An Ontology-based Framework for Itembank Integration and Knowledge Sharing

Shih-Pang Tseng, Ming-Chao Chiang, Chu-Sing Yang, and Chun-Wei Tsai

Abstract

The Internet is abounding with information, but there is simply no easy way to find all the information we need in a single place, well organized and easily accessible, let alone share. For this reason, we present a novel framework for resolving these issues. The proposed framework incorporates and unifies techniques from a variety of disciplines such as information extraction, information retrieval, fuzzy clustering, and ontology. To evaluate the performance of the proposed framework, we use it to solve the itembank integration and knowledge sharing problem faced by many e-Learning systems. Our experimental results show that the proposed framework can not only effectively integrate information available in the Internet, but it can also properly organize the information into a unified itembank of an e-Learning system, thus making it possible for the users of such a system to not just find but share all the information.

Keywords: Information Retrieval, Ontology, Itembank.

1. Introduction

Over the past decades, the increase in the computing power of computers and the advance in the internet technology have been phenomenal. As such, a tremendous amount of traditional printed material has been transformed into digital material. One recent study [1] found that over 90% of the information in use today was generated in digital form. This fundamentally changes the way information is disseminated as well as the way knowledge is acquired.

Researchers [2], [3] even argue that the spatial and temporal limitations of information will be broken when the era of the Internet comes. As a consequence, we, the human kind, will have abundant of information available for use any time and any where. The reality, however, is that the Internet is abounding with information, but even with the help of search engines, there is simply no easy way to find all the information we need in a single place, well organized and easily accessible, let alone share. For this reason, how the information we need is found and integrated and how the information we found and integrated is shared and reused have become probably two of the most important issues to be resolved, especially in the era of the Internet. In other words, we need an efficient method to find, integrate, share, and reuse all the information available in the Internet.

The purpose of this paper is thus to present such a method to resolve these issues. However, instead of a single method, we go one step further and present an ontology-based framework. To make the proposed framework more concrete, we will use the itembank integration and knowledge sharing problem faced by many e-Learning systems as an example for a very simple reason: Itembanks are the soul of an e-Learning system, but it is generally very difficult to integrate all the itembanks, even on the same e-Learning system, because the content structures of different itembanks are usually different. Unfortunately, this is more often than not a norm because itembanks are by and large created from a variety of sources. This is easily justified by the following two observations.[1]

1. *Resource integration*: In general, it is very difficult to find and integrate all the resources available in the Internet because the Internet is made up of all kinds of resources and the structures the resources use may be quite different. For example, the content structure of the web site *a* may be quite different from that of the web site *b*. If we consider each web site as a database, this is similar in principle to why something like Open Database Connectivity (ODBC) is needed to access data from different databases. However, integrating resources available in the Internet is apparently much more difficult than integrating databases using ODBC because the content structures of resources available in the Internet are even more versatile.

Corresponding Author: Ming-Chao Chiang is with the Department of Computer Science and Engineering, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan.
E-mail: mcchiang@cse.nsysu.edu.tw
Manuscript received 15 Nov. 2009; revised 19 Mar. 2010; accepted 25 Mar. 2010.

[1]Throughout the rest of the paper, since there is no confusion, we will use the terms *information*, *content*, *resource*, and *item* interchangeably to mean the same thing.

2. *Content sharing*: A problem that is relevant to resource integration is how the resources are shared and reused once they are found and integrated. Recently, more and more of the systems or applications in the Internet begin to provide options to output their contents in the XML format because XML provides a common way for exchanging or sharing the contents. However, most of them are incapable of classifying the contents (or items in the context of an e-Learning system) they have into similar groups (e.g., chapters or sections, again in the context of an e-Learning system). Thus, we are again faced with the same problem; that is, the content structures are different, thus making it extremely difficult to share the contents. As such, many researches [4] have recently started using ontology to define, describe, and preserve the content.

The observations above show that the content integration and sharing have become a problem faced by the users of the Internet. On one hand, if we are able to solve the content integration problem, all the e-Learning systems can benefit from such an integration because such an integration can provide the users of an e-Learning system more contents than ever before. On the other hand, if the content we have can be easily shared, then the users of the Internet will benefit from such a sharing. Thus, if we can solve these two problems, we can eventually construct a much more flexible e-Learning environment on which resources can be easily found, integrated, shared, and reused.

Now, the question is, how do we *define, extract, integrate,* and *share* the contents? The proposed framework—which we refer to as an ontology-based framework for itembank integration and knowledge sharing (or IIKS for short) — is thus aimed at extracting and integrating as many kinds of resources as there are parsers into a unified itembank of whatever e-Learning systems in question. However, as far as this paper is concerned, instead of an abstract e-Learning system, we will use HGLS [5], an e-Learning system we develop, as a testbed of the proposed framework, which incorporates and unifies techniques from a variety of disciplines such as information extraction, information retrieval, and ontology to accomplish these tasks. The information extraction technology is used to extract resources from the Internet; the information retrieval technology is used to integrate these resources; and the ontology technology is used to define and share the resources.

