A Knowledge-based Approach to Retrieving Teaching Materials for Context-aware Learning

Wen-Chung Shih

Department of Information Science and Applications and Department of Computer Science and Information Engineering, Asia University, Taichung 41354, Taiwan // wjshih@asia.edu.tw

Shian-Shyong Tseng[*]

Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan // sstseng@cs.nctu.edu.tw // Department of Information Science and Applications, Asia University, Taichung 41354, Taiwan // sstseng@asia.edu.tw

[*]Corresponding author

ABSTRACT

With the rapid development of wireless communication and sensor technologies, ubiquitous learning has become a promising solution to educational problems. In context-aware ubiquitous learning environments, it is required that learning content is retrieved according to environmental contexts, such as learners' location. Also, a learning content retrieval scheme should be able to work with various instructional strategies for different learning activities. To solve the context-aware learning content retrieval problem, we propose a strategy-driven approach to derive content retrieval strategies from instructional strategies. Moreover, we construct a knowledge-based system to expand query keywords based on the derived strategies and then select relevant keywords according to geographical distance between entities of concept and learners. This system is composed of four components: knowledge transformation, query expansion, content retrieval and user interface. Besides, ontology construction algorithms, designed for teachers to easily build up ontology from course outlines, are applied to generate the rules of query expansion and the taxonomic index of learning object repository. The experimental results indicate that the proposed approach can increase the learning performance of students.

Keywords

Ubiquitous learning, Context-aware learning, SCORM, Information retrieval, Ontology, Knowledge-base

Introduction

With the rapid development of wireless communication and sensor technologies, ubiquitous learning (u-learning) has become a promising solution to educational problems which can sense the situation of learners and provide adaptive supports to students (Chen et al., 2007; Hwang et al., 2008; Kuo et al., 2007; Si et al., 2006; Yang, 2006). Context-awareness is the major characteristic of u-learning, and the situation or environment of a learner situated can be sensed. There are twofold advantages of context-aware learning. One can alleviate environmental limitations, and the other can utilize available resources to facilitate learning.

There are several types of applications for context-aware u-learning. A typical scenario is "learning with on-line guidance" as presented in Hwang's research (2006), which focused on the "identification of plants" unit of the Nature Science course in an elementary school. The context is in campus, and the human-system interaction is described as follows:

- System: Can you identify the plant in front of you?
- Student: Yes.
- System: What is the name of this plant?
- Student: Ring-cupped oak.
- System: Do you see any insect on it?
- Student: Yes.
- System: Can you identify this insect?
- Student: No.
- …

The assumption is that the system is aware of the location of the student and her/his nearby plants by sensor technologies and built-in campus maps.

Learning activities in ubiquitous environments are directed by instructional strategies which are general approaches instead of specific methods. As shown in Figure 1, instructional activities are generated according to instructional strategies originated from pedagogic theories. Designers of learning activities should utilize the advantages of u-learning environments to realize pedagogic goals.
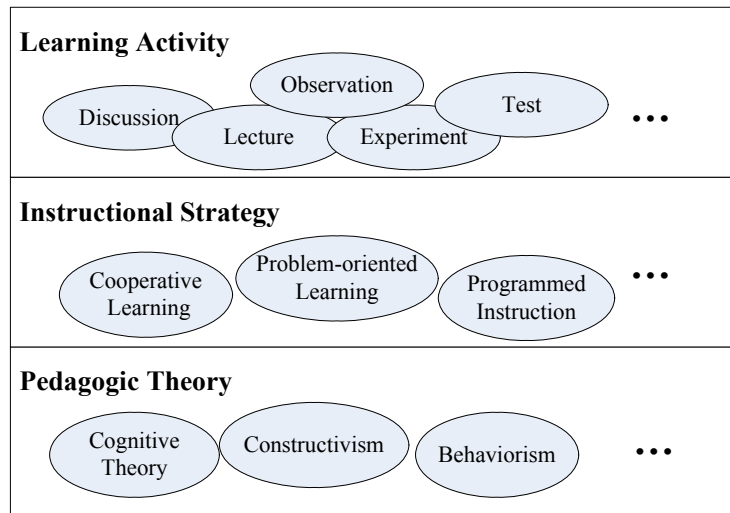

Figure 1. Layered relation of instructional activities, strategies and theories

The context-aware learning content retrieval problem is motivated by the following assumptions:

- Students' learning performance can be improved by providing right content at right time and right place.
- During ubiquitous learning, students' queries are usually related to knowledge of their nearby objects.

Retrieval of learning content, hereafter named Content Retrieval (CR), is an important activity in u-learning, especially for on-line data searching and cooperative problem solving. Furthermore, both teachers and students need to retrieve learning content for teaching and learning respectively. However, conventional keyword-based content retrieval schemes do not take context information into consideration, and therefore they cannot fulfill the basic requirements of u-learning to provide users with adaptive results. To support context-aware learning, learning contents need to be provided according to learners' contexts. For example, when a student can not identify an insect in the u-learning course, s/he can access a learning object repository for more information by submitting a query. Thus we can imagine that queries are most likely ambiguous and need refinement. If context information can be applied to refine the original query, it will be easier for learners to retrieve relevant contents.

As shown in Figure 2, we classify the schemes of content retrieval into static and dynamic types according to the adaptability of the retrieved results. For static CR, the retrieved result only depends on the query regardless of users and contexts. Nevertheless, dynamic CR can be further divided into personalized, context-aware, and other schemes according to the factors considered by the adaptive mechanisms of CR. The static CR is adapted to subjective factors of learners, such as user profile, preference, etc. In other words, the same query submitted by different persons could result in different results retrieved. On the other hand, context-aware CR is adapted to objective factors of learners, like time, place, device, activity, and peers etc. Hence, the same query run in different contexts could obtain distinct results.

Learning content retrieval is a universal requirement for many learning scenarios, such as Intelligent Tutoring Systems, and Zone of Proximal Development etc. However, each scenario has its own needs for content retrieval. In particular, an important characteristic of context-aware ubiquitous learning is to provide right contents to learners at right place and right time. In other words, it is required for a retrieval system to acquire the contents adapted to the learners' contexts.
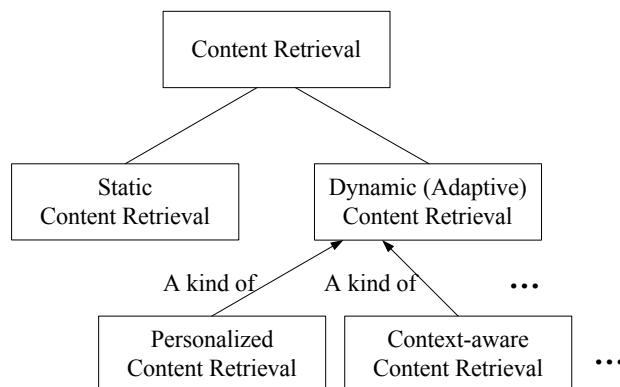
Figure 2. Classification of Content Retrieval

In this research, we investigate the context-aware learning content retrieval problem, which is to retrieve relevant learning contents from a repository according to the given query and context information, to improve the precision and recall of retrieval. However, there are some difficulties in the process of our research. First, context information needs to be taken into account for context-aware retrieval. Therefore, traditional information retrieval schemes have to be enhanced as context-aware. Second, needs for teaching materials are related to pedagogic factors, such as instructional strategies and goals. It is required to design a retrieval scheme which is flexible enough to adapt to various instructional strategies. Third, characteristics of standardized learning content must be considered to improve the accuracy of similarity comparison, such as metadata and structural information. Accordingly, the acquisition of context information requires extensive deployment of sensors. In this paper, we assume that the module of context acquisition is available and focus on the previous three issues.

To overcome the aforementioned difficulties, we propose a strategy-driven approach enabled by a knowledge-based query expansion method. First, we intend to expand the original query by acquired context information to retrieve content which is adapted to learners' contextual environments. We adopt the technique of query expansion because most queries in web search are short, ambiguous, and refinement-needed. Second, we propose a knowledge-based approach to expand queries based on instructional strategies. According to our observation on ontology, such as Wordnet, basic strategies of query expansion include generalization, specialization, association, and their combination etc. For example, when the educator aims to encourage the learners to do high-level thinking, it may be appropriate to adopt an expansion technique of generalization which offers more general keywords for content retrieval. In this study, we assume that the content about entities near the learner is more relevant than that far from the learner. For example, when walking by a fern plant, we may want to find some content introducing the fern. Also, this work assumes the instructional strategy and the strategy mapping are defined by experts in advance. Otherwise, that will be our future work and we may focus on applying retrieval strategies to realize context-aware content retrieval.

Based on the aforementioned idea, we designed a system consisting of four components: knowledge transformation, query expansion, content retrieval and user interface. In the knowledge transformation component, algorithms of ontology building and rule generation are proposed, and therefore teachers can easily construct ontology from course outlines. The purposes of the ontology are to generate rules of query expansion and to construct taxonomic index of learning object repository. Besides, the remaining three components work as follows. In user interface, the user submits a query, and the context information is extracted by sensors. Next, in query expansion component, candidate keywords are inferred for query expansion, and keywords with entities closer to the learner are selected. Finally, in content retrieval component, results are retrieved according to the expanded query, and they are ranked by a similarity function considering characteristics of learning content. To speed up the searching process, we use a taxonomic index, which is generated by reorganizing the existing documents based on a bottom-up approach (Shih et al., 2008).

