

This article was downloaded by: [National Chiao Tung University 國立交通大學]

On: 24 April 2014, At: 23:02

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Interactive Learning Environments

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/nile20>

Finding trustworthy experts to help problem solving on the programming learning forum

Shian-Shyong Tseng^{a b} & Jui-Feng Weng^a

^a Department of Computer Science, National Chiao Tung University, Hsinchu, 30010, Taiwan

^b College of Computer Science, Asia University, Taichung, 41354, Taiwan

Published online: 07 Dec 2009.

To cite this article: Shian-Shyong Tseng & Jui-Feng Weng (2010) Finding trustworthy experts to help problem solving on the programming learning forum, *Interactive Learning Environments*, 18:1, 81-99, DOI: [10.1080/10494820903195264](https://doi.org/10.1080/10494820903195264)

To link to this article: <http://dx.doi.org/10.1080/10494820903195264>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Finding trustworthy experts to help problem solving on the programming learning forum

Shian-Shyong Tseng^{a,b*} and Jui-Feng Weng^a

^aDepartment of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan;

^bCollege of Computer Science, Asia University, Taichung 41354, Taiwan

(Received 12 August 2008; final version received 14 December 2008)

The most important thing for learners in Programming Language subject is problem solving. During the practical programming project, various problems may occur and learners usually need consultation from the senior programmers (i.e. the experts) to assist them in solving the problems. Thus, the inquiry-based learning with learning forum is applied to assist the programming problem solving. However, even if the learning community of the forum is provided, the finding of trustworthy and available experts for improving the quality of social interactions is still difficult for students. Therefore, the idea of applying the social network service of Web 2.0 with the trustworthy experts finding service is proposed to actively consult the experts based on their topic of interest, trustworthiness, and availability. The experts' topic interest and trustworthy degree are obtained from the experts' posting documents on the forum. To maintain the dynamic discussion topics and avoid the synonym problem on the forum, the Programming Capability Ontology is constructed as the consensus taxonomy by the distributed clustering algorithm. The self-organized ontology maintenance scheme is also proposed to maintain and update the new topic keywords in the ontology. The availability of experts can be approximately calculated from their online presence log. Moreover, the updating of the experts' presence value is modeled by the fading probability function of the Ant Colony Algorithm. Finally, the quality of the expert finding service and learners' satisfaction with the expert finding service has been evaluated. The experimental result shows that the feasibility and the effectiveness of the proposed approach are satisfactory.

Keywords: trustworthy social network; ontology; collective intelligence; programming learning; forum

1. Introduction

The most important thing for learners in Programming Language subject is problem solving. During the practical programming implementation project, various problems such as how to use program library, what the meaning of the program is, what is wrong with the program source code, etc., may occur. Thus, the skill of problem solving through social interactions with the experts in the knowledge society

*Corresponding author. Email: sstseng@cs.nctu.edu.tw

(i.e. the forum) to find the solution is important for learners. In this article, the inquiry-based learning on the Web forum is applied to assist the training of problem solving for Computer Science students. However, how to identify the problem and find the right experts are important for students to improve the quality of social interactions.

With the growth of Web 2.0, the social network service is emphasized as one of the important factors of the Internet applications. The Internet reduces the cost of communications among peers around the world and motivates the emergence of virtual learning communities in cyberspace. It also motivates several researches about how to facilitate the social network service to enhance the learning collaborations.

In the programming learning communities, the learning forum is commonly used as a collaborative learning platform. In the forum, the inquiry-based learning begins when the learner encounters a problem. Thus, the learner would logon to the programming learning forum to start the inquiry by posting his/her question about the problem. The senior programmers (i.e. the experts) or other learners with similar topic interest will give feedback to the learner's problem. The inquiry and feedback cycle is repeatedly executed until the learner is satisfied with the solution. However, how to construct a social network and consult suitable experts to answer the question is usually difficult for learners. Therefore, how to assist the learner to bridge the 'problem solving social network' in the virtual learning community is investigated.

With our observations of senior programmers, it is identified that the right experts for consultation should be the ones who have interest and trustworthiness regarding the student's problem and who still have presence on the forum. Therefore, the trustworthy experts finding service is proposed to facilitate the social network service of Web 2.0 by actively consulting the experts based on their topic interest, trustworthiness, and availability.

Because experts on the learning forum discuss by posting documents, their posting documents may reflect their topic interest. However, there is a synonym issue in these documents. To solve this issue, the ontology-based approach is applied to represent the profile of the experts by consensus taxonomy. Accordingly, the Programming Capability Ontology (PCO) is proposed as shown in Figure 1. It should be noticed that there are four layers in PCO, where the first is Category Layer including different programming platforms to classify the programming learning topics into predefined categories; the second is the Topic Layer including predefined different types of questions to represent possible topic interests of experts such as 'what's the meaning' or 'what's wrong'; the third is Issue Layer which denotes the frequently discussed keywords in each topic, and the fourth is Document Layer which denotes the forum documents raw data.

With the defined ontology, the issues of how to judge the trustworthiness and presence of the experts still have to be solved.

As mentioned earlier, the quality of experts finding service highly relies on the indicators of the topic interest, trustworthiness, and availability of the experts. When a question is posted on the forum, the experts who meet the required criteria will be actively invited to join the discussion. Accordingly, the objective functions of trustworthiness and availability are defined as follows.

Firstly, the experts' topic interest is modeled by the keywords of topic layer in the PCO and the values of keyword vector are obtained from experts' posting documents