The remainder of the paper is organized as follows. Section 2 gives a brief introduction to the development, the learning theory, and the technologies of e-Learning systems. The problem we consider in this paper is defined in Section 3.1. Section 3 gives a detailed description of the proposed framework, especially how it is designed and implemented. Simulation and experimental results are resented in Section 4. Conclusion and future work are given in Section 5.

2. Related Work

In this section, we begin with a list of notations, followed by an introduction to the development of e-Learning and the learning theory for enhancing the performance of e-Learning. Then, we present several technologies that can be used to help the instructor or manager of an e-Learning system analyze the various factors that can influence the performance of learning.

A. *Notations*

To simplify our discussion that follows, throughout the rest of this paper except where no confusion is possible, the following notations will be used.

I	The set of input itembanks. That is, $I = \{I_1, I_2, \ldots, I_m\}$, where m is the number of itembanks.
I_i	The i-th input itembank (also known as item set or data set).
I_r	The reference itembank, i.e., the initial unified itembank into which all the other itembanks will be integrated.
T	The unified itembank, i.e., the itembank of itembanks.
U	The membership partition matrix $U_{c \times n} = [u_{ij}]$ where $u_{ij} \in [0, 1]$ represents the membership of the j-th pattern in the i-th cluster. Moreover, for all $j, \sum_{i=1}^{c} u_{ij} = 1$.
M	The set of means. That is, $M = \{m_1, m_2, \ldots, m_c\}$.
c	The number of groups (chapters or sections).
n	The number of patterns.
d_{ij}	The distance between the j-th item and i-th cluster.
α	A parameter of fuzzy c-means whose value is generally greater than 1. Usually, 2 is used.

B. *The e-Learning*

Many researches on the initial stage in the development of e-Learning have focused on changing the learning behavior of human beings from traditional paper and classroom to e-Learning. Many have focused on designing a more efficient method for managing the learning objects such as Learning Management System (LMS). For instance, Ismail [6] pointed out that the focus of the first-generation e-Learning systems is on the management and measurement of the training processes. Recently, the courseware rescue and adaptive education system have become the emerging researches on e-Learning [7]. Of course, in the ideal case, the learning objects or course contents should be able to be moved and reused among these e-Learning systems. However, the fact is that these learning objects and course contents can not be moved and reused as easily as we would expect.

The education theories that can be applied to an e-Learning system fall into the following three categories:

- *Exam System*: The traditional theories on the test or exam such as IRT [8] are one kind of the researches on the e-Learning exam system. How the itembanks are managed or integrated is, however, another research issue on e-Learning systems [9].
- *Learning Content*: How to design [10] reusable learning contents (reusability) and how to share the learning contents (interoperability) are probably two of the most important issues in the research on the learning contents.
- *Learning Behavior*: Several studies [11] pointed out that the online learning environment could even be better than the classroom. This is because the e-Learning environment provides us a much easier way to record all the learning path and behavior of students. As a result, it becomes much easier to analyze and review the learning results.

C. *Fuzzy Theory for e-Learning*

Web mining [12] is another technology that affects the performance of learning. For example, since many of the learning contents are extracted from the Internet, to use the traditional data mining technologies to analyze these web documents, we need to change the way the similarity of the web documents is measured. Among the researches on data clustering and e-Learning, the fuzzy theory is obviously one of the most important. It can be divided into two groups [13]. The first group includes the fuzzy c-means (FCM) and its variants. This group has been well developed and applied to many problems [14], [15] such as e-Learning, information retrieval, and image classification because FCM [16] allows the input patterns not to separate into "hard" (crisp) clusters that minimize the cost function

$$J(U,M) = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^{\alpha} d_{ij}, \quad (1)$$

where all the notations U, M, c, n, u_{ij} , and d_{ij} are as defined in Section 2.1. In [17], the authors pointed out that FCM is one of the most popular fuzzy clustering algorithms realized by Bezdek [16]. The second group includes the fuzzy methods for computing similarity. Unlike the traditional information retrieval, which uses Vector Space Model [18] or Probability Model [19] to compute the similarity between documents, the fuzzy information retrieval uses the fuzzy set theory to construct a correlation matrix between terms for the information retrieval system [20]. Due to these advantages of fuzzy-based methods, many researches have tried to apply them to the real world problems (e.g., information retrieval or e-Learning) to enhance the quality of the end result.