We think the proposed context-aware retrieval method can benefit the ubiquitous learning scenario by providing suitable content adapted to learners' context and instructors' strategies. Experiments have been conducted to show the evidence of this claim. First, an experiment involving 20 elementary school students is conducted to show the

learning performance affected by the proposed retrieval method. Next, a survey involving 15 elementary school students is performed to understand their degree of satisfaction for this system. The results show that this system can streamline the retrieval process and facilitate the learning activity as well. In addition, the comments of teachers indicate that this system can effectively find suitable contents adapted to context and instructional strategies.

The contributions of this paper can be summarized as follows. First, we propose a strategy-driven approach to solve the context-aware learning content retrieval problem. This new approach integrates pedagogic requirements and technical solutions, and thus it can benefit both parties. Second, a knowledge-based system is designed to support query expansion, which can increase maintainability. Moreover, the flexibility of the knowledge-based approach facilitates future integration with educational strategies. In addition, the distance-based keyword selection can achieve context-awareness. Third, knowledge transformation algorithms are designed for automatic derivation of ontology and query expansion rules, and thus it can resolve the difficulties for teachers to manually construct ontology and code rules. Finally, we build up a prototype and obtain experimental results to show that this approach can increase accuracy, and it is helpful to context-aware learning.

Preliminaries and Related Work

In this section, we review background knowledge about u-learning, Sharable Content Object Reference Model (SCORM), context-aware information retrieval and query expansion. Moreover, related researches about this work are surveyed.

Ubiquitous Learning

The rising of u-learning results from the convergence of e-learning and ubiquitous computing. However, this topic is too new to get a well accepted definition. Hwang et al. (2008) in their paper compared u-learning systems with m-learning systems and proposed twelve models for u-learning activities. Illustrative examples in that paper help readers understand what u-learning is like. In Si et al.'s paper (2006), existing u-learning applications were categorized as shown in Table 1, and a frame-based model was proposed to represent context-aware applications.

Table 1: Categorization of ubiquitous learning applications

Application Type	Example
Location-aware learning guidance	• Museum guide (Oppermann & Specht, 1999) • Tour guide (Abowd et al., 1997) • Conference assistant (Dey & Futakawa, 1999)
Correlation-aware collaborative learning	• Japanese polite teaching (Yin et al., 2005) • Knowledge awareness map (El-Bishouty & Ogata, 2006) • P2P content access and group discussion (Yang, 2006)
Task-aware supervised learning	• Requirement satisfied learning (Cheng et al., 2006)

Sharable Content Object Reference Model (SCORM)

To share and reuse teaching materials, several standards have been proposed recently. Among those standards, SCORM (http://www.adlnet.org/) is the most popular standard for learning contents sharing. It was proposed by the U.S. Department of Defense's Advanced Distributed Learning (ADL) organization in 1997. This standard consists of several specifications developed by IEEE LTSC (Learning Technology Standards Committee, http://ltsc.ieee.org/wg12/), IMS (Instructional Management System, http://www.imsproject.org/), and AICC (Aviation Industry CBT Committee, http://www.aicc.org/) etc. The SCORM specifications are a composite of several specifications developed by international standards organizations. In a nutshell, SCORM is a set of specifications for developing, packaging and delivering high-quality education, and for training materials whenever and wherever they are needed (Nitto et al., 2006). In SCORM, content packaging scheme is proposed to package the learning objects into standard teaching materials. The content packaging scheme defines a teaching materials package consisting of four components: 1) Metadata, which describes the characteristics or attributes of this learning content; 2) Organizations, which describe the structure of the teaching material; 3) Resources, which denote the

physical files linked by each learning object within the teaching material; and 4) the (Sub) Manifest, which describes this teaching material, consisting of itself and other teaching materials. SCORM Metadata refers to the IEEE's Learning Object Metadata (LOM), and describes the attributes of teaching materials. IEEE LOM v1.0 includes nine categories: General, LifeCycle, Meta-Metadata, technical, educational, rights, relation, annotation, and classification.

Information Retrieval

Inverted file indexing has been widely used in information retrieval (Baeza-Yates & Ribeiro-Neto, 1999). An inverted file is used for indexing a document collection to expedite the searching process. The structure of an inverted file consists of two components: the vocabulary and the posting list. The vocabulary is composed of all distinct terms in the document collection. For each term, a list of all the documents containing this term is stored. The set of all these lists is called the posting list. However, the structure of a document is not counted in this model.

Storage requirements of inverted indices (Lee et al., 1996) have been evaluated based on B^+-tree and posting list. Five strategies of the index term replication were discussed. This approach is extended to analyze the storage requirement of the proposed approach in this paper. In Cambazoglu & Aykanat's paper (2006), 11 different implementations of ranking-based text retrieval systems using inverted indices were presented, and their time complexities were also investigated.

Query Expansion

A number of researches have focused on query expansion and tried to solve the ambiguity problem of short queries. Cui et al. (2003) divided a work on automatic query expansion into two classes: global analysis and local analysis. In Bhogal et al.'s paper (2007), query expansion approaches were reviewed and classified into three categories: relevance feedback, corpus dependent knowledge models and corpus independent knowledge models. Actions of query expansion can be based on various ideas. The point is how to determine the correlation between each pair of keywords. Relevance feedback uses top-ranked items in the returned set. Corpus independent models make decision by thesaurus. Corpus dependent models calculate co-occurrence in the corpus. For educational purposes, teacher-oriented schemes are needed. Our idea is based on geographical proximity. The recommended keyword is mainly related to the learner's location instead of the original query.

There were few researches focused on query expansion for education. Kumela et al. (2004) used a query expansion technique implement advanced search features in support of Constructivist Learning. They claimed that learners need a mechanism to retrieve learning materials to support Constructivist Learning. Lee et al. (2008) presented an ontological approach to retrieve learning objects. They proposed an ontology-based query expansion algorithm for inferring user intention according to the original query. Experiments were conducted in terms of Precision, Recall and F-measure. These two researches indicate that applying query expansion to the search of educational resources is promising. However, they did not address the learning performance brought by the advanced search technology.

Related Work

SCORM-compliant teaching materials can be seen as structured documents. For fast retrieval of information from structured documents, Ko et al. (2002) proposed a new index structure integrating element-based and attribute-based structure information to represent a document and presented three retrieval methods: top-down, bottom-up and hybrid methods. Although this index structure takes information of elements and attributes into account, it is not suitable for management of a huge amount of documents due to its time and storage complexity. There have been numerous studies on Structured Document Retrieval (Trotman, 2004, 2005). Researches showed that structured searching can increase precision. Previous work mainly addresses XML and SGML documents. Besides, XML Query Languages, such as XIRQL, XQL, etc., were proposed. However, intra-document structural modeling is not suitable for SCORM-compliant documents.

For sharing and reusing teaching materials in different e-learning system, the Sharable Content Object Reference Model (SCORM) has become the most popular international standard among the existing ones. In the LOR, a huge

amount of SCORM teaching materials, including associated learning objects, will result in management problems. Su et al. (2005) proposed a management approach called the *Level-wise Content Management Scheme* (LCMS) to efficiently maintain, search and retrieve learning contents from a SCORM compliant LOR. LCMS includes two phases: the *Construction phase and Search phase*. In the beginning, the content structure of SCORM teaching materials (Content Package) is first transformed into a tree-like structure called a Content Tree (CT) to represent each piece of teaching material. Based on Content Trees (CTs), the proposed *Level-wise Content Clustering Algorithm* (LCCAlg) then creates a multistage graph showing relationships among learning objects (LOs), e.g., and a Directed Acyclic Graph (DAG) called the *Level-wise Content Clustering Graph* (LCCG). This level-wise approach employs the structural information of SCORM-compliant document for retrieval, but the metadata of content packages has not been utilized to increase its precision. Besides, its time and storage complexity have not been addressed in depth.

With the flourishing development of e-Learning, more and more SCORM-compliant teaching materials are developed by institutes and individuals in different sites. In addition, the e-Learning grid is emerging as an infrastructure to enhance traditional e-Learning systems. Therefore, information retrieval schemes supporting SCORM-compliant documents on grid environments are gaining its importance. To minimize the query processing time and content transmission time, a bottom-up approach was proposed to reorganize documents in these sites based on their metadata and to manage these contents in a centralized manner (Shih et al., 2008). An indexing structure named Taxonomic Indexing Trees (TI-trees) as shown in Figure 3 was designed. It is a taxonomic structure with two novel features: 1) reorganizing documents according to the Classification metadata such as queries by classes can be processed efficiently and 2) indexing dispersedly stored documents in a centralized manner which is suitable for common grid middleware. This approach is composed of a Construction phase and a Search phase. In the beginning, a local TI-tree is built from each Learning Object Repository, and then all local TI-trees are merged into a global TI-tree. In the end, a Grid Portal processes queries and presents results with estimated transmission time to users. Experimental results show that the proposed approach can efficiently retrieve SCORM-compliant documents with good scalability. We extend the index scheme to support context-aware learning content retrieval.
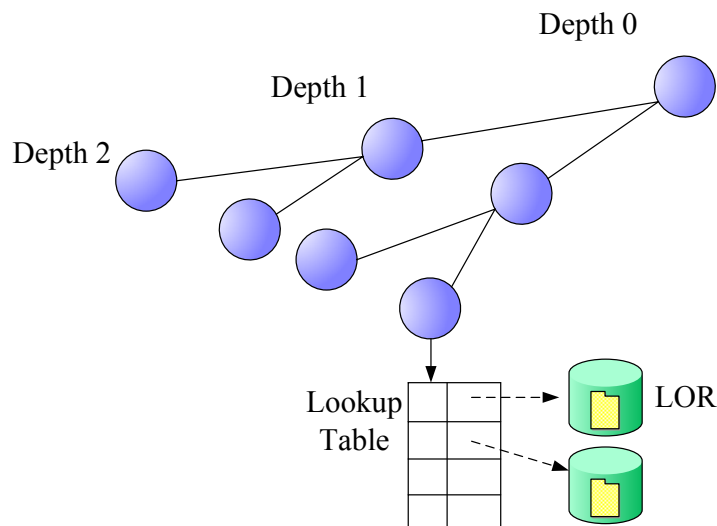

Figure 3. A Taxonomic Index Tree for learning content organization

A Knowledge-based Approach

The main difficulty for a learner to accurately retrieve relevant teaching material is how to formulate a good query. Usually, the query submitted by a novice user is short and ambiguous, and thus many irrelevant results are retrieved. Accordingly, a knowledge base system (KBS) is very suitable for solving this problem. After the expertise of query formulation is transformed into the knowledge base, the inference engine of the KBS can infer other appropriate keywords to expand the original query and get more relevant results. Furthermore, the knowledge can be reused by others and easily adapted to different scenarios.