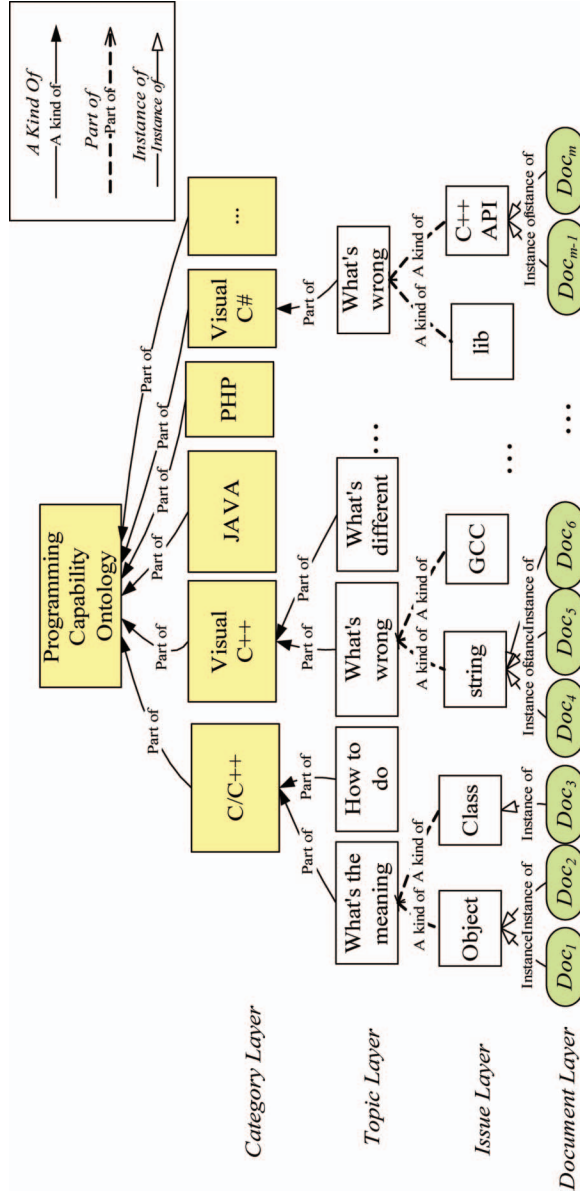


Figure 1. The programming capability ontology.

on the forum. The trustworthy value is computed by the experts' reputation degrees given by other community members. Thus, the trustworthiness is computed from the similarity values of the student's question with the expert's topic interest and associated trustworthy value. Secondly, the experts' availability is heuristically obtained by the weighted average of frequency of their presence online. Moreover, the fading of presence value is modeled by weighting with probability function update of Ant Colony Algorithm (Dorigo & Stützle, 2004). Accordingly, the objective function of trustworthiness and availability can be formulated to easily support the expert finding service. With the configuration of the parameters in objective function, three expert finding strategies including trustworthiness first, topic interest first, or availability first are further proposed to meet the different requirements.

To support ontology construction, the documents of the forum are firstly represented as the keyword space model. Next, the keyword vector-based clustering approach is applied to construct the ontology. Because the documents of the forum are entered incrementally, the Self-Organized Ontology Maintenance Scheme is proposed to support the maintaining of the PCO. To enhance the efficiency of the ontology construction, the documents can be classified into predefined categories and topics first. And then, the distributed clustering algorithm is proposed to maintain the concepts in issue layer of PCO. Although the new documents are added, only corresponding parts of ontology need to be updated.

Finally, to evaluate the performance of our scheme, the online programming learning forum includes around 14,000 forum documents which have been collected and analyzed. The experimental results show the higher feasibility and effectiveness of the expert finding service.

2. Related works

2.1. Computer-assisted programming learning

In the research of computer assisted programming learning, previous studies proposed the algorithm animations (Crews & Ziegler, 1998; Garner, 2003) or intelligent tutoring systems with model tracing approaches (Anderson, Corbett, Koedinger, & Pelletier, 1995; Ramadhan, 1997; Kumar, 2003) to assist the novice learners in understanding the program execution processes. In general, these works focused on the syntactic level of programming learning.

Besides the syntactic level learning, the training of problem solving skills such as the learning through project-based learning (Chen & Cheng, 2007; Clark, et al. 2007) or problem-based learning (Ryoo, Fonseca, & Janzen, 2008) with innovative programming laboratories were investigated. The interesting learning context such as game design was adopted to motivate the learners' engagement. As even small projects are usually implemented by teamwork, the collaboration among members becomes a new issue. Thus, researches based on the social-culture constructivism were proposed to provide the collaborative programming environments (Preston, 2005; Chen, 2005; Moreno, Myller, & Sutinen, 2004) or the peer assessment activities (Lin, Liu, & Yuan, 2001; Bhalerao & Ward, 2001; Sitthiworachart & Joy, 2004). The collective, collaborative learning tools such as discussion board, e-mail, etc. are integrated in the learning platform.

In summary, from syntactic level programming learning to collaborative programming learning, the training of programming has progressed to a more open-ended problem solving learning environment.

2.2. Social network to support the learning

The term Web 2.0 describes the changing trend of web platform design's aim to enhance the social networking sites, content sharing sites, wikis, blogs, and folksonomies. 'The Web 2.0 is the business revolution' described by Tim O'Reilly (2005) in the first O'Reilly Web 2.0 conference 2004 (O'Reilly, 2005) makes the term notable. With the growth of the Web 2.0, the social network service is emphasized as one of the important factors of the Internet applications. The Internet reduces the cost of communications among peers around the world and motivates the virtual learning communities in the cyberspace. Thus, the emergence of virtual learning communities has stimulated several researches about how to facilitate the social network service to enhance the learning collaborations.

The learning forum is commonly used as collaborative learning platform to show the power of social network in supporting the learning of higher education. Hou, Chang, and Sung (2008) analyzed the problem-solving-based online discussion and derived the sequential pattern and behavior of students. The results showed that, compared with the single topic appointed by teacher, the problem solving online discussion activity is more helpful for students' knowledge construction. Simpson, Reynolds, Light, and Attenborough (2008) analyzed the innovative project involving mental health service users in the education of preregistration mental health nursing students through enquiry-based learning. Zhu (2006) discussed the different levels of cognitive engagement for students' online discussions. Dengler (2008) discussed the analysis of critical thinking and problem solving using the learning forum and the results showed that it enhanced the participation of students who may feel more inhibited to engage in the discussion. The D.I.A.S. system (Bratitsis & Dimitracopoulou, 2005) analyzed the learners' interactive behaviors and posting activities in the learning forum based upon the scoring approach. Other researches also showed the effectiveness of applying collaborations on learning forum for discourse training, critical thinking, and collaborative learning (Rourke, Anderson, Garisson, & Archer, 2001; Caspi, Gorsky, & Chajut, 2003; Cazden & Beck, 2003; Hu & Yang, 2005).

Several studies have suggested that enhancing social presence in an e-learning environment can induce and sustain learners' motivation to create the impression of a quality learning experience on the learner (Newberry, 2001; Tu, 2001; Tung & Deng, 2006; Aragon, 2003). The research (Wasko & Faraj, 2005) also found that the knowledge sharing is still the motivation for participation in virtual communities. These studies have provided evidence that demonstrates the importance of social presence and knowledge exchange in enhancing learning performance. However, knowledge sharing requires mutual-trust collaboration between learners (Yang, Chen, Kinshuk, & Chen, 2007). The research (Bulu & Yildirim, 2008) also found that groups with different trust levels show different communication behaviors throughout the study.