D. *Ontology for e-Learning*

Recently, more and more researches on e-Learning have focused on ontology because the web mining and information retrieval technologies only provide a tool for analyzing and reviewing the learning contents. They, however, take into account neither the content sharing nor the content relationships when the contents are created. Unfortunately, the definition of ontology [21], [22], [23] is not unique and even confusing. In other words, different researchers give ontology different definition, especially researchers from different areas. In general, the so-called ontology contains individual, concept, and attributes to define and describe the knowledge content. Recently, several ontology languages [24], [25], [26] have been presented, such as Resource Description Framework (RDF), XML-based Ontology Exchange Language (XOL), Ontology Markup Language (OML), and Web Ontology Language (OWL). Moreover, a number of organizations have attempted to provide software to help the researchers and users build the ontology environment, such as Protégé [27]. The real world approaches based on ontology have been steadily and continuously increased. In [28], Lehti and Fankhauser focus on the integration of XML data sources using OWL.

3. The Problem Definition and Proposed Framework

A. *The Problem Definition*

In the research of e-Learning, the heterogeneous itembank integration problem is a problem the solution of which must provide an efficient way to combine itembanks either in the same e-Learning system or from different e-Learning systems. In [9], we present an efficient method for integrating a set of heterogeneous itembanks. Although that method is capable of integrating a set of heterogeneous itembanks with a high accuracy rate, the resources are limited to the so-called items. In this paper, our focus is on expanding the contents of the itembanks. In other words, the method described herein is aimed at integrating not only the itembanks containing solely of items, be it in the same e-Learning system or from different e-Learning systems, but also the resources available in the Internet such as Wiki or blogs. That is, our focus is on providing a more suitable method to make it capable of integrating not just itembanks but also resources available in the Internet into a unified itembank of an e-Learning system.

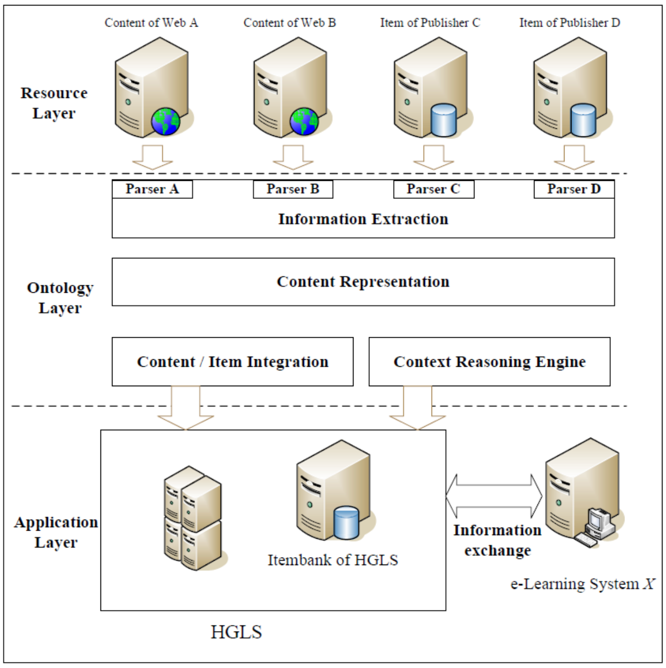


Fig. 1. The proposed framework.

In brief, the problem we consider in this paper is formulated as follows: Given a set of resources such as a set of brief introductions or descriptions $X = \{x_1, x_2, \ldots, x_n\}$ that are extracted from possibly different e-Learning systems or the Internet, find an efficient method to put each of the resources x_i into the most "appropriate" position of the unified itembank; that is, find an item t_j in the unified itembank to which the resource xi is most similar.

B. *The Proposed Framework*

In this paper, we present an ontology-based framework for itembank integration and knowledge sharing, called IIKS. IIKS can not only integrate itembanks from different e-Learning systems or publishers, but it can also integrate the learning contents available in the Internet. IIKS will create the OWL representation for each content integrated into the unified itembank. The OWL representation of such a content makes it possible to exchange the content with other systems by simply importing and exporting the content to be exchanged. In this section, we will discuss how the system is designed and implemented (see Fig. 1). As Fig. 1 shows, the proposed framework is composed of four modules: *information extraction*, *information integration*, *context representation*, and *context reasoning engine*.

