfaction with some technology (King & He, 2006; Venkatesh & Davis, 2000). The questionnaire used in this study is based on the TAM model and responses are collected using five-point Likert scales. Six teachers were invited to complete a questionnaire eliciting information concerning attitudes and responses; the results are shown in appendix A. The findings show that the mean score for most of the items is greater than 4, except for 10th and 15th items. Four of the teachers mentioned the disfluency of the interfaces, but suggested that their problems could be overcome by reading the manual or using the system more frequently. The results are overall very positive. After completing the questionnaire, teachers were interviewed for about 30 min on their views on using the system. Some examples of the feedbacks are as follows:

> This system is very novel, I can use it to analyze my post test sheets and get some important information to help me to comprehend or regulate my ideas on the assessment work (teacher B).
>
> This system assists me in compositing a proper test sheet, in addition to helping me select test items, I could see the concept distribution percentage by the knowledge maps helping (teacher F).

Fig. 6. The item selection interface.

Furthermore, several comments in the interviews regarding the presentation of user interface and knowledge map are as follows:

> Although the interfaces seem to be a little bit rough, they did not affect my operations, in contrast, I was looking forward to it helping my assessment works in the future (teacher A).
> I thought that this system is very meaningful for teachers' assessment works, especially the knowledge map. I suggest that the concept node might attach color labels in next version as well (teacher C).

In the evaluations, in general, all teachers felt that the system, which provided them with knowledge maps and weight appraisal of concepts, was very novel; they reported favorably on the value of the system for assisting with their assessment works. The knowledge maps, by aggregating the concepts and presenting the whole assessment structure, enable teachers to make inferences and to reflect on test items and allowing them to make good decision on item selection.

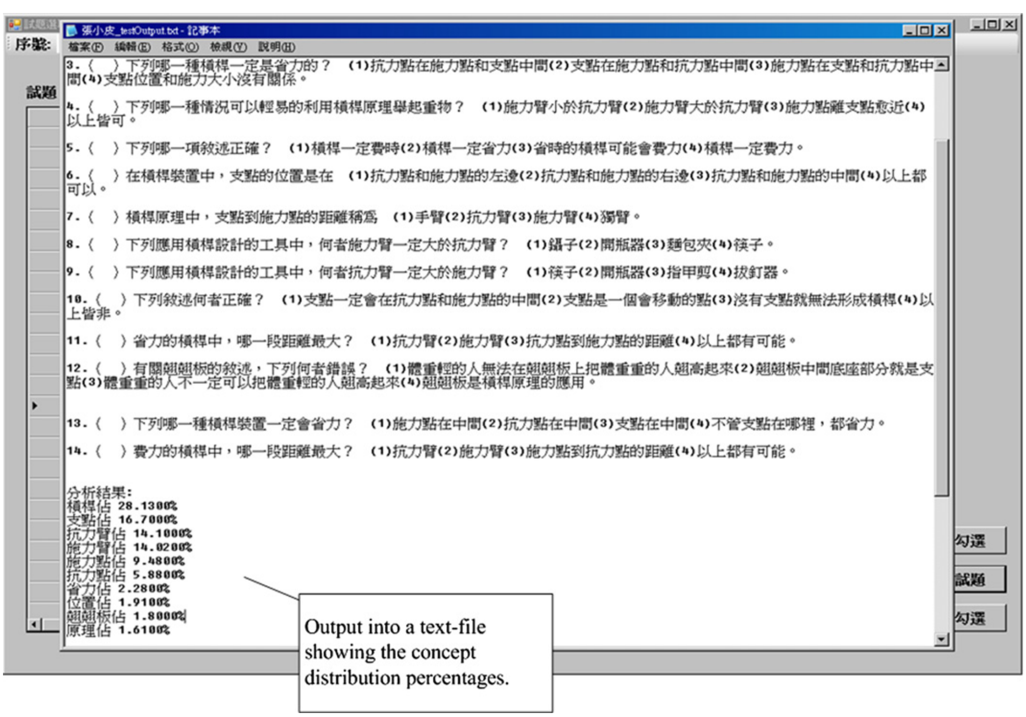

Fig. 7. The text file presents the exported items and the proportion of concepts.

4.1. Limitations of the paper

The evaluation objective set for this study is to let participants use KMAAS for their actual assessment works when teaching 5th and 6th grades primary school “natural science and life technology” courses in order to get more precise feedbacks. Owing to the short overlap period between the semester and the evaluation, the first limitation of this paper so far concerns the amount of time that the teachers spent on using the system. Teachers have busy work days, and could not spend too much of their time for evaluations; they were pleaded to use the system as much as possible within the period of the evaluation. The second limitation is the number of participants because the evaluation wanted to invite teachers with actual experience in teaching targeted courses, which in this case is the “levers and forces” courses, and it was not easy to fulfill the criterion. So a small number of teachers who participated in this study does not meet this basic conditions for statistical analysis, which may possibly affects the results of the questionnaire analysis. The final limitation concerns the interface design which was then preliminary and rough. However, the result of this research has the undeniable merit of offering valuable insights into this fast growing ICT-based educational assessment domain, and some of the limits may be offset by the use of TAM based questionnaire for evaluation surveys.