To apply a knowledge-based approach to context-aware retrieval, the most important thing is knowledge acquisition. In this section, we present the knowledge transformation model for teachers to use an ontology-driven method to transform their course expertise into query expansion rules. This study assumes that a teacher plans the course outline before s/he teaches a course and lists the concepts to be learned in the class. The Ontology Building Algorithm is designed to assist teachers to transform a course outline into ontology, and then the ontology is transformed into rules by the Ontology_to_Rule Algorithm. In other words, once the course outline is planned, it can be semi-automatically transformed into rules. Therefore, it is a helpful tool for teachers to rapidly generate rules for query expansion.

Problem Formulation

We assume that a context detection module is available, and it can extract users' context information. Subsequently, several definitions will be introduced, including teaching materials, learning object repository, a query, context and a similarity function.

The symbols in Table 2 are used throughout this paper.

Table 2: The notation used in this paper

Symbol	Description
CP	Content Package
LOR	Learning Object Repository
w_i	Weighting element i of CP vector
V	Set of terms in the vocabulary
Q	Query
v_Q	Vector representing the query
LC_x	The x-coordinate of location context
LC_y	The y-coordinate of location context
$sim()$	Similarity function

In the SCORM standard, a Content Package (*CP*) is defined as a package of learning materials, and a Learning Object Repository (*LOR*) is a database where the *CP*s are stored. In this paper, a *CP* is modeled as a tree to represent the structural information of a CP. To enable content-based retrieval, the well-known Vector Space Model is applied to represent the text content. Besides, SCORM metadata is included in this model of *CP* as an attribute to utilize the metadata given by authors. Afterward, teaching material, learning content, SCORM-compliant documents and Content Packages are used interchangeably in this paper.

Definition 1 (Content Package). A Content Package (CP) is modeled as a rooted tree where the leaf nodes contain the content and the internal nodes represent the structural information. The content of each level is represented by a vector. In addition, a CP is associated with a set of Metadata.

Example 1
An example CP is modeled as a trinary tree with three levels as shown in Figure 4. The leaf nodes contain the content, and the internal nodes represent the structural information. In addition, a *CP* is associated with a set of Metadata.

Traditionally, similarity is measured by the Vector Space Model (VSM) in information retrieval domain (Baeza-Yates & Ribeiro-Neto, 1999). In the VSM-based model, a document is represented as a vector. In general, a limited vocabulary of keywords is adopted to denote important words in documents. Each element of the vector corresponds to a keyword of the vocabulary. Therefore, the length of the vector for a document is equal to the size of the vocabulary. The value of each element is a weight denoting the importance of the keyword to the document. There are a number of methods to determine the weights. Among them, TF-IDF (Salton & McGill, 1983; Salton et al., 1975) is the most well-known method to assign weights. The TD-IDF method is based on two findings. First, words frequently used in a document, except stop words, are important keywords to the document. Second, words frequently appearing in many documents are not important for the purpose of differentiation.
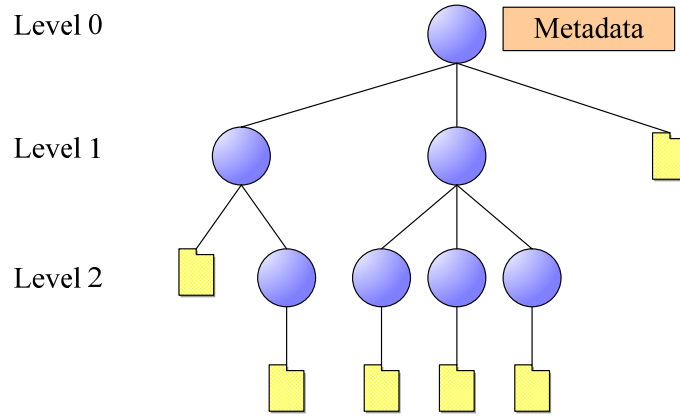
Figure 4. The model of a Content Package (CP)

As standardization of teaching materials becomes a trend, we model TMs by the SCORM standard, where a Content Package (CP) is defined as a package of learning materials. The content package is represented by a vector of keyword weights. To enable content-based retrieval, the Vector Space Model is applied to represent the text content and to calculate the keyword weights. Also, the Educational category of the LOM metadata is included in this model of CP. Therefore, a TM representation consists of a vector of weights and a category of metadata. Figure 4 illustrates an example of TM representation, where the vector represents weights of 10 keywords and metadata denotes educational attributes. Afterward, teaching materials, SCORM-compliant documents and Content Packages are used interchangeably.

Definition 2 (Learning Object Repository). A Learning Object Repository is a set of Content Packages located in the same site.

We use weight vector to represent the query, and its formal definition is as follows.

Definition 3 (Query). A Query is used by a user to specify the TMs s/he wants. Users can express their queries in two forms: keyword-based and metadata-based. A keyword-based query is a vector of keyword weights which mean the concepts about the desired contents. A metadata-based query is a list of (Attribute, Value) pairs, which describe the properties of TMs.

We will now define the notion of similarity between a query and a content package which means the relevance of the content package to the query.

Definition 4 (Similarity). Let Q be a query with query vector v_Q, and TM be a content package. The similarity function is denoted by sim(Q, TM).

In order to determine the degree of relevance for a query and a teaching material, the similarity function has to be defined. Conventional similarity functions, such as the cosine function, are not suitable for SCORM-compliant teaching materials which are characterized by textual content, metadata and structural information. Hence, a similarity measure Sim between a query Q and a teaching material TM is proposed by combining a keyword-based similarity and a metadata-based similarity. The keyword similarity $Sim_{Keyword}$ adopts a cosine function to measure the text similarity between a query and a TM. The metadata similarity $Sim_{Metadata}$ is defined as the number of matched attributes divided by the number of all attributes. Therefore, the range of these two similarity terms, $Sim_{Keyword}$ and $Sim_{Metadata}$, are both in [0, 1]. The similarity measure Sim is defined in (1).

$$Sim(Q, TM) = \alpha \times Sim_{Keyword}(Q, TM) + (1-\alpha) \times Sim_{Metadata}(Q, TM) \tag{1}$$

where the factor α, $0 \leq \alpha \leq 1$, is used to control the relative weighting of keyword similarity and metadata similarity.

In terms of ubiquitous learning, it is widely accepted that the contexts include personal and environmental information sensed by the system or/and retrieved from the databases. Five types of contexts are listed (Hwang et al., 2008; Yang et al., 2007). Among these contexts, "location" may be one of the important information for describing a learner's status (Lonsdale et al., 2003). In this research, we focus on the location context sensed by the system, which is represented by (x, y)-coordinates. Location-awareness means the system is aware of the learner's location by means of technologies such as the Global Position Services (GPS). We assume that context information of location can be acquired by extensively deployed sensors and built-in maps.

Definition 5 Location Context. The Location Context is represented by a two-dimensional coordinate, (LC_x, LC_y), where LC_x is the x-coordinate, and LC_y is the y-coordinate. These coordinates correspond to a map of the campus.

Based on the definitions above, the Context-aware Learning Content Retrieval Problem (CALCRP) can be defined as follows.

Definition 6 Context-aware Learning Content Retrieval Problem (CALCRP). Given a query and context information, this problem is to retrieve relevant learning contents from a repository ranking by a similarity function. The goal is to improve precision and recall of retrieval.

Ontology Building

Ontology building has been considered as a craft rather than an engineering activity. Traditionally, the process of ontology building requires the participation of domain experts and knowledge engineers. Although a number of automatic technologies of ontology construction have been proposed, it is still not easy for teachers and domain experts to build up ontology. Therefore, the ontology building algorithm is proposed for teachers to easily derive ontology from course outline. In this algorithm, an "expert" means an educator who is also good at knowledge engineering.

Before describing the process of ontology building, we give a general definition of ontology.

Definition 7 (Ontology). Ontology is a conceptualization of a domain, which is defined as a quadruple O= (C, A, R, X), where
C is a set of concepts;
A is a collection of attributes sets, one for each concept;
R is a set of relations on C×C;
X is a set of axioms.