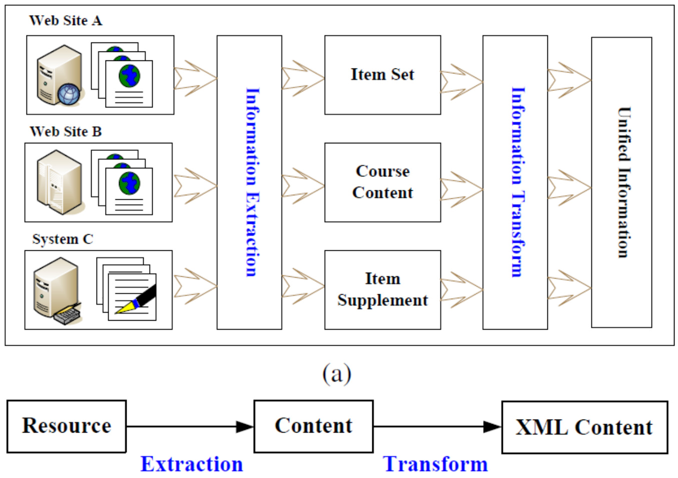


Fig. 2. (a) The Information Extraction Module. (b) The process of transforming the extracted information to the XML format.

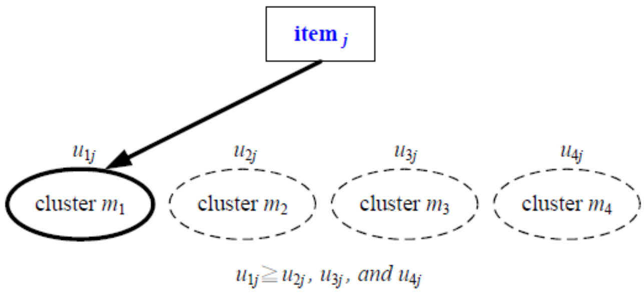


Fig. 3. An simple example illustrating how the proposed algorithm assigns each item to a cluster.

The information extraction module is responsible for extracting and parsing the learning contents from different systems such as blogs, Wikipedia, or itembanks. The information integration module takes the responsibility of combining these contents into a unified itembank using the method described in [9]. After that, the context representation module is used to transform these contents into the OWL ontologies so that they can be easily accessed by other systems. Finally, the context reasoning engine is responsible for handling predefined events. For instance, it can be used to send a notification to the manager of the itembank when a brand new item enters the system or when the unified itembank is modified. The following sections will discuss each of the modules in detail.

Fig. 4. Outline of MHIIS.

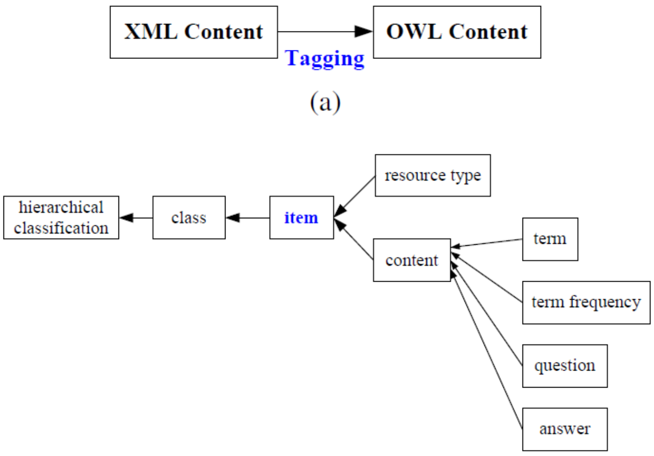
```
1   /* Input: T is the unified itembank to which the set of items t is to be integrated. */
2   /* Output: T. */
3   Procedure(T, t)
4   Begin
5     /* Integration */
6     l = 0.
7     For each item t_j in t          /* for each item */
8       For l = 0 to L-1              /* for each level */
9         Compute the similarity between item (content) t_j and the abstractions of T.
10        Use the most similar node to determine the catalog of t_j at level l and to reduce the search space.
11      End
12      Add t_j to the leaf of T.
13    End
14    /* Fuzzy Clustering */
15    For i = 1 to t_max
16      For each item t_j in t
17        Update the membership partition matrix U using Eq. (3).
18      End
19      Update the clustering means using Eq. (2).
20    End
21    For each item t_j in t
22      Find the most similar clusters to determine the catalog in T to which it belongs.
23    End
24  End
```

Information Extraction: As Fig. 2(a) shows, the information extraction module (IEM) is responsible for extracting and parsing the content from different e-Learning systems or the Internet. In other words, the IEM works as follows: first, the IEM will extract and parse the content from the Internet or e-Learning systems using whatever parsers are required for that content. For instance, the parsers we developed for extracting the content will first remove the tags and redundant content. Then, the IEM will remove the redundant information using stopword list and Stemmer's algorithm [29]. Because the structure of Chinese differs from that of English, a Chinese segmentation system [30] is generally required to segment the content in Chinese and to annotate all the terms segmented. Then, the IEM will record the title, content, description, and other information extracted. Finally, the IEM will transform the acquired content to the unified structure imposed by HGLS. Fig. 2(b) shows the process of transforming the extracted information to the XML format. That is, IEM first converts the extracted information to the XML format before storing them on the system.