In summary, the value of social presence and virtual community to support the problem solving is obviously high. However, how to enhance the consideration of trustworthiness is still an interesting and important issue to facilitate the social network service of Web 2.0 in programming learning.

2.3. *Ontology building approaches*

To obtain the problem solving knowledge on the forum for showing the topic interests of learning peers, the ontology building approaches are reviewed. The Dictionary-based approach (Khan & Luo, 2002) was proposed to construct the ontology based on a traditional dictionary, which presents the related concepts of words, including synonyms, etymology, etc. The association rules mining (Maedche, 2001) was proposed to construct the ontology by computing the frequency of an association of terms in the text repositories. If the frequency of the association is close to the occurrence of individual terms, the association is transformed into an ontological relation. The formal concept analysis (Weng, Tsai, Liu, & Hsu, 2006) was proposed with the formal method defined for representation, analysis and management of data and knowledge. The hierarchy of terms in ontology can be built by this method. The conceptual clustering (Hotho, Maedche, & Staab, 2001) was proposed to construct the ontology by grouping the concepts according to a semantic distance between each other to generate hierarchical relations. In summary, the idea of analyzing the concept structure of the specific topic from contents was proposed in these approaches. It motivates us to obtain the topic interest of experts in the learning forum community from their posting of documents.

3. *The scenario and method*

In online learning communities, the discussion activities for problem solving rely on community members' interactions. In this article, our aim is to help the questioner find the right persons to stimulate the problem solving discussions. Thus, how to find the trustworthy and presence experts on the learning communities becomes an interesting and important issue. To simplify the discussion, we focus on the user communities of the web-based learning forum in the rest of the article. The detailed descriptions of our ideas to assist the bridging of problem solving social network are given as follows.

3.1. *The social network service for inquiry-based learning*

In the course of Programming Language subject, the inquiry-based learning (Bruner, 1961) is usually applied to train learners' practical problem-solving capability. The inquiry-based learning is an effective strategy that helps learners to link the theory to the practice and develop teamwork collaborative learning skills (Simpson et al., 2008). The learning context for the students in this article is based on the implementation project of algorithm problems.

In the programming learning forum, the inquiry-based learning begins when learners identify the encountered problem. Next, the learner can logon the programming learning forum to start the inquiry by posting his/her question or problem. The senior programmers (i.e. the experts) or other learners with similar topic interest will provide feedback by replying to the learner's question. The inquiry and feedback cycle is repeatedly executed until the learner is satisfied with the solution.

To stimulate the problem solving activities in the community, the social network service of Web 2.0 with trustworthy experts finding is proposed. As shown in Figure 2, while the questioner posts a question, the main keywords of the question are first identified by the interaction with the questioner. The expert finding service will find trustworthy experts based on their topic interest with respect to the posted question.

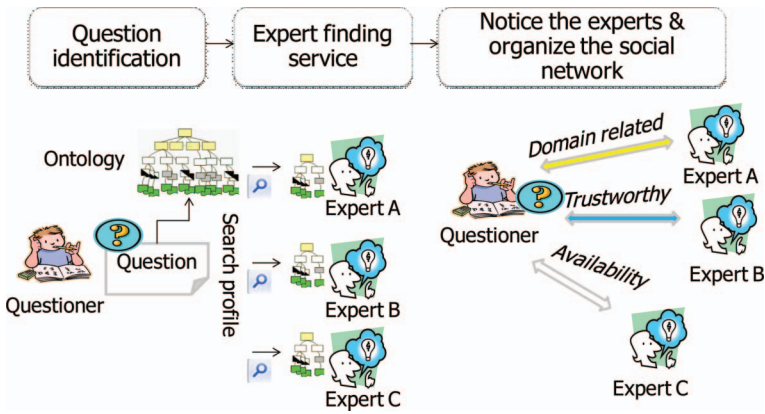


Figure 2. The trustworthy expert finding service to bridge the social network for problem solving.

Next, the questioner can configure the parameters to change the priority of the recommendation to fit their required trustworthiness and availability. The trustworthiness means that the experts may have topic interests to the posted question and have a good reputation based on their portfolio on the forum. The availability means that the experts are still present and keep visiting the forum in recent months. Thus, with the recommended experts list, the system can actively organize the social network from the questioner to these experts by inviting them to help solve the posted question on the forum.

Because the trustworthiness of the service is based on the posted forum documents, it may result in the phenomenon of ‘the more discussions you post, the more your social network can be explored’. Thus, the service on the forum can facilitate the collective intelligence and the social network of Web 2.0 to enrich the programming problem solving in the learning community.

3.2. The model of expert’s topic interest

As described earlier, the expert finding service aims to bridge the social network on the web forum. Because the domain interests of the expert should be concerned in the service, the ontology-based approach is applied to construct the consensus taxonomy for the programming learning topics. Thus, the PCO is defined to represent the problem solving topics discussed on the forum. There are four layers in PCO which are the Category Layer, Topic Layer, Issue Layer, and Document Layer. Assuming there are n predefined categories; thus the definition of PCO is given as follows.

Definition 1. The PCO is defined as $PCO = (P, T, V, D, R)$, where P , T , V , and D are concepts in different layers and R is a set of relations among concepts as shown in Figure 3.

$P = \{p_1, p_2, \dots\}$ is a finite set of category nodes in the Category Layer to represent a predefined category of the topics.

For each purpose p_i , the topics $T_i = \{p_i t_1, p_i t_2, \dots\}$ is a finite set of topic nodes in the Topic Layer to represent different topics discussed in the forum. Topic nodes with similar categories are linked below the corresponding category node by the ‘A Part Of’ relations.