5. Conclusion and further investigation

This study proposes a novel methodology that harnesses knowledge maps with weight appraisal of concepts and concept relationships and builds an assessment system for assisting primary school teachers in creating assessments. KMAAS enables teachers to analyze and reorganize their collection of test sheets and test items, and provides conceptually-organized methods for retrieving test items. It does so by performing learning concept extractions and automatic classifications on each collected test item to establish the associations between concepts and test items. When compiling a test sheet, teachers may refer to an automatically-constructed knowledge map presenting the importance of course concepts and interrelationship among course concepts and test items. This map helps teachers to understand and evaluate test item distributions according to the weight appraisal of concepts. It also serves as a useful blueprint for assisting teachers as they design various assessment projects. Moreover, teachers can use KMAAS to create and manage their own assessment e-portfolios and to share them with each other via the network. For the evaluation of KMAAS, evaluation have been carried out in terms of the accuracy of learning concepts extraction by KMAAS and user satisfaction, measured by a questionnaire given to teachers who tested the system. The results promise that KMAAS has a high potential and high practical value for succeeding in ICT-based educational assessment domain, and can definitely help teacher with their assessment creation works.

Although KMAAS was proved successfully in a science course subject with its related items or assessments, it is not necessarily that KMAAS is confined to such a limited domain. The methodology and technologies KMAAS employed are all well developed and indeed perform automatically with very few human intervention. For other subject course material, test items, and test sheets, using KMAAS would require no further labor work than download/upload and creating files. In the near future, we will extend the depth and breadth of the KMAAS in several technical ways, such as increasing the volume for test items, covering diverse course materials to cater for different teachers’ requirements, providing information about test item difficulty and discrimination, and presenting a two-way specification table to assist teachers in selecting test items. Improving the user interface to make the system more appealing to teachers is also an urgent work. The techniques used and reviewed in this paper are mainly based on knowledge maps and computer-assisted assessment. In the future, other techniques may be investigated to further improve the effectiveness of using KMAAS by improving the formulae of calculating concept weights and relationship strengths and by considering the accumulated teachers assessment knowledge.

Appendix A

Items	Mean	Standard deviation
Perceived usefulness		
1. Using the KMAAS improves my educational assessment work performance.	4.333	0.51640
2. Using the KMAAS increases my making educational assessment productivity.	4.333	0.51640
3. Using the KMAAS enhances my effectiveness in my assessment work.	4.333	0.51640
4. I find the KMAAS to be useful in my assessment work.	4.333	0.51640
5. Using the KMAAS would make my assessment work easier.	4.0000	0.00000
6. Using the KMAAS would enable me to make assessment quickly.	4.0000	0.63246
Perceived ease of use		
7. It would be easy for me to become skillful at using the KMAAS.	4.1667	0.75277
8. My interaction with the KMAAS is clear and understandable.	4.0000	0.89443
9. I believe that it is easy to get the KMAAS to do what I want it to do.	4.1667	0.40825
10. Interacting with the KMAAS does not require a lot of my mental effort.	3.8333	0.40825
11. Overall, I find the KMAAS to be easy to use.	4.3333	0.51640
Attitude		
12. I like using the KMAAS.	4.1667	0.40825
13. The KMAAS is fun to use.	4.1667	0.40825
14. The KMAAS provides an attractive working environment.	4.0000	0.63246
Technological complexity		
15. Learning to use the KMAAS takes up too much of my time.	3.8333	0.40825
16. Using the KMAAS involves too much time.	4.0000	0.63246
17. It takes too long to learn how to use the KMAAS.	4.0000	0.63246

(*continued on next page*)

Appendix (*continued*)

Items	Mean	Standard deviation
Concentration		
18. I was absorbed intensely in the educational assessment activity.	4.3333	0.81650
19. My attention was focused on the educational assessment activity.	4.1667	0.75277
20. I was deeply engrossed in the educational assessment activity.	4.1667	0.40825
Usage intentions		
21. Assuming I have access to the KMAAS, I intend to use it.	4.5000	0.54772
22. Given that I have access to the KMAAS, I plan to use it.	4.5000	0.54772
23. I intend to increase my use of the KMAAS in the future.	4.5000	0.54772
24. It is worth to use the KMAAS.	4.5000	0.54772
25. I will frequently use the KMAAS in the future.	4.3333	0.51640

$n = 6$.