Information Integration: The information integration module (IIM) described herein is built on an efficient algorithm to integrate the heterogeneous itembanks and resources, which we refer to as MHIIS in this paper. MHIIS is designed specifically for classifying items extracted from possibly heterogeneous itembanks and the Internet. In other words, the IIM receives the contents in XML from different web resources and rearranges them in a content tree. The similar contents will be put into the same group (chapter) in the content tree. The proposed system can enrich the contents in the content tree by adding contents available in the Internet and other resources to the tree. More precisely, the proposed algorithm leverages the strengths of the item integration algorithm [9] and the fuzzy c-means clustering algorithm [16]. As such, it is composed of two phases: *integration* and *fuzzy clustering*.

As indicated in Fig. 4, the input to MHIIS is the unified itembank T and the set of items t to be integrated. All the input items no matter where they came from will be put into the most appropriate place of the unified itembank T one by one. In the first phase, MHIIS computes the similarity between all the new items and the abstractions of T, on line 9. The abstractions of T refer to the Term Frequency (TF) and Inverse Document Frequency (IDF) precomputed for all the items in the itembank T to speed up the computation of the similarity between all the new items and items in T. In addition, as far as this paper is concerned, it is the Vector Space Model (VSM) [18] that is used in the computation of the similarity.

On line 10, MHIIS seeks the most suitable category at level l to which the item is to be assigned. The process will be repeated until the leaf node is found. After all the items are integrated, the new unified itembank will be taken as the input to the fuzzy clustering phase. In the second phase, MHIIS employs the fuzzy-based clustering algorithm (the fuzzy c-means, or FCM for short) to fine-tune the end result to enhance its quality, on lines 15 to 20. In general, the goal of FCM is to minimize the cost function defined by Eq. (1). The output of the inte gration phase becomes the input of the FCM. That is, the input to FCM are the set of items, the set of catalogs in T, and the number of catalogs. Then, the FCM will first update the membership partition matrix U using Eq. (3) and the clustering means using Eq. (2).



(b)

Fig. 5. (a) The transform to the OWL content. (b) Part of ontology about item.

$$m_i = \frac{\sum_{j=1}^{n_i} (u_{ij})^{\alpha} x_j}{\sum_{j=1}^{n_i} (u_{ij})^{\alpha}}, \text{for} \quad i = 1, \ldots, c, \tag{2}$$

$$u_{ij} = \frac{1}{\sum_{i=1}^{c} \left(\frac{d(m_k, x_j)}{d(m_k, x_j)} \right)^{\frac{2}{\alpha - 1}}}. \tag{3}$$

After the fuzzy clustering phase, MHIIS will re-compute the most similar clusters[2] for the set of items in t to determine the new catalogs to which they belong. Finally, the proposed algorithm uses the membership u_{ij} to determine the cluster. or catalog i to which the item t_j belongs, on lines 21 to 23. More precisely, the item t_j belongs to the cluster (catalog) i if and only if $u_{ij} \geqq u_{i'j}$, $\forall i \neq i'$. Fig. 3 gives an example to illustrate how the proposed algorithm determines which item belongs to which cluster. The example assumes that the proposed algorithm will assign the item t_j to the first cluster, denoted by its centroid m_1 because $u_{1j} \geqq u_{i'j}$ for $i' = 2, 3,$ and 4.

Context Representation: The Context Representation Module (CRM) plays the role of creating the OWL representation for all the items in the itembank of HGLS. In general, many researchers employ the Protégé [27] API to create the OWL representation, as we did in our pre vious work [31]. However, in this paper, due to the performance consideration, we first create the OWL representation of the existing items using CRM. As Fig. 5(a) shows, upon receiving the content from other systems, the CRM will automatically create the OWL representation associated with it. More precisely, CRM adds more tags to the XML content to ensure that they conform to the OWL format, as shown in Fig. 5(b). The CRM can also parse the OWL ontologies from other systems. For this reason, the CRM can be considered as a layer of software for reading and writing the OWL ontologies, though it is designed specially for HGLS. An example is given in Fig. 6 to show that HGLS is capable of exporting the items it owns to the OWL ontology, which can then be displayed by Protégé [27]. In this paper, we assume that the contents on the proposed system can be shared by the other ontology systems if they have access to the proposed system and use Protégé and the system we develop.

