Recommendation of Desired Learning Path to Provide Optimized Concept Access

D.SANDHIYA
Assistant Professor,
Department of Computer Science and Engineering,
Tagore Institute of Engineering and Technology,
Salem, Tamilnadu,
sandhiya.sasi@gmail.com

A.SURYA
Department of Computer Science and Engineering,
Tagore Institute of Engineering and Technology,
Salem, Tamilnadu,
surya.cse89@gmail.com,

G.SURESH
Head Of the Dept, Dept of Electrical and Electronic Engineering,
Thirumala Polytechnic college,
Tiruvannamalai, Tamilnadu
suresha3686@gmail.com

Abstract— Now a day's browser becomes a significant tool for extracting information from the internet. End-users are gathering a fragmented knowledge through the net. But still user facing some difficulties that are, there is no sequential order in the materials users are moving back and forth to figure out which page has to read first and what is to be read next, they are wasting their useful time to ordering the materials. To avoid such a problem we prefer a special path and recommending some objects to learn. Special path construction is based on the data mining concepts to select the keywords. Then by using formal concept analysis we extract the significant keywords to produce the suitable learning path for end-user. This will lead to an efficient surfing to surfers.

Index Terms—Fragmented knowledge, Data Mining, Formal Concept Analysis (FCA), Learning Material Recommendation, TF/IDF.

I. INTRODUCTION

In current scenario most of the peoples are expanding their knowledge by consuming fragmented information from Internet. Educationists or enthusiasts, seeing the benefit of on-line material, feverishly set up their web sites to share their knowledge of specific domains. Though being versatile, these web sites generally follow no standards for content organization and presentation order. They might be just web pages or some learning objects embedded in web pages. When posted on to the Internet, collected and indexed by robots using keywords, and returned by powerful search engines, usually, a vast amount of them Homepages or learning objects is returned directly to a user with no sequential order. The return snippets by the search engine really may related to each other but still uses move back and forth to figure out which page has to read first. Although a user might have some intuitions about the domain but these intuitions are yet to be connected. Thus, this methodology will lead to a effective sequential learning of fragmented knowledge and which is helpful to a user who are novel to a desired domain.

II. SYSTEM ARCHITECTURE

In this paper, several approaches are used to automatically construct a suitable learning path and recommend suitable contents from gathered materials. The architecture of the anticipated system is shown in Fig. 1. The major components carries out the procedure are the Learning Interface, the Candidate Course Generator, the Learning Object Content Preprocessor, the Learning Object Correlative Weighting Generator and the Learning Object Recommendation Module.

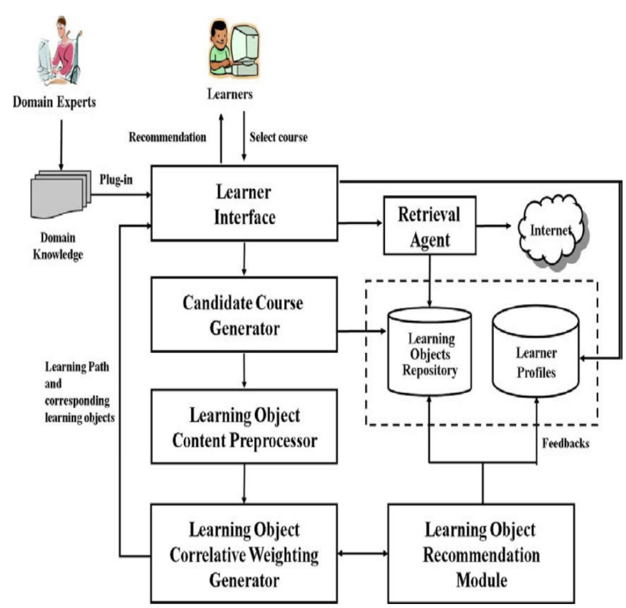

Fig. 1. The system architecture.

A. System description

A learner logins to the Learner Interface to give a subject (a learning goal or a short query) he/she desired to learn. If it is a novel subject that is not yet queried before, the novel subject will send through the retrieval manager to collect the knowledge about appropriate subject entered by the user to the web .That exact information of the novel subject will be fed in to the repository. Domain expert can plug in an object which provides information about specific knowledge according to a user's intended subject. Base on a subject, the Candidate Course Generator discover several courseware unit patterns by data mining on the collected learning resources using the well-known Apriori algorithm. Then the Learning Object Content Preprocessor is notified to find out all the significant keywords of every learning object, using an Adapted TF–IDF (ATF–IDF) algorithm. After that, the Learning Object Correlative Weighting Generator constructs a hierarchy of relationships between all the concepts represented by the keywords and uses FCA to further compute the mutual relationships among all the learning objects or documents to decide a suitable learning path. After the path is constructed the appropriate path will chose, the Learning Object Recommendation Module uses both the preference-based and the correlation based algorithms, which order the typical learning object related to the documents to a learner's intension and preference, for recommending the most related learning objects or documents for each unit of the courses in order to facilitate more efficient learning for all the learners. Let's see the details of the components under the architecture.

B. System Components

1).Learning Interface: To develop a secure and stable learning environment, the open source and web-oriented course management system Moodle is employed. Moodle, is short form for Modular Object-Oriented Dynamic studying Environment, which allows easy learner management, courseware construction, as well as high teaching activity via browsers (Moodle). To build a sophisticated studying environment we have a learner interface we are that learner gives subject to learn. User profile database created as soon as user entered his/her intended subject, wherein the user profile database consist of user intention, feedback and preferred subject. With help of profile database the learning path construction will be building and recommended the appropriate subject according to the user intention. The learner interface using some retrieval distributor to pick the information/documents/materials from the web when it is novel to the system. The retrieved learning materials then are put into the learning object repository.