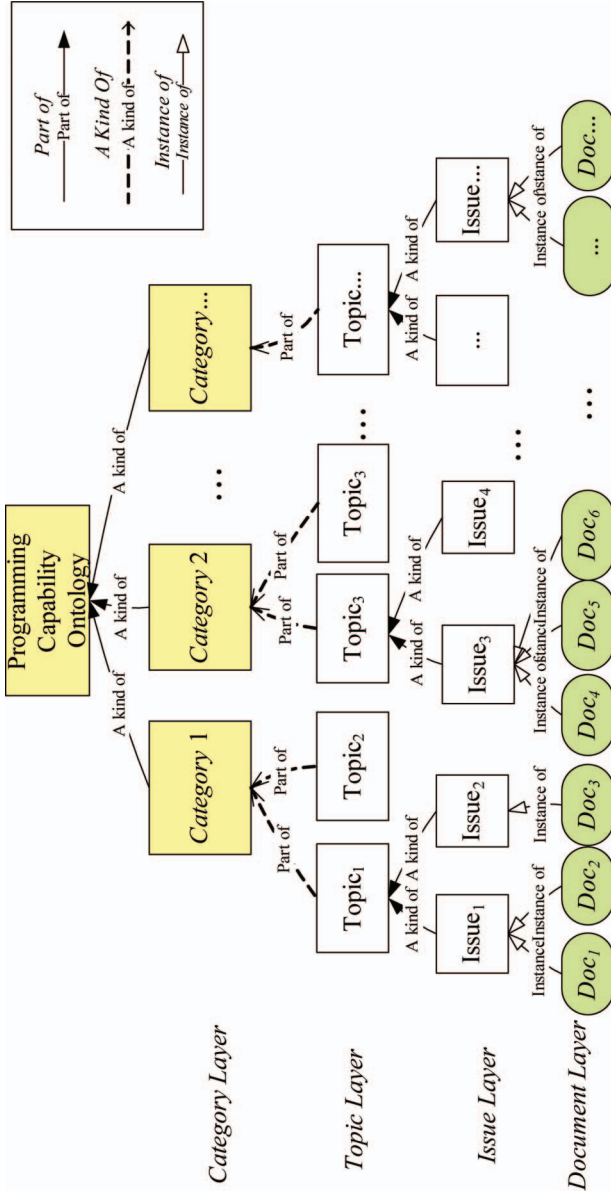


Figure 3. The four layers of programming capability ontology.

For each category p_i and topic t_j , the issues $V_{ij} = \{p_i t_j v_1, p_i t_j v_2, \dots\}$ is a finite set of issue nodes in the Issue Layer to represent the discussion keyword features of forum documents. Issue nodes with similar topics are linked below the corresponding topic node by the ‘A Kind Of’ relations.

For each category p_i , topic t_j and issue v_k , the documents $D_{ijk} = \{p_i t_j v_k d_1, p_i t_j v_k d_2, \dots\}$ is a finite set of document nodes in the Document Layer to represent the linkages associated to the original forum documents. The document nodes with similar issues are linked below the corresponding issue node by the ‘Instance Of’ relations.

With the defined ontology, different experts’ topic interests can be represented and annotated by the concepts in the issue layer. We assume that the users’ posted contents on the forum can represent their interests. Thus, the expert’s profile including topic interest, trustworthy and presence is defined. An example is shown in Figure 4.

- Topic interest: referred to the number of concepts in the issue layer of PCO, the topic interest is a vector of Boolean values where k -th element is assigned to 1 if the expert has posted the documents related to the k -th issue before.
- Trustworthy value: it is also a vector with the same length as that of topic interest to represent the reputation of the expert in the specific topic. The k -th element of trustworthy value is represented by the ratio of the number of satisfied questioners to the number of all questioners with respect to the expert’s historical replies. The larger the value, the more trustworthy the expert is.
- Presence value: it is a list of array which records the ratio of the number of online days to the number of all days in each month. The M_1 is the ratio of that in the last month; M_2 is the ratio of that 2 months ago, etc.

3.3. The trustworthy expert finding

With the defined expert profile, the aim of the expert finding service is to retrieve the relevant experts whose profiles are related to the posted question. It can be formulated as the objective indicators as follows.

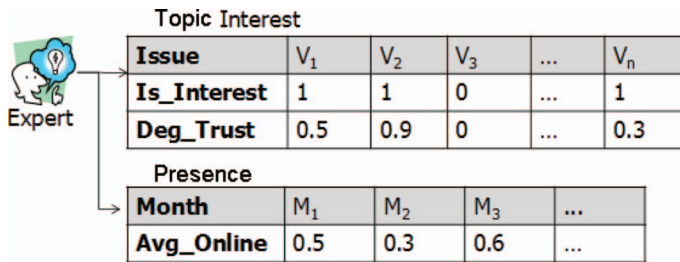


Figure 4. An example of expert’s profile representation.

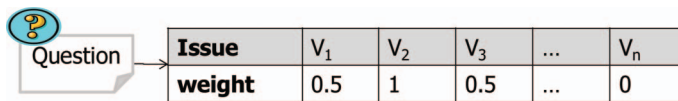


Figure 5. A posted question is transformed into keyword vector.

A *Question Q* is inputted by a questioner to express his/her programming problem with the concept weight vector as shown in Figure 5. When a learner inputs a sentence of question description, the predefined thesaurus is applied to extract the keywords appeared. Thus, the question is transformed into the keyword vector where the length of the *Q* is limited to the number of issues in PCO. Next, the weight values, from 0 (not related), 0.5 (partially related), to 1 (highly related), can be adjusted by the questioner to represent the relation degree of his/her question to the issues. In general, the keywords of similar meaning are recognized as the same concept. Because the documents in the forum are short sentences, the length of concept weight according to our experiment can be limited to the vector with less than 50 keywords.

In order to determine the degree of relevance of a query and experts, the indicators of objective function are defined. Assume that we are given a query *Q* and an expert *E*. Let *E.Interest* represent the interest vector and let *E.Trust* represent the trustworthy vector of expert's profile. Here, an objective function *Obj* for measuring the correlation between query and expert is proposed by combining the objective functions of *Obj_{Trust}* and *Obj_{Available}*.

3.3.1. Trustworthiness

The correlations of query vector *Q* with vectors *E.Interest* and *E.Trust*, respectively, are firstly calculated by the inner product represented as $Q \bullet E.Interest$ and $Q \bullet E.Trust$ each of which represents the similarity of two vectors. Thus, the trustworthiness value is measured by the weighted sum of two inner products with the α factor to control the importance of weighting between trustworthy or topic interest. The objective function of trustworthiness is defined in Equation (1).

$$Obj_{Trust}(Q, E) = \alpha \times (Q \bullet E.Interest) + (1 - \alpha) \times (Q \bullet E.Trust) \quad (1)$$

where the factor α , $0 < \alpha < 1$, is used to control the importance weighting between trustworthy or topic interest.