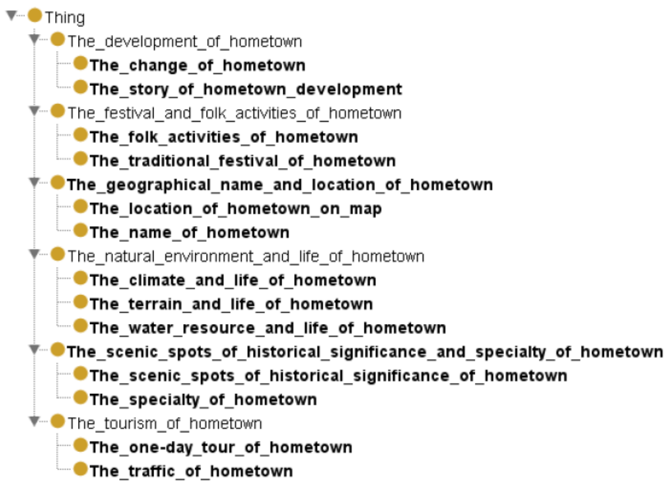


Fig. 6. An example of items of HGLS displayed using Protégé.

In addition, the proposed framework provides the functionality to export the content we want to share with others. By design, the proposed framework also allows the user or administrator of an e-Learning system to select the contents or items to be shared with others. In addition, the CRM is also responsible for exporting the items or contents in different formats such as XML, OWL, or text.

Context Reasoning Engine: Moreover, we also design a Context Reasoning Engine Module (CREM) for managing the items, contents, and supplements. For example, the CREM will update the items and contents when a new item enters HGLS or in a predefined time schedule. In summary, the CREM is event driven, and we predefine several events that will trigger the CREM as follows:

- **New items**: When a new item enters HGLS, the CREM will send it to the IIM and classifies it to the most suitable category. In addition, the CREM will also send a message to the itembank manager so that the itembank manager can keep track of all the relevant information.
- **New contents**: These are the kinds of resources extracted from the Internet. When a new content

[2] The most similar clusters are ranked in terms of the distances between t_j and all the nodes in T.

enters the IEM, the IIM will compute the similarity between the new content and the items in the unified itembank of HGLS. In addition, the users can determine if the new content is actually a new item or a supplement either by themselves or based on the suggestion given by the proposed framework. In addition, the CREM will also send the content to the manager for the bookkeeping purpose.

- **Event notification**: In this case, when the CREM detects that the unified itembank of HGLS is modified, it will send a message to the manager, again for the bookkeeping purpose.

4. Simulation and Experimental Results

In this section, we will start with an introduction to the environmental settings of HGLS. Then, we will present the experimental and simulation results and analyze the performance of the proposed framework.

A. *Environment and Parameter Settings*

All the empirical analysis was conducted on a 2.8GHz Intel Pentium4 machine with 512MB of memory running Ubuntu 8.04 with Linux kernel 2.6.24. Moreover, all the programs are written in C++, PHP, and Perl. In this paper, we use HGLS [5], an e-Learning system we develop, as a testbed of the proposed framework. HGLS has been put into use by National Sun Yat-sen University and Educational Network Center and several elementary schools in the city of Tainan, Taiwan. In addition, the system integrates several techniques such as web mining, Web 2.0, information retrieval, grid computing, and so forth.

B. *Simulation Results*

To evaluate the performance of the proposed framework, we use two different itembanks. One is from Kang Hsuan Publishing [32]; the other is from Han Lin Publishing [33]. In addition, we also use the content extracted from the Internet. More precisely, there are 697 items from Han Lin Publishing, 464 items from Kang Hsuan Publishing, contents from Wiki, and contents from blogs.[3] All the data are segmented into terms by the Chinese word segmentation system [30].

In Table 1, the accuracy rates in percentage are defined by

$$E_i = \frac{\sum_{j=1}^{n_i} E_{ij}}{n_i} \times 100\%, \quad (4)$$

$$\overline{E} = \frac{\sum_{i=1}^{m} E_i}{m}, \quad (5)$$

[3] They are available for download at http://cwtsai.ee.ncku.edu.tw/ct/benchmark.txt.

Table 1. The accuracy rate of adding all the resources given in the table into the unified itembank.

Measurement	Method	Wiki	Blog	K. H.	Web	Average
Accuracy Rate(%)		E_1	E_2	E_3	E_4	$\overline{E}$
	OHIBI	92.31	75.0	98.09	84.62	87.5
	MHIIS	92.31	91.67	98.42	92.31	93.67
SSE	OHIBI	1.92	2.28	2.73	2.43	2.34
	MHIIS	1.45	1.92	2.62	2.19	2.04

where E_i indicates the accuracy rate of resource i and $\overline{E}$ indicates the average accuracy rate of all the resources. Moreover, m denotes the number of resources, n_i denotes the number of contents resource i has, and E_{ij} takes the value of 1 and 0 to indicate if the content c_{ij} is assigned to the suitable category of the itembank, with 1 being correct and 0 being wrong. Moreover, we also use sum of squared errors (SSE) defined by

$$\text{SSE} = \frac{1}{\sum_{i=1}^{m} n_i} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\| x_{ij} - m_i \right\| \quad (6)$$

to measure the proposed framework where m_i denotes the centroid of cluster i (e.g., section of course content); x_{ij} the j-th input of cluster i.