Example 3
A Campus_Plant_Course Ontology O_{CPC} = (C, A, R, X) is an ontology where its components are endowed as follows.
C = {"Plant," "Structure," "Fern," …}
A = {Keyword, Type, Location, Level}
R = {"is_a," "related_to"}
X = { IF is_a("A", "B") and is_a("B", "C") THEN is_a("A", "C")}

In this study, the ontology is derived from a pre-defined course outline which reflects the content of the course to be taught by the teacher. The course outline is usually organized by the teacher before the class begins. Moreover, a course outline is defined as a two-level structure, chapters and sections.

Definition 8 (Course Outline). A Course Outline is a two-level tree-like representation of the table of content for a course. A course outline consists of a limited number of Chapters, which consists of a limited number of Sections.

Example 4
A Campus_Plant Course Outline CO_{CP} can be represented as follows.

Course Name: Plants in the Campus

Chapter 1. Introduction to Plants
        Section 1.1. What is plants?
        Section 1.2. Classification of Plants
Chapter 2. Structures of Plants
        Section 2.1. Flowers
        Section 2.2. Leaves
        Section 2.3. Fruits
Chapter 3. Growth of Plants
        Section 3.1. Budding
        Section 3.2. The Growing Process
Chapter 4. Identification of Plants in the Campus

After the course outline is determined, the teacher can follow the steps of the **Ontology Building Algorithm** to derive ontology from the course outline. This is a special-purpose algorithm which is designed for constructing the ontology of a course about plants in a campus. Teachers who teach this kind of courses can follow this algorithm to generate ontology.

Subroutine Topic_Concept_Extraction (Sub_TCE) extracts the topic concept according to the name and content of a chapter by calling a subroutine which is designed based on heuristics. First, the name of the chapter is scanned to extract a keyword representing the topic concept of the chapter. If the given name of a chapter does not represent or illuminate any related concept, the text content of this chapter is represented by the Vector Space Model, and the TF-IDF (Term Frequency-Inversion Document Frequency) scheme is used to find the keyword with the largest weight. Then, this keyword is assigned as the topic concept of the chapter. Likewise, the course concept can be extracted in a similar manner.

Subroutine Topic_Concept_Extraction (Sub_TCE)

Symbols Definition:
Chap_Name: the name of the chapter
Chap_Content: the Vector-Space-Model representation of the chapter
Topic_Concept: the extracted topic concept
Input: Chap_Name, Chap_Content
Output: Topic_Concept

Step 1: Extract the related Topic concept from the name of the chapter.
        Step 1.1: Remove the stop words.
        Step 1.2: Extract the keywords from the name of the chapter in stem forms.
        Step 1.3: Remove the Course concept from the keywords.
        Step 1.4: If only one keyword remains, then assign it as Topic_Concept, and return.
Step 2: Extract the related Topic concept from the content of the chapter.
        Step 2.1: Use the TF-IDF scheme to find the keyword with the largest weight.
        Step 2.2: Assign it as Topic_Concept, and return.

The Ontology Building Algorithm is listed as follows, which can be executed by computers after appropriate implementation.

Campus_Plant_Course (CPC) Ontology Building Algorithm

Symbols Definition:
Course_Outline: a two-level course outline as defined in Definition 8
CPC_Ont: an ontology for a course about campus plants, as defined in Definition 7
Input: Course_Outline
Output: CPC_Ont

Step 1: Build the skeleton CPC ontology.
        Step 1.1: Extract the Course concept from the name of the course.

Step 1.2: For each chapter, call Sub_TCE to extract the Topic concept.
Step 1.3: For each Topic concept, create a "related_to" relation to the Course concept.
Step 2: Develop concept hierarchy for the Course concept with "is_a" relations.
Step 2.1: Create one Category concept for each kind of plants appearing in the course content.
Step 2.2: For each Category concept, create an "is_a" relation to the Course concept.
Step 3: Develop concept hierarchies for Topic concepts.
Step 3.1: For each Topic concept, extract Sub-topic concepts from the names of sections.
Step 3.2: For each Sub-topic concept, create a "sub_topic" relation to the Topic concept.
Step 4: Experts verify the ontology.

Associating concepts with instances, also called Annotation, is an important task for retrieving learning content based on the ontology. However, it is really heavy work for teachers to annotate all plants in a campus. Therefore, we have reduced the workload of a teacher to a reasonable degree. For example, the teacher only has to identify instances of plants appearing in the course content. In other words, all the teacher needs to do is to identify one Instance concept in the campus for each category in the ontology. In the future work, we plan to propose a collaborative approach to annotating plants in a campus-wide manner. By an appropriate distribution of workload, a whole ontology of plants in the campus can be built up and annotated.

The process of ontology building is shown in Figure 5. Two types of relationships in the Campus_Plant_Course ontology (CPC Ontology) are described as follows. (1) "is_a" is a generalization relationship, which could be used to describe the concept taxonomies in the class hierarchy. For example, either a wooden plant or a fern plant is a kind of plants. (2) "related_to" denotes that there exists a "related to" relationship between two concepts. For example, we could use the "related_to" relationship to denote that the "plant" subject is related to the "structure" topic.
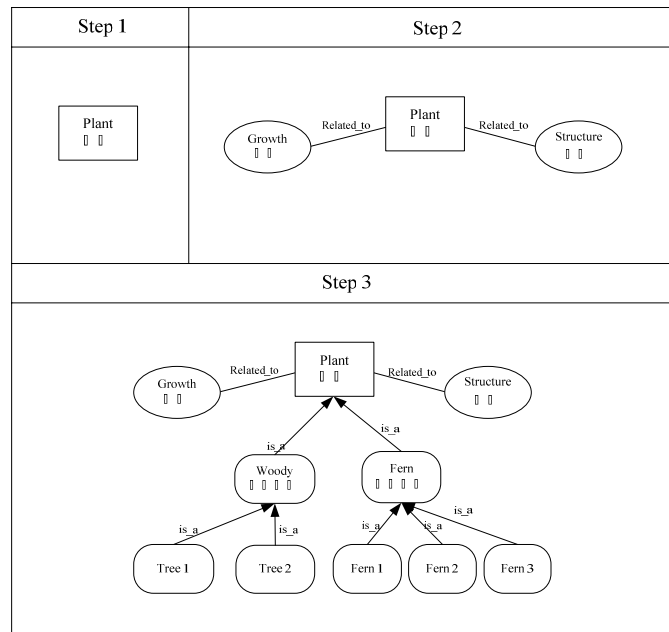

Figure 5. Process of ontology building

This section describes the process of constructing the knowledge base. Throughout this paper, we take the campus plants for example. Just like the concept of object orientation, we could view all of the entities in the campus as concepts and it is natural for us to model the campus using concept hierarchies. For example, a "woody plant" is a concept, and it contains attributes or slots: Leaf_Shape, Leaf_Color, Leaf_Size, etc. Furthermore, people tend to group the knowledge and build up structural information when they learn new concepts. The grouped knowledge could be viewed as a bigger concept as well. For example, both woody plants and fern plants are typical types of plants. Hence, the "woody plant" concept inherits the "plant" concept, and there exists a relationship between them. In essence, ontological representation is suitable for communications and natural for human thinking; meanwhile, rule-based representation is powerful for machine to manipulate the concepts. As described above, ontology could be

used to model the concept hierarchy and relationships between concepts. However, it is not easy to model the behavior of concepts using ontology only. When the problem domain is described clearly and well modeled, it is much easier to build up a rule-based expert system because many tools (called expert system shells) can offer assistances. Hence, in practice, rule-based representation is more suitable for building applications. On the other hand, since most real-world applications need complex rules to model, the meaning captured into ontology for the problem domain becomes very helpful for rule extractions when building complex systems.

System Overview

We have proposed a context-aware learning content retrieval system based on the knowledge transformation model. As shown in Figure 6, the overall system consists of four components:

- User Interface: query input and context detection
- Query Expansion: expanding a query by rule inference
- Content Retrieval: searching and results ranking
- Knowledge transformation: ontology building and rules generation

The flow of the system can be summarized as follows. First, the query and context information are transferred to the Query Expansion component to generate an expanded query. Next, the expanded query is sent to the Content Retrieval component for query processing and retrieving relevant content. Finally, the retrieved results are returned to the user. The search process is carried out by the Query Processor component, which receives expanded queries, processes the queries and presents results to the users.
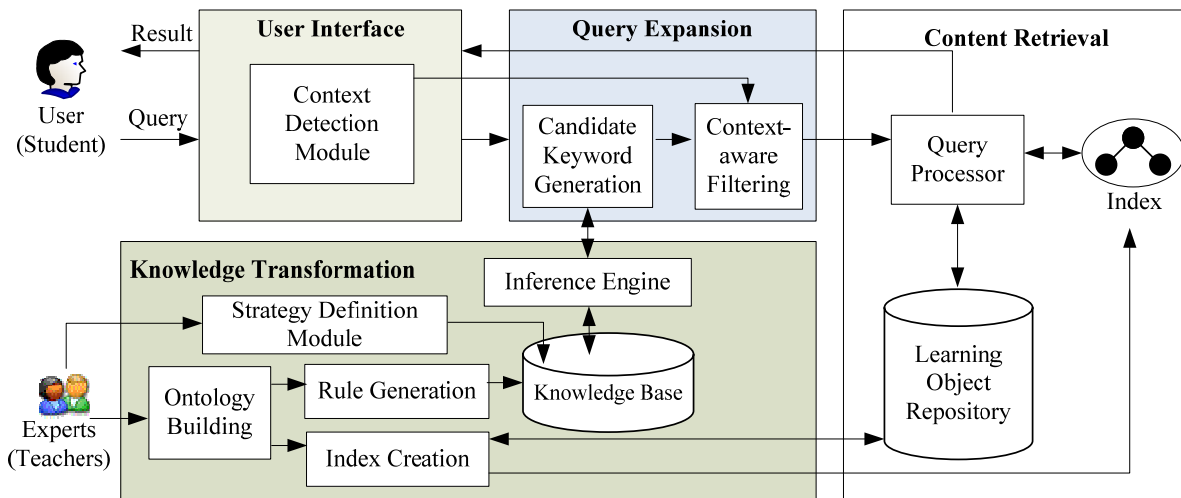

Figure 6. Overview of the system

In this work, we assume that the instructional strategies and corresponding retrieval strategies have been defined in the knowledge base by experts. Strategies of query expansion include:

- Specialization: Giving keywords belonging to subordinate concepts
- Generalization: Giving keywords belonging to super-ordinate concepts
- Association: Giving keywords belonging to related topics

Knowledge is represented by rules describing actions of query expansion. After the generation of ontology, we derive rules for query expansion using an ontology-driven method. Rules are automatically transformed from the ontology and extracted from two kinds of relations:

- "is_a" relation: For generalization/specialization strategies
For example, IF (Strategy="Specialization") and (term="Fern") THEN expand("Fern 1").