3.3.2. Availability

To reduce the problem of asynchronous, the existing experts can be invited to join the problem solving discussion with higher priority. Thus, the availability parameter is included in the objective function. The objective function of availability is measured by the weighted average of presence records in the expert's profile. Assume there are *N* records in the presence array and the *E.M_i* represents the *i*-th element in the array, the objective function is defined in Equation (2). Because the availability should be judged with recent period of time, the factor τ is proposed to annotate the fading of the behavior influence based on the probability pheromone update of Ant algorithm (Dorigo & Stützle, 2004).

$$Obj_{Available}(E) = \frac{\sum_{i=1}^N \tau^i E.M_i}{\sum_{i=1}^N \tau^i}, \quad (2)$$

where the τ , $0 < \tau < 1$, is the factor to reflect the fading of expert's behavior influence.

Therefore, the range of these two objective terms, Obj_{Trust} and $Obj_{Available}$, are both in $[0, 1]$. The objective measurement Obj for question Q and Expert E which is a linear combination of Obj_{Trust} and $Obj_{Available}$ is defined in Equation (3).

$$Obj(Q, E) = \beta \times Obj_{Trust}(Q, E) + (1 - \beta) \times Obj_{Available}(E), \quad (3)$$

where the factor β , $0 \leq \beta \leq 1$, is used to control the weight between trustworthiness and presence.

On the basis of the definition of objective measurement function, there are several heuristic strategies for the questioner to choose.

- Trustworthy experts first (e.g. set $\alpha = 0.2$, $\beta = 1$): recommend the experts who are highly related to the question and have a high reputation to help solve the posted question. It can be used for the difficult problem solving topics, such as program debugging, how to implement a new application, etc.
- Similar topic interest experts first (e.g. set $\alpha = 1$, $\beta = 0.5$): recommend the experts who actively reply to the related questions to help solve the posted question. It can be used for finding learning partners to discuss the topic, such as how to configure the developing platform, how to use the specific function or modules, etc.
- Expert's availability first (e.g. set $\alpha = 0.8$, $\beta = 0.2$): recommend the active users to reply to their opinions. It can be used for the need of quick feedback, such as the comparison of different SDK, opinion sharing for new technology, etc.

4. The construction of the programming capability ontology

As described earlier, the topic interest of experts are obtained from their posting documents on the forum. Because the inquiry documents in the learning forum may be related to several categories and topics, it raises the synonym issue. To solve the issue, the ontology is constructed as the consensus of taxonomy.

Because the forum documents are incrementally posted to the forum, the construction and maintenance of ontology become important issues. Therefore, the Self-Organized Ontology Maintenance Scheme is proposed to avoid the rerunning of the whole construction process whereas new documents are posted. The detailed description of the construction and maintenance of ontology is given as follows.

4.1. The category and topic of programming capability ontology

- In the category layer of PCO, the different programming languages or SDKs can be used to classify the topic interests into different categories. Next, in the topic layer, different question types can be used to classify the purposes of problem. In our observation, the forum document usually consists of the question words to represent the purpose of question. According to researches about question analysis (Wang, Wu, Liang, & Chang, 2005), most of the question patterns can be represented as *question word + topic keywords*, where the *question word* is one of the interrogatives (e.g. What, How, and Why) and

the *topic keywords* represents the keywords in the subsequent chunks that tend to reflect the intended answer more precisely. In the programming learning forum, there are different inquiry types.

- The inquiries of programming learning may concern the usage of the programming tool, the syntactic level problem, the semantic level problem analysis, the execution time trouble shooting, etc. On the basis of the frequently asked questions in the legacy C++ learning forum, the topics of documents can be categorized into different question types as shown in Table 1.

4.2. The topic clustering to update the programming capability ontology

In general, the information retrieval models would consult the existing thesaurus to obtain the keywords. Because the taxonomy used in this article is specialized to the C/C++ programming learning topic, we build our own thesaurus of C/C++ topic from the taxonomy of the textbooks, technical documents of the compiler, online resources such as MSDN website, etc. Next, we apply the *Term Frequency – Inverse Document Frequency* (TF-IDF) weighting scheme (Avancini, Lavelli, Magnini, Sebastiani, & Zanolini, 2003; Debole & Sebastiani, 2003; Wang, Lei, Cheng, & Tseng, 2003) to represent the main issue of each document. Each document can be represented by a vector $\langle tf_1 \times idf_1, tf_2 \times idf_2, \dots, tf_n \times idf_n \rangle$, where tf_i is the frequency of the i -th term, $idf_i = \log(n/df(i))$ is the IDF of the i -th term in the document, n is total number of documents and $df(i)$ is the number of documents that contains the term.

After transforming the forum documents into keyword vectors, our aim is to obtain the topics and issues through clustering analysis. With the predefined category layer and topic layer in PCO, the distributed clustering algorithm is then

Table 1. Different question types of problem solving topics.

Question types	Description	Example issues
What's the meaning	Questions about the definition of function library	Concept about template and data member revision; Concept about static object, etc.
What's wrong	Questions about what's wrong with the bug or specific programming error	Problem about free and delete from memory; Why cannot it pass-by-reference, etc.
What's different	Questions about what's the difference between two or more domain specific concepts	Conflict about dynamic class creation and overloading; differences between structure in C and Class in C++, etc.
How to do	Questions about how to implement the required functionality	How to use constructor in class; how to initiate the array in construct, etc.
How to use	Questions about how to use the function library or program statements	How to compile the class in another directory; how to use winsock.h in dev C++, etc.
Other experience sharing	Other discussion topics such as quiz, experience sharing, etc.	Best practices of OOP; bibles of C++, etc.

applied to obtain the discussion topics forum documents. Firstly, the keywords of the concepts in category layer and topic layer are used to classify the documents into parts of the predefined categories by the interrogative's patterns. Secondly, for each part, the related discussion concepts can be discovered by clustered documents with similar keyword vectors. The similarity of keyword vectors can be calculated by the inner product space measurement. Because the number of issues is unknown so far, the ISODATA clustering algorithm (Ball & Hall, 1965) is used here, which can adaptively divide and merge the clusters to find the most suitable cluster number for the given data distribution. The distributed clustering algorithm is proposed as follows.