To measure the accuracy rate of the proposed framework, we conduct several experiments and compare the proposed algorithm with OHIBI [9]. First, we use Han Lin itembank as the unified itembank to which items will be integrated and the contents of Kang Hsuan itembank as the items to be integrated to the unified itembank. Moreover, we use the tree-structured itembank to represent the contents. As Table 1 shows, the accuracy rate for adding all kinds of contents given in Table 1 into the unified itembank is 93.67% on average. For the FCM, SSE is computed by using the centroid with the highest membership to which a pattern belongs. Our simulation also shows that the proposed algorithm can provide a better result than OHIBI in terms of SSE.

C. *Experimental Results*

In this paper, we integrate the learning contents, which include itembanks from several publishers and resources from the Internet, to provide the users of HGLS a much better learning environment. Fig. 7 shows an example of the content extracted from a small town in Central Taiwan named Wuri by the IEM. We also design and implement parsers for contents from other resources and itembanks from publishers. The IIM will compute the similarity between the new input content and those in the itembank. Then, it will return the suggestions to the user to help them classify these contents into the itembank. Finally, the user can determine if this content is a new item or a supplement and assign it to the suitable position either according to the suggestions of the system or based on their own judgment.

Content	The first step to recognize Wuri is to realize the characteristics of its geographical position and administrative divisions.
http://content.edu.tw/local/taichun/yuan/h013/1/10.htm	

Fig. 7. An example of content extraction of the proposed system.

5. Conclusion

In this paper, we presented a novel framework, called an ontology-based framework for itembank integration and knowledge sharing (IIKS), for integrating resources available in the Internet into a unified itembank with a high accuracy rate. Moreover, the proposed framework uses ontology to make the learning content of HGLS easily exchanged and thus shared. The context reasoning engine is designed to provide services such as event notification to the user of such a system. As a result, the students and instructors can now have more learning contents than ever before to enhance the performance of learning. Our focus in the future is to generalize the proposed framework to make it possible to automatically create the itembank using all the information available in the Internet. We will also add more rules to the context reasoning engine to make HGLS a much better e-Learning system, especially from the user's perspective.

Acknowlegdment

The authors would like to thank the editors and the anonymous reviewers for their constructive comments and suggestions on the paper. The authors would also like to thank Kang Hsuan Publishing and Han Lin Publishing for providing their itembanks and the Chinese Knowledge and Information Processing (CKIP) group, Institute of Information Science, Academia Sinica for providing their Chinese Word Segmentation System to support this research. This work was supported in part by National Science Council, Taiwan, ROC, under Contract Nos. 95-2221-E-006-509-MY2 and 98-2218-E-006-003.

References

[1] K. Consulting, "Electronic discovery costs: Managing systems and information to control costs and improve results," *Kahn Consulting Inc.*, pp. 1-12, 2005.
[2] S. Graham, "The end of geography or the explosion of place? conceptualizing space, place and information technology," *Prog Hum Geogr*, vol. 22, no. 2, pp. 165-185, 1998.
[3] A. S. Tanenbaum, *Computer Networks, Fourth Edition*, 4th ed. Prentice Hall, 2002.
[4] L. Stojanovic and S. Staab, "e-learning based on the semantic web," *World Conference on the WWW and Internet*, 2001.
[5] HGLS. Available: http://hgls.tn.edu.tw/hgls/
[6] J. Ismail, "The design of an e-learning system: Beyond the hype," *The Internet and Higher Education*, vol. 4, no. 3-4, pp. 329-336, 2001.
[7] N. Friesen, "Three objections to learning objects and e-learning standards," in Online Education Using Learning Objects, R. McGreal, Ed. Routledge, London, pp. 59-70, 2004.
[8] H. S. Ronald K. Hambleton, *Item response theory :principles and applications*. Boston: Kluwer Academic Publisher, 1983.
[9] C. W. Tsai, S. P. Tseng, Y. S. Yang, M. C. Chiang, and C. S. Yang, "An ontology-based method for integrating heterogeneous itembanks," in *Proceedings of the International Conference on Intelligent Systems Design and Applications*, pp. 919-924, 2009.
[10] M. Angel Sicilia and E. Garc´ıa, "On the concepts of usability and reusability of learning objects," *The International Review of Research in Open and Distance Learning*, vol. 4, no. 2, 2003.
[11] M. M. Thompson, "Distance delivery of graduate-level teacher education: Beyond parity claims," *Journal of Continuing Higher Education*, vol. 44, no. 3, pp. 29-34, 1996.
[12] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer, 2007.
[13] N. Clara, "Clustering fuzzy relational data base on fuzzy cardinality," *International Journal of Fuzzy Systems*, vol. 9, no. 4, pp. 220-229, 2007.
[14] M. Mendes, E. Martinez, and L. Sacks, "Knowledge-based content navigation in e-learning applications," in *Proceedings of the LCS* 2002, 2002.
[15] C. H. Li, W. C. Huang, B. C. Kuo, and C. C. Hung, "A novel fuzzy weighted c-means method for image classification," *International Journal of Fuzzy Systems*, vol. 10, no. 3, pp. 168-173, 2008.
[16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.
[17] R. Xu and D. W. II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645-678, 2005.
[18] G. Salton, A. Wong, and C. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613-620, 1975.
[19] S. E. Robertson and K. Sparck Jones, "Relevance