- "related_to" relation: For association strategies

For example, IF (Strategy="Association") and (subject="Plant") THEN expand("Structure").

The algorithm is listed as follows.

Ontology_to_Rule Algorithm
Input:
CPC ontology
Output:
Rules for query expansion

Step 1: Extract rules by "is_a" relations.
1.1: for each "A 'is_a' B" relation, generate:
IF (Strategy='Generalization') and (term="A") THEN expand("B")
1.2: for each "A 'is_a' B" relation, generate:
IF (Strategy='Specialization') and (term="B") THEN expand("A")
Step 2: Extract rules by "related_to" relations.
2.1: for each "A 'rel' B" relation, generate four rules:
IF (Strategy='Association') and (Subject="A") THEN expand("B")
IF (Strategy='Association') and (Subject="B") THEN expand("A")
IF (Strategy='Association') and (Topic="A") THEN expand("B")
IF (Strategy='Association') and (Topic="B") THEN expand("A")
Step 3: Verify the generated rules by domain experts and ask the knowledge engineers to modify the rules if needed.

To support the pre-defined instructional strategy and enable context-awareness, we propose to divide the process of query expansion into two phases:

- Phase 1: Candidate keywords generation. Based on the retrieval strategy derived from the instructional strategy, candidate keywords are recommended. This phase extends the instructional strategy.
- Phase 2: Context-aware filtering. This phase focuses on filtering the candidate keywords to realize context-awareness. Among a variety of methods, we adopt a distance-based method to determine relevance of keywords.

The main advantage of dividing the process of query expansion is the separation of pedagogic design and technical implementation. While the first phase generates keywords based on instructional consideration, the second phase is related to technical factors. Various technologies can be applied if necessary. Actions of query expansion can be based on various ideas. The point is how to determine the correlation of two keywords. Conventional methods of calculating keyword correlation include: thesaurus-based, co-occurrence in the corpus and top-ranked in the returned set. Our idea is based on geographical proximity. The expanded keyword is mainly related to the learner's location, instead of the original query.

The distance of two entities is defined as follows.

$$Dist(C, L) = \sqrt{(x_C - x_L)^2 + (y_C - y_L)^2}$$ (2)

Dist(C, L) means the geographical distance between C and L, where
- C: a concept;
- L: a learner;
(x_C, y_C), (x_L, y_L): coordinates of C and L.

As shown in Figure 7 and 8, for specializing "Fern," there are three choices: Fern 1, Fern 2 and Fern 3. We choose the nearest one to the learner.
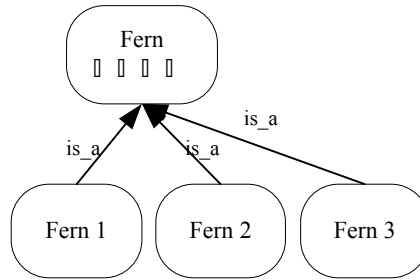
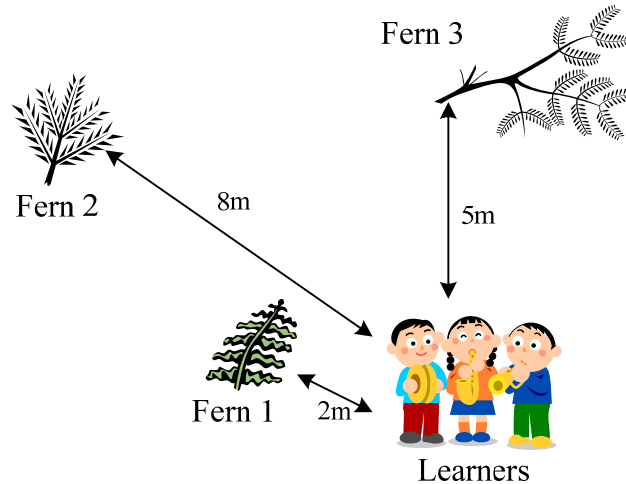*Figure 7.* The Campus_Plant_Course Ontology (partial)

Figure 8. Geographical proximity of plants and learners

The User Interface receives users' queries and context information extracted by the Context Detection Module. Another task of the User Interface component is initialization of facts for inference. For example, the fact of the strategy defined by teachers is initialized and loaded into the knowledge base for inference. Another important task is to initialize the campus map for default reasoning. The campus map records coordinates of primary plants, buildings and other entities. Although we assume that a lot of sensors have been installed in the u-learning environment, it is probable that the user walks by an area without a sensor. In this situation, coordinates in the campus map can serve as default context information to prevent the inference process from being failed.

An Illustrative Scenario

This section presents a scenario of a context-aware ubiquitous learning environment, where some communication and context-sensing devices have been installed to enable context-aware retrieval of teaching materials. For an "identification of plants" class of an elementary school, the teacher sets up five learning corners in the campus. The students in class are divided into five groups, and the teacher arranges an on-line tour guidance for these students. When a student equipped with a PDA goes to the first corner, the system interacts with the student:

- System: Can you identify the plant in front of you?
- Student: Yes.
- System: Can you identify the name of this plant?
- Student: Yes.
- System: What is the name of this plant?
- Student: Fern.
- System: Can you describe the characteristics of a fern?

- Student: No.
- System: Do you need the Search service?
- Student: Yes.
- System: Please input the query:
- Student: fern, characteristic

After the student submits the query, the system conducts the context-aware retrieval and then presents relevant content to the student.

In this scenario, the system retrieves relevant teaching material according to the location context of the student. For example, assuming the fern plant in the first corner is some kind of fern. We may call it Fern 1. We assume that the contents in the repository include teaching materials about Fern 1, Fern 2 and Fern 3. The system will deliver content about Fern 1 to the student in accordance with the context information. The content describing the plant seen by the student will strengthen the impression of the learner and therefore enhance the learning effect.

Also, we think the similarity can improve the learning performance. The query reflects the things the student wants to learn. The more questions of the learner can be solved when he/she is provided with more relevant content, the more the learning effect can be increased. Context-aware retrieval can provide students with relevant teaching materials and shorten the learning process.

The interactive query mechanism is a common process proposed in ubiquitous learning activities. However, the design of a meaningful question base has not been widely discussed in the literature. In our opinion, the system can ask questions of the learner according to a pre-defined scenario generated from the course objectives and the teaching material. For example, if the course objective is to identify the name of some plant mentioned in the teaching material, the scenario will contain such a question as "What is the name of the plant?" In the future work, we plan to make the course objectives converted automatically and fit the teaching material into the query scenarios which can be refined by teachers if necessary.

Method

Our proposed approach has been implemented and the prototype was provided to several teachers of an elementary school for evaluation. The corpus is composed of SCORM-compliant teaching materials adopted from those of repositories built by Taiwan Ministry of Education (http://nature.edu.tw). The experiments investigate the accuracy of searching. Also, some questionnaires involved in such issues are used to respond to the user satisfaction.

Prototype

To evaluate the proposed approach, we have implemented a web-based prototype. As shown in Figure 9, users can submit queries in this web page. Then, the original query is expanded by the knowledge-based system. After that, the required contents are retrieved from repositories and delivered to the user. The retrieved content packages are ranked by their similarities to the query. All programs are implemented in the Java language.