Algorithm 1: The distributed clustering algorithm

Input: Keyword vectors of forum documents, PCO

Output: New version PCO

- Step 1.** Initially, the concepts in topic layer of PCO are used as different classifications and the associated keywords are used to construct the classifier.
- Step 2.** Classify the new posted document into different topics. For each topic, compute the similarity of the documents and the most similar issue node in PCO using keyword-based inner product. Add the 'instance_of' relation from document node to the issue node. Repeat the Step 2 until all documents are inserted into the ontology.
- Step 3.** For each altered part, apply the ISODATA algorithm to re-cluster the documents into new clusters. Update the concepts in issue layer of PCO by cluster centers. Update the keywords vector of each topic by averaging keyword vectors of new issues in PCO.
- Step 4.** Update and output new version PCO.

There are two reasons for performing the distributed clustering. The first is to implement our heuristic of document analyzing and the second is to efficiently reduce the computation complexity. To implement our heuristic of 'classification first, and then clustering' which is usually applied in the analysis of Q&A systems, the distributed clustering algorithm is proposed with cascading of classifier and clustering analysis module. Besides, the ISODATA algorithm is sensitive to the data size in clustering. To reduce the computation complexity, the distributed clustering algorithm can efficiently reduce the data size in clustering analysis.

Currently, the constructed thesaurus in this research is customized for specific forums. Although the thesaurus may need to be extended to cover more possible vocabulary of different communities if more forums are considered, the proposed methodology can be reused in different forum.

4.3. The self-organized ontology maintenance scheme

With the distributed clustering algorithm, the PCO ontology and experts' profile can be incrementally maintained as shown in Figure 6.

There are four processes in the self-organized ontology maintenance scheme. Although new documents are inserted, the document category classification process initially classifies each document into one of the categories and inserts it into the

document layer of PCO. Next, the documents of the altered parts are reclustered by the distributed clustering algorithm. Because only specific parts of the concepts in the topic layer and issue layer are updated, only corresponding parts of the expert's topic interest in profile need to be renewed.

5. Experiments

In this section, the experiment on the programming learning forum 'Programmer-Club' is presented. In the beginning, the contents of the learning forum are extracted and analyzed to construct the PCO. Next, the prototype of expert finding service with the learning forum is provided. The evaluation of feasibility and effectiveness of the proposed methodology are discussed later on.

5.1. The feasibility evaluation

5.1.1. Training set for ontology construction

The data of programming learning forum 'Programmer-Club' consisting of about 14,000 forum documents and 1734 user accounts are collected from year 2001 to 2007 as the test data. The characteristics of the forum test collection are listed in Table 2.

5.1.2. Sample questions

To compute the precision of the proposed approach in different questions, four frequently asked hot topics which are issues of 'Q₁: the object-oriented programming', 'Q₂: the string processing', 'Q₃: the array processing', and 'Q₄: the loop statements' are collected as sample questions.

5.1.3. Expert finding service configurations

Three expert finding strategies with different configurations of parameter values are listed in Table 3.

For each question proposed earlier, the precision of retrieved top-*k* experts is evaluated. In this way, the *precision* measure is judged by the human experts who are

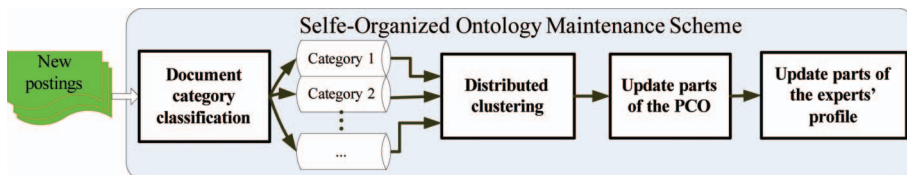


Figure 6. The self-organized ontology maintenance scheme.

Table 2. Characteristics of the test forum documents database.

Forum name	No. of postings	No. of community members	Subject
Programmer-club	14,183	1734	C/C++ programming

instructors of programming language course in universities. The precision is defined in Equation (4).

$$\text{Precision} = \frac{N_Acceptable_Expert}{N_Retrieved}, \tag{4}$$

where $N_Retrieved$ is the number of recommended experts and $N_Acceptable_Expert$ is the number of acceptable experts judged by the human experts.

Therefore, with the test data mentioned earlier, the objective values of different experts are ranked and the top 20 of them are retrieved. The precision measures are shown in Figure 7. The Q_1 – Q_4 in x -axis represents different questions, and the data in y -axis represent the precision value. For each question, the measurements of three different expert finding strategies are shown.

As shown in Figure 7, we found that the precision values obtained by *Trustworthy experts first* strategy and *Experts presence first* are relatively low. With further observation held later on, the documents of Q_2 and Q_4 lacked sufficient number of trustworthy values in our training data, even though the average precision values are higher than 50%. In summary, it may be concluded that the proposed expert finding service is feasible in general, where the similar topic interest experts first strategy can have the highest feasibility.

5.2. The effectiveness evaluation

In addition to the feasibility evaluation, the effectiveness of the proposed social network services is investigated. The inquiry-based learning process is based on the existing web forum and learning community treated as learning context. The prototype of the social network service is provided as the add-on functionality to recommend the trustworthy experts based on learner’s question. The experiment was

Table 3. Parameter values of the three experts finding strategies.

Experts finding strategy	Topic interest: α	Trustworthy: β
S ₁ . Trustworthy experts first	Low ($\alpha = 0.2$)	High ($\beta = 1$)
S ₂ . Similar topic interest experts first	High ($\alpha = 1$)	Median ($\beta = 0.5$)
S ₃ . Experts availability first	High ($\alpha = 0.8$)	Low ($\beta = 0.2$)

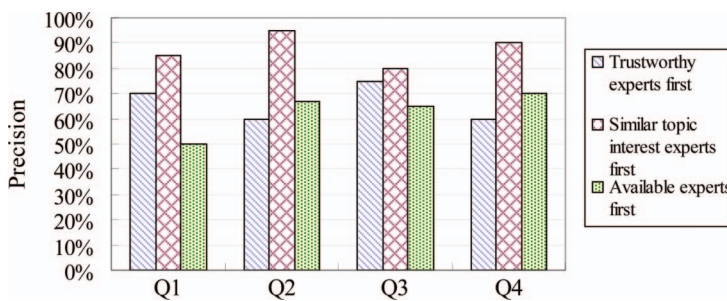


Figure 7. Precision measure for the three expert finding strategies.

held by involving 21 university students who are majoring in computer science, all with programming experience, and who participated in the evaluation. The questionnaire analysis is applied to evaluate the students' satisfaction degree of the provided services in different inquiry problems and in different expert finding strategies as shown in Tables 4 and 5, respectively. The items are measured by the 5-point Likert scale ranging from 5, 'strongly agree' to 1, 'strongly disagree'. The mean and standard deviation of the questionnaire statistical results are shown as follows.