weighting of search terms," *Journal of the American Society for Information Sciences*, vol. 27, no. 3, pp. 129-146, 1976.

[20] Y. Ogawa, T. Morita, and K. Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method," *Fuzzy Sets Systems*, vol. 39, no. 2, pp. 163-179, 1991.

[21] T. R. Gruber and T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, pp. 199-220, 1993.

[22] W. N. Borst, "Construction of engineering ontologies for knowledge sharing and reuse," Ph.D. dissertation, Universiteit Twente, Enschede, September 1997.

[23] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *ACM Sigmod Record*, vol. 33, no. 4, pp. 65-70, 2004.

[24] T. Berners-Lee, "Why rdf model is different from the xml model," 1999. Available: http://www.w3.org/DesignIssues/ RDF-XML.html

[25] O. Lassila and R. R. Swick, "Resource description framework (rdf) model and syntax specification," W3C Recommendation, 1999. Available: http://www.w3c.org/TR/REC-rdf-syntax

[26] W3C, "Owl web ontology language," W3C Recommendation. Available: http://www.w3.org/TR/owl-ref/

[27] The Protégé Team, "The protégé ontology editor and knowledge acquisition system." Available: http://protege.stanford.edu/

[28] P. Lehti and P. Fankhauser, "Xml data integration with owl: Experiences and challenges," *International Symposium on Applications and the Internet*, pp. 160-170, 2004.

[29] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130-137, 1980.

[30] Chinese Document Segmentation, Available: http://ckipsvr.iis.sinica.edu.tw/.

[31] W. J. Jeng, K. W. Huang, C. W. Tsai, and C. S. Yang, "An ontologybased intelligent project management system," in *Proceedings of the National Computer Symposium (NCS 2009)*, 2009.

[32] Kang Hsuan Publishing. Available: http://www.knsh.com.tw/

[33] Han Lin Publishing. Available: http://www.hle.com.tw/

Shih-Pang Tseng was born in 1968. He received the M.S. degree in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan in 1992. He is currently a Ph.D. candidate with the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan as well as an instructor of the Department of Computer Science and Information Engineering, Tajen University, Pingtung, Taiwan. His current research interests include learning technology, and evolutionary computation.

Ming-Chao Chiang received the B.S. degree in Management Science from National Chiao Tung University, Hsinchu, Taiwan in 1978 and the M.S., M.Phil., and Ph.D. degrees in Computer Science from Columbia University, New York, NY, U.S.A. in 1991, 1998, and 1998, respectively. He had over 12 years of experience in the software industry encompassing a wide variety of roles and responsibilities in both large and start-up companies before joining the faculty of the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan in 2003, where he is currently an Assistant Professor. His current research interests include image processing, evolutionary computation, and system software.

Chu-Sing Yang received the B.S. degree in Engineering Science, and the M.S. and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan in 1976, 1984 and 1987, respectively. He joined the faculty of the Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, as an Associate Professor in 1988. Since 1993, he has been a professor of the Department of Computer Science and Engineering, National Sun Yat-sen University. He was a Chair of Department of Computer Science and Engineering, National Sun Yat-sen University from August 1995 to July 1999. He was a director of Computer Center, National Sun Yat-sen University from August 1998 to October 2002. He was a Program Chair of ICS-96 and Program Co-Chair of ICPP-2003. He joined the faculty of the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, as a Professor in 2006. His research interests include multimedia communication, parallel and distributed systems, scalable web servers, and wireless communication.

Chun-Wei Tsai was born in 1978. He received the M.S. degree in Management Information System from National Pingtung University of Science & Technology, Pingtung, Taiwan in 2002, and the Ph.D. degree in Computer Science from National Sun Yat-sen University, Kaohsiung, Taiwan in 2009. He is currently a postdoctoral fellow with the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan. His research interests include web information retrieval, evolutionary computation, and combinatorial optimization.