References

Agarwal, R., Edwards, S., & Perez-Quinones, M. (2006). *Designing an adaptive learning module to teach software testing*. Paper presented at the SIGCSE'06, Houston, Texas, USA.

Alshawi, H. (1992). *The core language engine*. Cambridge: MA The MIT Press.

Barak, M., & Rafaeli, S. (2004). On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning. *International Journal of Human-Computer Studies, 61*(1), 84–103.

Barki, H., Rivard, S., & Talbot, J. (1993). A keyword classification scheme for IS research literature: an update. *MIS Quarterly* 209–226.

Basili, R., & Pazienza, M. (1997). Lexical acquisition for information extraction. *Lecture Notes in Computer Science* 44–72.

Bonham, S., Beichner, R., Titus, A., & Martin, L. (2000). Education research using web-based assessment systems. *Journal of Research on Computing in Education, 33*(1), 28–45.

Canas, A., Coffey, J., Carnot, M., Feltovich, P., Hoffman, R., Feltovich, J., et al. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. *Final Report to CNET*. (Retrieved).

Chamoso, J. M., & Ca'ceres, M. J. (2009). Analysis of the reflections of student–teachers of mathematics when working with learning portfolios in Spanish university classrooms. *Teaching and Teacher Education, 25*(1), 198–206.

Chang, J. -C., Chiu, Y. -P., Lin, Y. -Y., & Heh, J. -S. (2003). Learning diagnosis process with SCORM Compatible learning materials. Paper Presented at the WISCS 2003.

Chen, K. -J., & Ma, W. -Y. (2002). Unknown word extraction for Chinese documents. Paper Presented at the Proceedings of the 19th International Conference on Computational Linguistics.

Chen, N., Wei, C., & Chen, H. (2008). Mining e-Learning domain concept map from academic articles. *Computers & Education, 50*(3), 1009–1021.

Chou, S., & Liu, C. (2005). Learning effectiveness in a web-based virtual learning environment: a learner control perspective. *Journal of Computer Assisted Learning, 21*(1), 65–76.

Conca, L., Schechter, C., & Castle, S. (2004). Challenges teachers face as they work to connect assessment and instruction. *Teachers and Teaching, 10*(1), 59–75.

Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J, 13*(1), 17–31.

Crampes, M., Ranwez, S., Villerd, J., Velickovski, F., Mooney, C., Emery, A., et al. (2006). Concept maps for designing adaptive knowledge maps. *Information Visualization, 5*(3), 211–224.

Edwards, D., & Hardman, L. (1999). mapping and navigation in a hypertext environment. *Hypertext: Theory Into Practice, 90*.

El-Alfy, E.-S. M., & Abdel-Aal, R. E. (2008). Construction and analysis of educational tests using abductive machine learning. *Computers & Education, 51*(1), 1–16.

Goodison, T. (2002). Learning with ICT at primary level: pupils' perceptions. *Journal of Computer Assisted Learning, 18*(3), 282–295.

Grey, D. (1999). Knowledge mapping: a practical overview. From: http://www.impactalliance.org/file_download.php?location=S_U&filename=10383546681Knowledge_Mapping.htm Retrieved 9.14, 2009.

Hall, R., Balestra, J., & Davis, M. (2000). *A navigational analysis of linear and non-linear hypermedia interfaces*. Paper presented at the the annual meeting of the American Educational Research Association, New Orelans, LA.

He, Q., & Tymms, P. (2005). A computer-assisted test design and diagnosis system for use by classroom teachers. *Journal of Computer Assisted Learning, 21*(6), 419–429.

Hsu, C., Kuo, R., Chang, M., & Heh, J. (2002). *Implementing a problem solving system for physics based on knowledge map and four steps problem solving strategies*. Paper presented at the IEEE 2nd international conference on advanced learning technologies, Kazan, Russia.

Huang, C.-J., Chen, C.-H., Luo, Y.-C., Chen, H.-X., & Chuang, Y.-T. (2008). Developing an intelligent diagnosis and assessment E-learning tool for introductory programming. *Educational Technology & Society, 11*(4), 139–157.

Huang, Y.-M., Lin, Y.-T., & Cheng, S.-C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers & Education, 52*(1), 53–67.

Hwang, G., Tseng, J., & Hwang, G. (2008). Diagnosing student learning problems based on historical assessment records. *Innovations in Education and Teaching International, 45*(1), 77–89.

Hwang, G. J., Chu, H. C., Yin, P. Y., & Lin, J. Y. (2008). An innovative parallel test sheet composition approach to meet multiple assessment criteria for national tests. *Computers & Education, 51*(3), 1058–1072.

Ives, B., & Hoy, C. (2003). Graphic organizers applied to higher-level secondary mathematics. *Learning Disabilities Research & Practice, 18*(1), 36–51.