In this paper, we use DRAMA (Lin et al., 2003) as an expert system shell because of its client–server architecture and the object-oriented knowledge base structure. The purpose of the DRAMA's server is to load, manage, and use the knowledge bases according to the knowledge service that users need. DRAMA's server contains many different rulebases and provides different APIs for the application servers to connect. The application server employs DRAMA's APIs to provide user-friendly web pages for users to use expert systems. Based on the client-server architecture, it becomes very easy for us to develop a KBS for supporting context-aware u-learning. The Lucene search engine (http://lucene.apache.org) was adopted to perform basic keyword indexing, search and retrieval in the prototype. This IR engine was open-source software developed by the Apache Lucene project. Lucene is characterized by the ease to enable applications to index and search documents.
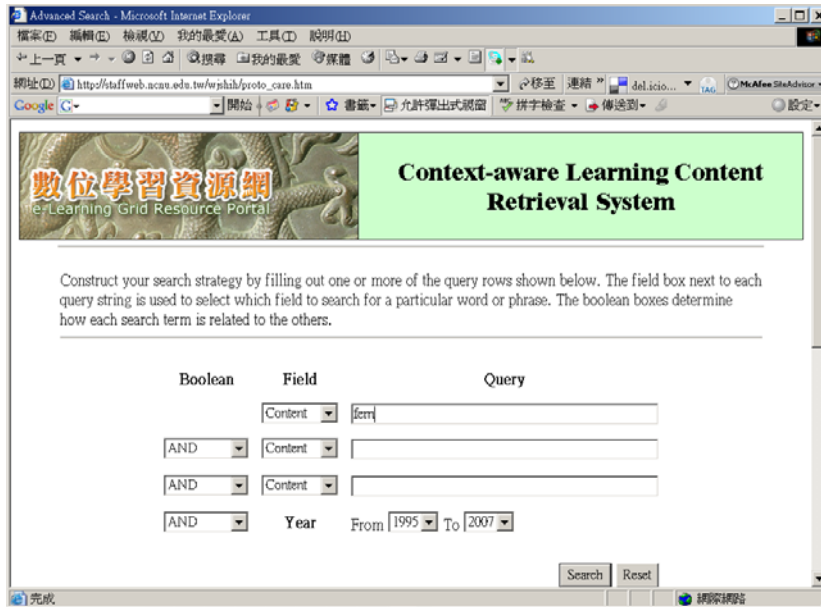
Figure 9. Query submission of the prototype

There are nearly one hundred Content Packages in the LOR, which are retrieved and adopted from existing repositories on the Internet like http://learning.edu.tw/mainpage.php. Currently, the LOR is only available to elementary school teachers who participate in this evaluation. However, in the near future, we plan to place the prototype and the LOR on the web for public access and large-scale evaluation.

The concept of "relevance" of a TM to a query has to be defined before the retrieval method can be evaluated. It is somewhat subjective, depending on the users, to judge whether a document is relevant to a query or not, especially in a dynamic changing environment. For example, a document about fern is usually thought of lowly relevant as a query of "tree." However, while the user is standing in front a fern plant, this document will be highly relevant to a query. In this experiment, we try to make an objective, reliable and fair measure for the relevance. First, we define that a document and a query is relevant if their similarity value calculated by (1) exceeds the threshold value assigned in advance in the system. Next, for 5 places in the campus, we manually generate a query for each of them. Then, the documents in the repository are manually judged its relevance to the 5 queries by teachers and experts.

In this experiment, we use two well-known metrics of information retrieval, precision and recall, to measure performance of the proposed approach. We define precision and recall as follows.

$$Precision = \frac{R\_ret}{Ret} \tag{3}$$

$$Recall = \frac{R\_ret}{R\_LOR} \tag{4}$$

where
- R_ret is the number of relevant documents in the retrieved documents;
- Ret is the number of retrieved documents;
- R_LOR is the number of all relevant documents in all repositories.

To optimally set up the parameters of the similarity function, α is not easy. In this experiment, we set $\alpha = 0.5$. This setting means equal weighting for keyword similarity and metadata similarity. Other settings of the parameter will be considered in future work.

Study Design

In this section, three experiments/surveys are conducted to address respectively as follows:
* the performance of knowledge-based query expansion;
* the learning performance of using the proposed system; and
* the satisfaction survey of using context-aware retrieval.

The purpose of this first experiment is to evaluate the performance of knowledge-based query expansion with respect to the precision and recall of context-aware content retrieval. The participants are 15 fourth-grade students invited from an elementary school in Nantou of Taiwan. After one-week usage training, they use this system to retrieve relevant teaching material. At each learning corner in the campus, every student submits a query. Then, the original query is transformed by the system into a specialized query and a generalized query respectively. Next, the three queries are used to retrieve teaching materials, and the precision and recall values are calculated. Figure 10 illustrates the average precision and recall values for the original query, the specialized query, and the generalized query. The results show that the expanded queries perform better than the original one. The main reason may be that the original queries are usually short and ambiguous, and therefore they are insufficient to represent the intention of users. In addition, we found that generalization can improve recall, and specialization can improve precision. This is consistent with the cognition of precision and recall.

The second experiment was conducted to evaluate the effects of the location-aware learning content retrieval system. This experiment specifically investigated the following question:

"Can location-aware learning content retrieval contribute to students' learning?"

This experiment was conducted at an elementary school, and the subjects were 20 fourth-grade students from the same class taught by the same teacher. The experimental course is named "Introduction to Campus Plants," which aims to help students learn knowledge about plants in the campus. Before the experiment conducted, students were trained to be familiar with the retrieval tool. The training consists of two phases. First, all students learned in a conventional classroom and used the learning content retrieval system without the location-awareness feature. Then, they were educated how to access the retrieval tool through a PDA. After training, the 20 students were separated into two groups: A (control group) and B (experimental group), each of which contained 10 students. The students in Group-A had access to learning content retrieval system without the location-aware feature while those in Group-B had access to the proposed location-aware learning content retrieval system. However, the location-awareness feature is transparent to students. That is, users do not know that the tool can retrieve learning content according to their location context. Before the course began, the 20 students were given a pre-test to ensure that they had the equivalent performance for the course. After the course finished, a post-test was given to respond to the comparison of the learning achievements. The contents of the two tests cover those learning materials related to the "Introduction to Campus Plants" course in the Learning Object Repository. Common threats to internal validity include: Confounding, Selection (bias), History, Maturation, Repeated testing, Instrument change, Regression toward the mean, Mortality/differential attrition, Selection-maturation interaction and Experimenter bias. In this experiment, we are careful to avoid these threats. The statistical results from applying SPSS to analyze the tests are presented in the next section.

The items of the pre-test include 25 single-choice questions, and sample items are shown in Appendix A. In this study, the items were presented in Chinese when undertaking this study, and the test items, shown in the Appendix, were translated by the authors. We went through five stages during the process of test construction: item construction, preliminary test, item analysis, item selection and validity analysis. First, a two-way specification table was constructed according to learning content and Bloom's Taxonomy of Educational Objectives (Bloom, 1956) as shown in Table 3. The content of the test covers the following levels of expertise: knowledge, comprehension, application, analysis and synthesis. The draft of the pre-test items includes 30 single-choice questions. 30 fourth-grade students who did not participate in the pre-test and the post-test are invited to take the preliminary test. According to the results of the preliminary test, we eliminate 5 questions which have low difficulty and discrimination. The two-way specification table shows the learning units and the educational objectives. After the initial construction of the pre-test items, two teachers in the fields of Nature Science were invited to make comments on it for face validity.

Table 3: The two-way specification table of the pre-test

Learning Unit \ Level	Knowledge	Comprehension	Application, Analysis and Synthesis	Total
Functions of root/stem/leaf	3	4	2	9
Functions of flower/fruit/seed	3	4	2	9
Other	2	4	1	7
Total	8	12	5	25

The items of the post-test were constructed in a way similar to the pre-test, as shown in Appendix B. The two-way specification table of the post-test is shown in Table 4.

Table 4: The two-way specification table of the post-test

Learning Unit \ Level	Knowledge	Comprehension	Application, Analysis and Synthesis	Total
Nutritional organs of plants	3	3	2	8
Reproduction of plants	3	6	2	11
Other	2	3	1	6
Total	8	12	5	25

The two variables that may contribute to the difference between the two groups are carefully controlled in the experiment.

- The learning environment.

The students of the two groups used the same retrieval system in the same outdoor campus. The only difference is that the context-awareness feature of the retrieval system was disabled for the control group.

- Teachers and learning activities.

Students were taught by the same teacher during the experiment. Additionally, the instruction or facilitation received by students was not different.

Next, a survey was conducted with respect to the 15 students in the first experiment. The questionnaire contains questions of four categories, each of which has five questions. The purpose is to contemplate their comments on the retrieved contents and learning effect. Besides, a five-point Likert scale with anchors ranging from strongly disagree (1) to strongly agree (5) is used for this survey.

Experimental Results

In this section, the results of three experiments/surveys are reported.

The purpose of this first experiment is to evaluate the performance of knowledge-based query expansion with respect to the precision and recall of context-aware content retrieval. At first glance, the results shown in Figure 10 seem to conflict with conventional information retrieval practice which indicates the trend of decreasing precision along with the rise of recall. In fact, there is no conflict. To generate a typical precision-recall plotting for a given query, the set of retrieved documents are listed. Next, the precision and recall are calculated accumulatively from the first document to the last document. Finally, these pairs of precision and recall values are plotted in a 2-dimentional coordinate figure with the precision against the recall. In this kind of figures, the following trend usually holds, as the recall rises, the precision decreases. However, Figure 10 is not obtained in this manner. Given a collection and a query, the precision and recall values are calculated by (3) and (4). That is, while we focus on one query, conventional experiments illustrate the results of multiple queries. Therefore, it is possible for the expanded query to outperform the original query in both precision and recall.
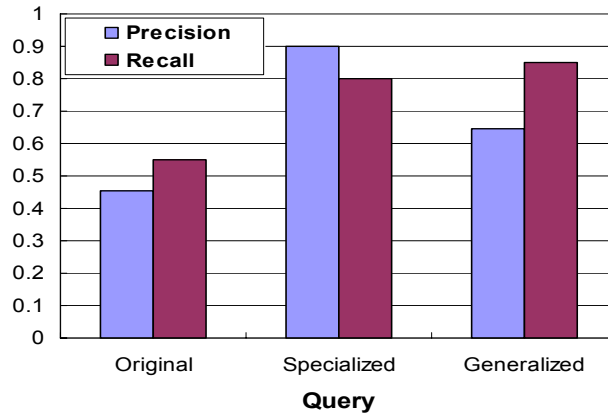
Figure 10. Comparison of Precision and Recall

The second experiment was conducted to evaluate the effects of the location-aware learning content retrieval system.