As shown in Table 4, the mean of Q₁ item is larger than 4.0. Thus, the expert finding service is helpful in general. The items from Q₂ to Q₅ show the satisfaction value of different inquiry problems discussed on the forum. The highest mean value occurred in Q₅. It shows that the inquiry with experts is most helpful in 'new experience sharing'. The mean value of Q₄ is relatively lower than others. With the further feedbacks from learners for the problems of 'how to use', some of them would like to read the technical documents by themselves rather than asking from social interactions. The mean values of items Q₂ and Q₃ are higher than 4.0 which represent the helpfulness of the services.

As shown in Table 5, questionnaire items of learners' satisfaction in different expert finding services have been investigated. The item Q₁₀ was asked in opposite ways compared with others. In average, from the mean value of Q₆, Q₇, Q₈, and Q₉, the satisfaction evaluations of proposed services are larger than 3.0 which means acceptable. Among them, the item Q₈: 'similar topic interest experts finding service'

Table 4. Questionnaire of learners' satisfaction in different inquiry problems.

Questionnaire item	Mean	SD
Q ₁ . I think the inquiry-based learning with experts on the forum is helpful for the programming problem solving	4.05	0.80
Q ₂ . I think the inquiry is especially helpful in the problems of 'what's the meaning' or 'what's the different'	4.10	1.00
Q ₃ . I think the inquiry is especially helpful in the problems of 'what's wrong'	4.19	0.98
Q ₄ . I think the inquiry is especially helpful in the problems of 'how to use' or 'how to do'	3.95	0.86
Q ₅ . I think the inquiry is especially helpful in the discussions of 'new experience sharing'	4.33	0.66

Table 5. Questionnaire of learners' satisfaction in different expert finding strategies.

Questionnaire item	Mean	SD
Q ₆ . I think the 'trustworthy experts finding service' is helpful for my learning	3.95	0.80
Q ₇ . I think the 'availability experts finding service' is helpful for my learning	4.05	0.74
Q ₈ . I think the 'similar topic interest experts finding service' is helpful for my learning	4.24	0.70
Q ₉ . I think the 'automatic social networking service' is helpful for my learning	3.62	0.67
Q ₁₀ . The 'automatic discussion invitation services' of problems from other learners do disturb me	2.67	0.86

got the highest value. In addition, some interesting feedback of how to further improve the service was provided.

One of the learner's feedback said that the categories and topics can be more customized for their learning subjects. Thus, it can be easier for them to ask the right question to find the right experts.

In summary, the experiment result shows that most of the learners agreed that the proposed social network service is helpful for their learning. Especially, the 'similar topic interest experts finding service' is the most useful among them. Thus, it can be concluded that the effectiveness of the proposed approach is satisfactory.

6. Conclusion

In this article, the inquiry-based learning is applied to assist students' programming learning for problem solving on Web forum. The social network service of Web 2.0 with trustworthy experts finding service has been proposed to assist students to easily find the right experts to inquire about the solution. The issues of how to find the trustworthy experts for improving the quality of social interactions are investigated. Our ideas of finding experts based on their topic interest, trustworthiness, and availability have been proposed. The topic interest is annotated based on the expert's postings on the forum. The trustworthiness is calculated by accumulating the rating feedbacks from other community members. The availability is obtained by the frequency of their presence online. Accordingly, the objective function of trustworthiness and availability are formulated to easily support the expert finding service with more flexible configuration. Thus, three expert finding strategies including trustworthiness first, topic interest first, or availability first are further proposed based on different configurations to meet students' requirements. With regard to the issue of constructing the topic interest of experts and avoiding the synonym problem, initially the PCO is constructed by the distributed clustering algorithm. Next, the self-organized ontology maintenance scheme is proposed to maintain the ontology efficiently. To evaluate the proposed approach, some experiments have been done to show the feasibility and effectiveness of expert finding service based on predefined test cases and learners' satisfaction questionnaire analysis. Finally, the experimental result shows that the feasibility and the effectiveness of the proposed approach are satisfactory.

Currently, three parameters are concerned in the trustworthy experts finding service. The new referable parameters could be extended according to the extra functionality provided and required in different community platforms. In the future, the automatically fine tune mechanism for the expert finding services based on the students' feedback may be further investigated to provide better service.

Acknowledgments

This work was partially supported by National Science Council of the Republic of China under contracts NSC 96-2752-E-009-006-PAE, NSC95-2520-S009-007-MY3, and NSC95-2520-S009-008-MY3.

Notes on contributors

Shian-Shyong Tseng received a PhD in computer engineering from the National Chiao Tung University in 1984. From 1983 to 2009, he has been on the faculty of the Department of Computer and Information Science at National Chiao-Tung University. From 1992 to 1996,

he was the Director of the Computer Center at Ministry of Education and the Chairman of Taiwan Academic Network (TANet) management committee. In December 1999, he founded Taiwan Network Information Center (TWNIC). He is currently the Chairman of the board of directors of TWNIC. He is also the vice president of Asia University. His current research interests include expert systems, data mining, computer-assisted learning, and Internet-based applications. He has published more than 100 journal papers.

Jui-Feng Weng received an MS degree from the Department of Computer and Information Science, National Chiao Tung University in 2002. Currently, he is the PhD candidate at the Institute of Computer Science, National Chiao Tung University, Taiwan. His current research interests include e-learning platforms, game-based learning, knowledge engineering, expert systems, and data mining.