Johnson, R., & Wichern, D. (2002). *Applied multivariate statistical analysis* (5th ed.). New York: Prentice Hall.

Kim, S., Suh, E., & Hwang, H. (2003). Building the knowledge map: an industrial case study. *Journal of Knowledge Management, 7*(2), 34–45.

King, W., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740–755.

Knight, K. (1999). Mining online text. *Communications of the ACM, 42*(11), 58–61.

Kuo, R., Lien, W., Chang, M., & Heh, J. (2004). Analyzing problem's difficulty based on neural networks and knowledge map. *Journal of Educational Technology and Society, 7*(2), 42–50.

Lambiotte, J., Dansereau, D., Cross, D., & Reynolds, S. (1989). Multirelational semantic maps. *Educational Psychology Review, 1*(4), 331–367.

Li, X., & Roth, D. (2002). Learning question classifiers. In *Proceedings of the 19th international conference on computational linguistics*. Taipei, Taiwan.

Low, G. (2004). Graphviz – graph visualization software. From: http://www.graphviz.org/About.php Retrieved 10.12, 2008.

Luhn, H. (1999). The automatic creation of literature abstracts. In M. Inderjeet, & M. T. Maybury (Eds.), *Advances in automatic text summarization* (pp. 15–21). Cambridge, MA: The MIT Press.

Ma, W. -Y., & Chen, K. -J. (2003). Introduction to CKIP Chinese word segmentation system for the first international Chinese word segmentation bakeoff. Paper Presented at the Proceedings Of The Second SIGHAN Workshop on Chinese Language Processing.

McCagg, E., & Dansereau, D. (1991). A convergent paradigm for examining knowledge mapping as a learning strategy. *The Journal of Educational Research, 84*(6), 317–324.

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education, 39*(3), 299–312.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education, 32*(3), 49–64.

Novak, J. (1979). The reception learning paradigm. *Journal of Research in Science Teaching, 16*(6), 481–488.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge Univ Press.

Nuntiyagul, A., Naruedomkul, K., Cercone, N., & Wongsawang, D. (2008). Adaptable learning assistant for item bank management. *Computers & Education, 50*(1), 357–370.

O'Donnell, A., Dansereau, D., & Hall, R. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review, 14*(1), 71–86.

Phelps, J., & Reynolds, R. (1999). Formative evaluation of a web-based course in meteorology. *Computers & Education, 32*(2), 181–193.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77–115). Springer.

Popham, J. (2006). Needed: a dose of assessment literacy. *Educational Leadership, 63*(6), 84–85.

Prusak, L., & Davenport, T. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard business school press.

Rouse, W., Thomas, B., & Boff, K. (1998). Knowledge maps for knowledge mining: application to R&D/technology management. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 28*(3), 309–317.
Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill, Inc.
Shaw, R.-S. (2010). A study of learning performance of e-learning materials design with knowledge maps. *Computers & Education, 54*(1), 253–264.
Smith, M., Wood, W., & Knight, J. (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE-Life Sciences Education, 7*(4), 422–430.
Song, W., Wenyin, L., Gu, N., Quan, X., & Hao, T. Automatic categorization of questions for user-interactive question answering. *Information Processing & Management*, in press.
Speel, P., Shadbolt, N., Vries, W., Dam, P., & O'hara, K. (1999). Knowledge mapping for industrial purposes. Paper Presented at the 12th Workshop on Knowledge Acquisition, (Modeling and Management, Alberta, Canada).
Teo, T., Lee, C. B., Chai, C. S., & Wong, S. L. (2009). Assessing the intention to use technology among pre-service teachers in Singapore and Malaysia: a multigroup invariance analysis of the technology acceptance model (TAM). *Computers & Education, 53*(3), 1000–1009.
Thomas, A. (2009). Knowledge mapping. From: http://pmtips.net/knowledge-mapping/ Retrieved 12.19, 2008.
Trucano, Michael (2005). *Knowledge maps: ICT in education*. Washington, DC: Infodev/World Bank. Available at: http://www.infodev.org/en/Publication0.8.html.
Venkatesh, V., & Davis, F. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*(2), 186–204.
Wang, H.-C., Chang, C.-Y., & Li, T.-Y. (2008). Assessing creative problem-solving with automated text grading. *Computers & Education, 51*(4), 1450–1466.
Yin, P., Chang, K., Hwang, G., Hwang, G., & Chan, Y. (2006). A particle swarm optimization approach to composing serial test sheets for multiple assessment criteria. *Educational Technology and Society, 9*(3), 3–15.
Zhang, D., & Lee, W. S. (2003). Question classification using support vector machine. In *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 26–32), Toronto, Canada.
Zumbach, J. (2009). The role of graphical and text based argumentation tools in hypermedia learning. *Computers in Human Behavior, 25*(4), 811–817.