- Pre-test

The aim of the pre-test is to ensure that both groups of students had the equivalent performance for the course. Table 5 presents the t-test results of the pre-test. The mean scores for the pre-test reveal that Group A performed as well as Group B. That is, $|t| = 1.37 < t_\alpha(9) = 1.833$, which implies that the performance of Groups A and B in the pre-test does not differ significantly. Therefore, we can conclude that Group A performed as well as Group B in the pre-test conducted before performing the experiment.

Table 5: t-test of the pre-test results (α=0.05)

	Number	Mean	S.D.	t
A (Control Group)	10	52.30	6.67	
B (Experimental Group)	10	47.70	7.06	
A - B	10	4.60	10.64	1.37

- Post-test

The post-test aims to compare the learning achievements of the two groups of students after taking the course with different retrieval tools. Table 6 shows the t-test results for the post-test. From the mean value of the post-test, Group B performed better than Group A. $|t| = 1.98 > t_\alpha(9) = 1.833$, which implies a significant difference between the performance of Groups B and A in the post-test. Therefore, we can conclude that Group B achieved a significant improvement compared to Group A after using the location-aware learning content retrieval system.

Table 6: t-test of the post-test results (α=0.05)

	Number	Mean	S.D.	t
A (Control Group)	10	83.40	16.08	
B (Experimental Group)	10	91.70	5.83	
A - B	10	-8.30	13.23	-1.98

Next, a survey was conducted with respect to the 15 students in the first experiment. The mean value and standard deviation (SD) are calculated for each category, as shown in Table 7.

Table 7: The result of the survey for students

Category No.	Questions	Mean	SD
1	Satisfaction of the user interface of the system	2.9	1.06
2	Satisfaction of the retrieved contents and learning effect	4.3	0.89
3	Willingness to use this system for learning	3.7	1.03

For questions of Categories 1 and 3, the deviation of user satisfaction is slightly larger than other categories. The reason may be that the participants are not all familiar with the usage of the system and some English interfaces. Some participants comment that they are not used to studying on computers. However, some participants appreciate this idea and like to retrieve relevant TMs.

The results of Categories 2 show that most participants are satisfied with the expanded queries and the retrieved contents. Most students agree that the expanded queries can enhance their original queries, and the retrieved contents are helpful to their learning. In summary, the system can help students efficiently find relevant teaching materials for learning. However, the user interface has to be improved to attract more users.

Discussion

Based on the assumptions mentioned in the Introduction Section, the experimental results can be interpreted as follows. First, the location-aware retrieval tool can provide students with "right" content about their nearby plants, insects, etc. We think that the right content can promptly solve students' questions. Furthermore, the knowledge related to such nearby objects could be more impressive and interesting to increase users' learning performance.
Based on the assumptions, the experimental results can be interpreted as follows:

- The location-aware retrieval tool can provide students with "right" content about their nearby plants, insects, etc. We think that the right content can promptly solve students' questions and that the knowledge related to nearby learning objects could be more impressive and interesting to improve the learning performance.
- The integration of physical situation into instruction can enhance students' learning performance. This viewpoint can also be found in the model of "Situated Learning" proposed by Lave & Wenger (1991). Traditionally, learning has been viewed as transmission of knowledge from teachers to students. However, the Situated Learning Theory argues that learning should be situated in a specific contextual, social and physical environment.
- The "seamless" feature of ubiquitous learning makes the learning process more convenient than the conventional learning. For example, in an ideal u-learning environment, a student with an RFID-enabled device can experience "seamless learning at right time and right place."

The findings of experimental results are interpreted in terms of related literatures. First, we address the application of query expansion to context-aware retrieval. Query expansion has been investigated to improve recall of information retrieval and disambiguate the meaning of queries. A large number of researches have been devoted to this topic (Bhogal et al., 2007), but query expansion has not been widely applied to the e-learning domain and mentioned in the ubiquitous learning. Shih et al. (2008) indicated that efficient retrieval of teaching material could facilitate the learning process and then enhance the learning effects. In this study, the result of the first experiment implies that the proposed query expansion can improve the performance of context-aware retrieval, which can support the ubiquitous learning scenario.

Currently, researches of context-aware ubiquitous learning focus on the acquisition and modeling of context information (Oppermann & Specht, 1999; Yang, 2006; Yin et al., 2005), such as location, temperature, and humidity etc. In this study, we find that the context-aware retrieval tool is helpful to students' learning. They comment that the system can retrieve context-related contents, which saves their time to find relevant references.

The knowledge engineering like ontology building has been thought of as tough work which can only be dealt with by domain experts and knowledge engineers. Therefore, researches of ontology building focus on automatic or semi-automatic approaches (Liu et al., 2004; Tho et al., 2006) to alleviate the burden of the builders. This study proposes a teacher-guided approach to build simple ontology for educational usage. The result of survey shows that it is feasible for teachers to provide their expertise and help the system generate a simple ontology based on a pre-defined course outline.

In addition, the contribution of this paper about pedagogical feasibility and keyword association is clarified. First, to derive the important components of the proposed approach, the ontology and the knowledge base, an automatic approach is adopted to alleviate the burden of teachers and domain experts. For the proposed setting, teachers guide the automatic construction of the ontology by providing a course outline and instances of concepts existing in the campus. Next, domain experts verify the ontology generated by the proposed algorithm. We do not intend to require

teachers to manually build up the rules and ontology. Instead, teachers provide their knowledge about course outline and campus context, which is not difficult, and the system will transform them into ontology and rules. In this way, the proposed approach will be pedagogically feasible. Second, the proposed approach is distinct from existing searching engines. In particular, such as Google, we can not adaptively find results through users' context. Although collaborative filtering techniques have succeeded in suggesting contents of keyword association mining from users' query logs, the context-awareness has not been integrated into this technology.

Conclusion

With the flourishing development of e-learning, more and more learning object repositories are constructed and connected to share the content. Efficient content retrieval schemes, such as context-aware retrieval, can reduce the response time and thus attract more users to utilize the e-learning systems. The existing methods of information retrieval can not rapidly and accurately satisfy the request of retrieving desired learning contents in context-aware ubiquitous learning environments. The proposed approach attains rapidness and precision by a knowledge-based approach with which the desired learning contents can be efficiently retrieved and then advance the sharing and reusing of learning contents. For example, one application of this technology is to teach material design. To support individualized and adaptive learning, teachers are encouraged to develop various teaching materials by means of different requirements. However, traditional methodologies for designing teaching materials are time-consuming. To speed up the development process of teaching materials, teachers can reuse existing contents to rapidly generate context-aware contents.

We have designed a knowledge-based approach to query expansion for context-aware learning content retrieval. In addition, a prototype was built to implement the model. Experimental results show the proposed approach can enhance original queries to improve the precision. We have also shown that explicit context and content ontology can be used during knowledge transformation for dealing with the complexity of context-aware knowledge management, especially for guiding the construction of the knowledge base. Our main contributions are: (1) to design and implement a framework of context-aware learning content retrieval for supporting context-aware u-learning; (2) to propose an ontology-driven model for eliciting rules from a previously built ontology and constructing the knowledge base. According to the experimental results, the paradigm of using ontology to build up a context-aware learning content retrieval system works well and effective. This system will benefit from the sharing and reusing of built knowledge, the reduction of people's time to learn knowledge management, the ease of context-aware course design and the improvement on the precision of learning content retrieval.

The main findings from the experimental results can be summarized as follows: First, the proposed knowledge-based query expansion can improve the performance of context-aware retrieval. Second, using context-aware retrieval can efficiently retrieve relevant teaching materials which are helpful to learners. The results of the experiment in which 20 elementary school students participated reveal that some students, compared with the others in the control group, using the location-aware learning content retrieval system made significant progress in learning. Therefore, we conclude that the proposed approach can help students improve their learning performance.

On the issue of knowledge maintainability, the proposed approach is based on an object-oriented rule model (NORM), which facilitates knowledge maintenance and reuse. When the scenario or contextual environment is changed, the designers can update the original ontology and the new rules will be automatically generated. It is supposed that the same approach could be adaptively modified to other contextual scenarios for knowledge base construction. Moreover, teaching material design is a sustainable and evolving task that the contextual knowledge of the learning environment might need to be updated from time to time.

Location-aware learning content retrieval can be applied to various ubiquitous learning activities as shown in Table 1, such as museum guide and laboratory assistance etc. In these applications, the readers can easily apply the system to their teaching activities. For example, the learner can retrieve relevant content about the artist's work in front of her/him during museum guidance.

Contexts of learning status and learning requirements are very important clues for retrieving learning materials. Hence, we plan to consider these contexts acquired by Learning Management Systems for context-aware learning content retrieval. For example, a Learning Portfolio Mining (LPM) system was developed to extract learning features

from learning portfolio(Su et al., 2006). With this context information, the proposed knowledge-based approach can be easily extended to consider the educational contexts by substituting learner-related knowledge such as learning status and learning requirements and then to adaptively provide appropriate content for learners. Currently, this work focuses on the location context. However, this approach can be extended to consider more types of context information.