References

- Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207.
- Aragon, S.R. (2003). Creating social presence in online environments. In S.R. Aragon (Ed.), *Facilitating learning in online environments* (pp. 57–68). San Francisco: Jossey-Bass.
- Avancini, H., Lavelli, A., Magnini, B., Sebastiani, F., & Zanoli, R. (2003). Expanding domain-specific lexicons by term categorization. *Proceedings of ACM symposium on applied computing* (pp. 793–797). New York: ACM.
- Ball, G.H., & Hall, D.J. (1965). *ISODATA, a novel method of data analysis and pattern classification*. Menlo Park, CA: Stanford Research Institute.
- Bhalerao, A., & Ward, A. (2001). Towards electronically assisted peer assessment: A case study. *Association for Learning Technology Journal*, 9(1), 26–37.
- Bratitsis, T., & Dimitracopoulou, A. (2005). *Data recording and usage interaction analysis in asynchronous discussions: The D.I.A.S. system*. Paper presented at the AIED Workshop (AIED 05). Amsterdam, The Netherlands.
- Bruner, J.S. (1961). The act of discovery. *Harvard Educational Review*, 31(1), 21–32.
- Bulu, S.T., & Yildirim, Z. (2008). Communication behaviors and trust in collaborative online teams. *Educational Technology and Society*, 11(1), 132–147.
- Caspi, A., Gorsky, P., & Chajut, E. (2003). The influence of group size on non-mandatory asynchronous instructional discussion groups. *The Internet and Higher Education*, 6(3), 227–240.
- Cazden, C.B., & Beck, S.W. (2003). Classroom discourse. In A.C. Graesser, M.A. Gernsbacher, & S.R. Goldman (Eds.), *Handbook of discourse processes* (pp. 165–197). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chen, J.W. (2005). *Designing a web-based van Hiele model for teaching and learning, computer programming to promote collaborative learning*. Paper presented at the fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), Kaohsiung, Taiwan.
- Chen, W.K., & Cheng, Y.C. (2007). Teaching object-oriented programming laboratory with computer game programming. *IEEE Transactions on Education*, 50(3), 197–203.
- Clark, B., Rosenberg, J., Smith, T., Steiner, S., Wallace, S., & Orr, G. (2007). Game development courses in the computer science curriculum. *Journal of Computing Sciences in Colleges*, 23(2), 65–66.
- Crews, T. Jr., & Ziegler, U. (1998). *The flowchart interpreter for introductory programming courses*. Paper presented at the Frontiers in Education Conference. FIE'98 28th Annual, Tempe, Arizona.
- Debole, F., & Sebastiani, F. (2003). Supervised term weighting for automated text categorization. *Proceedings of ACM Symposium on Applied Computing* (pp. 784–788). New York: ACM.
- Dengler, M. (2008). Classroom active learning complemented by an online discussion forum to teach sustainability. *Journal of Geography in Higher Education*, 32(3), 481–494.
- Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*. Cambridge, MA: MIT Press.
- Garner, S. (2003). *Learning resources and tools to aid novice learners learn programming*. Paper presented at the Informing Science & Information Technology Education (InSITE 2003) Joint Conference, Pori, Finland.

- Hotho, A., Maedche, A., & Staab, S. (2001). *Ontology-based text clustering*. Paper presented at the IJCAI-2001 workshop Text learning: Beyond supervision, Seattle, WA.
- Hou, H.T., Chang, K.E., & Sung, Y.T. (2008). Analysis of problem-solving-based online asynchronous discussion pattern. *Educational Technology & Society*, 11(1), 17–28.
- Hu, B.Y., & Yang, J.T. (2005). *Analyzing critical thinking and factors influencing interactions in online discussion forum*. Paper presented at the fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), Kaohsiung, Taiwan.
- Khan, L., & Luo, F. (2002). Ontology construction for information selection. *Proceedings of the 14th IEEE international conference on tools with artificial intelligence* (pp. 122–127). Washington, DC: IEEE.
- Kumar, A.N. (2003). *A reified interface for a tutor on program debugging*. Paper presented at the 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03), Athens, Greece.
- Lin, S., Liu, E., & Yuan, S. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17, 420–432.
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), 72–79.
- Moreno, A., Myller, N., & Sutinen, E. (2004). JeCo, a collaborative learning tool for programming. *Proceedings of the IEEE Symposium on Visual Languages and Human Centric Computing* (pp. 261–263). Washington, DC: IEEE Computing Society.
- Newberry, B. (2001). Raising student social presence in online classes. In W. Fowler & J. Hasebrook (Eds.), *Proceedings of WebNet 2001 World Conference on the WWW and the Internet* (pp. 905–910). Norfolk, VA: Association for the Advancement of Computing in Education.
- O'Reilly, T. (2005). *What is Web 2.0*. O'Reilly network. Retrieved November 6, 2008, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- Preston, D. (2005). PAIR programming as a model of collaborative learning: A review of the research. *Journal of Computing Sciences in Colleges*, 20(4), 39–45.
- Ramadhan, H.A. (1997). Improving the engineering of model tracing based intelligent program diagnosis. *IEE Proceedings-Software*, 144(3), 149–161.
- Rourke, L., Anderson, T., Garrison, D.R., & Archer, W. (2001). Assessing social presence in asynchronous text-based computer. *Journal of Distance Education*, 14(2), 51–70.
- Ryoo, J., Fonseca, F., & Janzen, D.S. (2008). *Teaching object-oriented software engineering through problem-based learning in the context of game design*. Paper presented at the 21st Conference on Software Engineering Education and Training (CSEET '08), South Carolina, USA.
- Simpson, A., Reynolds, L., Light, I., & Attenborough, J. (2008). Talking with the experts: Evaluation of an online discussion forum involving mental health service users in the education of mental health nursing students. *Nurse Education Today*, 28(5), 633–640.
- Sithiworachart, J., & Joy, M. (2004). *Effective peer assessment for learning computer programming*. Paper presented at the 9th annual conference on Innovation and Technology in Computer Science Education, Leeds, UK.
- Tu, C.H. (2001). How Chinese perceive social presence: An examination of interaction in an online learning environment. *Educational Media International*, 38(1), 45–60.
- Tung, F.W., & Deng, Y.S. (2006). Designing social presence in e-learning environments: Testing the effect of interactivity on children. *Interactive Learning Environments*, 14(3), 251–264.
- Wang, C.Y., Lei, Y.C., Cheng, P.C., & Tseng, S.S. (2003). *A level-wise clustering algorithm on structured documents*. Paper presented at the NCS2003, Taiwan.
- Wasko, M.M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1), 35–57.
- Weng, S.S., Tsai, H.J., Liu, S.C., & Hsu, C.H. (2006). Ontology construction for information classification. *Expert Systems with Applications*, 31, 1–12.
- Yang, S.J.H., Chen, I.Y.L., Kinshuk, & Chen, N.-S. (2007). Enhancing the quality of e-learning in virtual learning communities by finding quality learning content and trustworthy collaborators. *Educational Technology and Society*, 10(2), 84–95.
- Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science*, 34(6), 451–480.