In the experiment, the only difference between the control and experimental group is the "location-aware feature." Namely, students of the two groups use the proposed knowledge-based tool with the same retrieval interface. While the experimental group is implicitly supported with the location-aware feature, the control group can only retrieve results which are not refined from the location context. The proposed knowledge-based approach is flexible and can be extended to process other kinds of context information such as personal profile, time, location and temperature etc. This work focuses on the location context. Therefore, the experiment was carefully set to avoid the interference of other features. Furthermore, the experiment addresses the impact of location-aware content retrieval on students' learning. We believe that other context features could also improve students' learning, which will be investigated in our future work.

In addition, our future work will investigate knowledge acquisition for mapping instructional strategies to retrieval strategies. Next, we plan to extend this approach to model personalized learning content retrieval and hence to facilitate an adaptive learning environment. Also, the proposed approach will be extended to model other types of context information, such as time, activities, and peers etc. In this paper, we focused on the management of learning contents which have three types of attributes: textual contents, metadata and structural information. SCORM is one of the standards which satisfy the requirements. In fact, improving the proposed method to consider the learning design of contents such as IMS-Learning Design is another interesting issue of our future work.

Acknowledgements

References

Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R., Pinkerton, M., & Centre, U. (1997). CyberGuide: A Mobile Context-aware Tour Guide. *ACM Wireless Networks, 3*(5), 421-433.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: ACM Press.

Bhogal, J., Macfarlane, A., & Smith, P. (2007). A review of ontology based query expansion. *Information Processing and Management, 43*(4), 866-886.

Bloom, B. S. (Ed.). (1956). *Taxonomy of Educational Objectives*. New York: David McKay Company Inc.

Cambazoglu, B. B., & Aykanat, C. (2006). Performance of query processing implementations in ranking-based text retrieval systems using inverted indices. *Information Processing and Management, 42*, 875-898.

Chen, C.-M., Li, Y.-L., & Chen, M.-C. (2007, 18-20 July). *Personalized Context-Aware Ubiquitous Learning System for Supporting Effectively English Vocabulary Learning.* Paper presented at the Seventh IEEE International Conference on Advanced Learning Technologies, Niigata, Japan.

Cheng, Z., Han, Q., Sun, S., Kansen, M., Hosokawa, T., Huang, T., & He, A. (2006). *A proposal on a learner's context-aware personalized education support method based on principles of behavior science.* Paper presented at the 20th International Conference on Advanced Information Networking and Applications, Vienna, Austria.

Cui, H., Wen, J.-R., Nie, J.-Y., & Ma, W.-Y. (2003). Query expansion by mining user logs. *IEEE Transactions on Knowledge and Data Engineering, 15*(4), 829-839.

Dey, A. K., & Futakawa, K. M. (1999). *The Conference Assistant: Combining context-aware with wearable computing.* Paper presented at the 3rd International Symposium on Wearable Computers, San Francisco, CA, USA.

El-Bishouty, M. M., & Ogata, H. (2006). *Personalized Knowledge Awareness Map in Computer Supported Ubiquitous Learning.* Paper presented at the International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands.

Hwang, G.-J. (2006, 5-7 June). *Criteria and Strategies of Ubiquitous Learning.* Paper presented at the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, Taichung, Taiwan.

Hwang, G.-J., Tsai, C.-C., & Yang, S. J. H. (2008). Criteria, Strategies and Research Issues of Context-Aware Ubiquitous Learning. *Journal of Educational Technology & Society, 11*(2), 81-91.

Ko, S.-K., & Choy, Y.-C. (2002). *A structured documents retrieval method supporting attribute-based structure information.* Paper presented at the 2002 ACM symposium on Applied computing, Madrid, Spain.

Kumela, D., Watts, K., & Adrion, W. R. (2004). *Supporting constructivist learning in a multimedia presentation system.* Paper presented at the the 34th Annual Frontiers in Education (FIE 2004), Savannah, GA.

Kuo, F.-R., Hwang, G.-J., Chen, Y.-J., & Wang, S.-L. (2007). *Standards and Tools for Context-Aware Ubiquitous Learning.* Paper presented at the Seventh IEEE International Conference on Advanced Learning Technologies, Niigata, Japan.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate peripheral participation*: Cambridge: University of Cambridge Press

Lee, M. C., Tsai, K. H., & Wang, T. I. (2008). A practical ontology query expansion algorithm for semantic-aware learning objects retrieval. *Computers & Education, 50*(4), 1240-1257.

Lee, Y. K., Yoo, S.-J., Yoon, K., & Berra, P. B. (1996, March). *Index structures for structured documents.* Paper presented at the the 1st ACM international conference on digital libraries, Bethesda, Maryland, United States.

Lin, Y. T., Tseng, S. S., & Tsai, C.-F. (2003). Design and implementation of new object-oriented rule base management system. *Expert Systems with Applications, 25*(3), 369-385.

Liu, C.-L., Tseng, S.-S., & Chen, C.-S. (2004). Design and implementation of an intelligent DNS management system. *Expert Systems with Applications, 27*(2), 223-236.

Lonsdale, P., Baber, C., Sharples, M., & Arvanitis, T. (2003, May 19-20). *A Context Awareness Architecture for Facilitating Mobile Learning.* Paper presented at the MLearn 2003 Conference, London,UK.

Nitto, E. D., Mainetti, L., Monga, M., Sbattella, L., & Tedesco, R. (2006). Supporting Interoperability and Reusability of Learning Objects: The Virtual Campus Approach. *Educational Technology & Society, 9*(2), 33-50.

Oppermann, R., & Specht, M. (1999, August). *Adaptive mobile museum guide for information and learning on demand.* Paper presented at the 8th International Conference on Human-Computer Interaction, Munich, Germany

Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. New York: McGraw & Hill.

Salton, G., Wang, A., & Yang, C. (1975). A Vector Space Model for Information Retrieval. *Journal of the American Society for Information Science, 18*, 613-620.

Shih, W.-C., Tseng, S.-S., & Yang, C.-T. (2008). Using Taxonomic Indexing Trees to Efficiently Retrieve SCORM-compliant Documents in e-Learning Grids. *Educational Technology & Society, 11*(2), 206-226.

Si, N.-K., Weng, J.-F., & Tseng, S.-S. (2006, September). *Building a Frame-Based Interaction and Learning Model for U-Learning.* Paper presented at the 3rd International Conference on Ubiquitous Intelligence and Computing, Wuhan and Three Gorges, China.

Su, J.-M., Tseng, S.-S., Wang, W., Weng, J.-F., Yang, J. T. D., & Tsai, W.-N. (2006). Learning Portfolio Analysis and Mining for SCORM Compliant Environment. *Educational Technology & Society, 9*(1), 262-275.

Su, J. M., Tseng, S. S., Wang, C. Y., Lei, Y. C., Sung, Y. C., & Tsai, W. N. (2005). A Content Management Scheme in SCORM Compliant Learning Object Repository. *Journal of Information Science and Engineering, 21*(5), 1053-1075.

Tho, Q. T., Hui, S. C., Fong, A. C. M., & Cao, T. H. (2006). Automatic fuzzy ontology generation for semantic Web. *IEEE Transactions on Knowledge and Data Engineering, 18*(6), 842-856.

Trotman, A. (2004). Searching structured documents. *Information Processing and Management, 40*, 619-632.

Trotman, A. (2005). Choosing Document Structure Weights. *Information Processing and Management, 41*, 243-264.

Yang, S. J. H. (2006). Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. *Educational Technology & Society, 9*(1), 188-201.

Yang, S. J. H., Zhang, J., & Chen, I. Y. L. (2007). Ubiquitous Provision of Context-aware Web Services. *International Journal of Web Service Research, 4*(4), 83-103.

Yin, C., Ogata, H., & Yano, Y. (2005, November). *Ubiquitous-learning system for the Japanese polite expressions.* Paper presented at the IEEE International Workshop on Wireless and Mobile Technologies in Education, The University of Tokushima, JAPAN.

Appendix A. Selected Sample Items of the Pre-test

1. (   ) Which is the main function of roots? (1) Water absorbing (2) Water transpiration (3) Nutrients making (4) Nutrients transportation.
2. (   ) Which function do potato and taros have in common? (1) Climbing (2) Breathing (3) Nutrients storage (4) Insect catching.
3. (   ) Which sentence is false? (1) Some plants may have metamorphosed leaves, which have special functions. (2) Stems of most plants are like long tubes for water transportation. (3) Roots of most plants are thin and branching for absorbing water. (4) Leaves of most plants are flat for absorbing water.
4. (   ) Which is the main function of seeds? (1) Photosynthesis (2) Nutrients transportation (3) Growing new plants (4) Water absorbing.
5. (   ) What is the purpose of blooming and fruiting? (1) Food for humans (2) Food for animals (3) Reproduction (4) Nutrients consumption.

Appendix B. Selected Sample Items of the Post-test

1. (   ) Most plants have vascular bundles and belong to (1) Liliopsida (2) Magnoliopsida (3) Non-Vascular plants (4) Vascular plants.
2. (   ) Which sentence is true? (1) The leaves of the cactus are thick and large. (2) The leaves of the cactus store a lot of water. (3) The cactus can grow well in dark and wet area. (4) The needle-like leaves of the cactus is to reduce the consumption of water.
3. (   ) Which sentence about potato and sweet potato is true? (1) Both are roots. (2) Both are stems. (3) The part we eat can store nutrients. (4) The part we eat can produce nutrients.
4. (   ) Which plant uses leaves for reproduction? (1) Pea (2) Air plant (3) Fern (4) Pine.
5. (   ) Which is not the feature of ferns? (1) Non-blooming (2) Non-fruiting (3) Reproduction using spores (4) Found at dry and warm places.