2). Candidate course generator: The primary function of the Candidate Course Generator is to find the patterns present in a course. From that it will form a course unit according to a learner's intended subject entered by the users. Based on that course unit generator pick some appropriate subject towards user's wish. Fig. 2 illustrates the processes of candidate course unit generator. Before find out a course pattern, a material is processed to have all its significant keywords extracted and tagged. Each learning material will record a set of course unit themes, which is called a themeset (set of themes). Next the generator collects set of themes of the course unit from these learning objects. After that it will finds other related i.e. virtually related candidate course unit in order to be more accordant with the user intended subject. To do so we are applying apriori algorithm from the concept of data mining to discover the association rules among all collected course unit themes. An itemset is called a large-item set if its support value is greater or equal to the user-specified support threshold (called minSupport). Assume an itemset I= {i1, i2, i3,…,im}is a set of items.

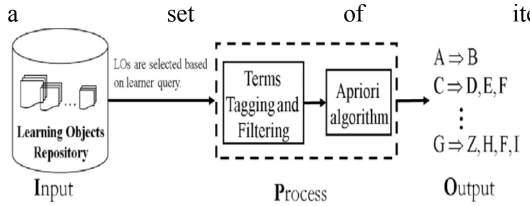

Fig. 2. Processes of candidate course generator

Apriori algorithm

Input: Learning objects repository (LOR), Threshold of minimum support value (minSupport).

Output: Large itemsets in learning objects repository (LI).

Procedure:

1: LI1 = find large 1-itemsets in LOR.

2: For (k=2; $LI_{K-1} \neq \phi$; K++){

Ck = apriori-gen(LI_{K-1}); // New candidates

3: for all of records r $\in$ LOR {

Ct = subset(Ck, r) // Candidates contained in r

for all of candidates c $\in$ Ct

c.count++;}

$$LI_K = \{c \in c_k \mid c.count \geq \min Support\}\}$$

4: Return $LI = \cup_k LI_k$;

*3). Learning Object Content Preprocessor:*After discovering a set of candidate course units and their associated learning materials by the Apriori algorithm, the learning object content preprocessor extract all the important keywords from the materials. A usual way of doing such keyword extraction is to use TF–IDF (Term Frequency/Inverse Document Frequency) that finds out for a document those specific keywords that distinguishes the document from the others. The less a keyword of a document appears in other documents, the more specific the keyword is for the document and the higher the TF–IDF value. The formula of TF– IDF is shown in following,

$$r_{ij} = \frac{\left(n_{ij} \big/ \sum_{l=1}^{k} n_{il}\right) \times \log(N/n_j)}{\sqrt{\sum_{j=1}^{t} \left(n_{ij} \big/ \sum_{l=1}^{k} n_{il}\right)^2 \times [\log(N/n_j)]^2}}$$

where rij is the significance of keyword j in document i with its value between 0 and 1, nij is the number of appearance of keyword j in document i,

$\sum_{l=1}^{k} n_{il}$ is the total meaningful term frequency in document i, N is the total number of documents, and nj is the number of document in which keyword j appears.

4). Learning Object Correlative Weighting Generator: After significant keywords of each material are extracted, several steps are involved in computing mutual relationships. In this paper, the Formal Concept Analysis approach can be used to derive implicit relationships between concepts using learning objects and attributes associated with these concepts. Three steps are required in this paper in using FCA to construct inter-relationships between significant keywords of learning objects. Furthermore, these relationships are used to calculate the correlative weighting that represents the coherences between learning objects in the final step.

*5). Learning Path Generation:*To fulfill a goal issued by a learner, e.g. to satisfy a learner's desired to understand a specific terminology in a domain, usually requires prerequisite knowledge for the terminology. Learning objects associated to the requisite knowledge must be prepared for such a learning activity and they must be thought in the front side of the learning path, if necessarily. This paper builds a suitable learning path for a learner according to the Correlative Weighting Metrics of learning object collected. To build a learning path satisfying a specific goal with better continuity, the learning object most matches the goal is chosen as the main subject of the learning, learning object has the highest weight of correlation to the goal is selected for the prerequisite knowledge and the all others are found and arranged subsequently according to that correlative weighting.

*6). Learning object recommendation module:*Every course unit may be associated with several learning objects in the repository. These learning objects may contain different phase of difficulties and different types of content. Anyhow among that learning materials some may be opt for a user's intention, some of the learning material may not. Hence it should prefer the most suitable course unit of a learning path for each material entered by user is the most significant phase. In order to provide such a significant i.e. to recommend the most suitable learning material we introduce technique namely learning material recommendation. Under the technique several sequence of phases are there. And this kind of recommendation using to algorithm namely, preference-based and correlation-based evolution approaches. which use a learners personal preference as well as his/her neighbor's suggestion to calculate the recommendation course for ranking all the learning objects chosen from the repository for each course unit of a learning path in order to select most suitable one's for the learning path Finally, learner feedbacks phase deals with the feedbacks of learners and updates their profile that, as a result will affect the next recommendation.

III. EXPERIMENTS

Several experiments have been conducted to show merits of the approach proposed in this paper. Some of them are presented as follows.

A. ATF–IDF vs. TF–IDF

This experiment is conducted in order to compare the precision of TF–IDF and ATF–IDF on extracting significant keywords for the proposed approach. Data used are learning objects taken from the Java Learning Object Ontology (JLOO), which was introduced in several human experts, are invited to pick out from these learning objects all the significant keywords that should be chosen by the approach and gather them into a relevant keyword set. The two significant keyword sets built by TF– IDF and ATF–IDF respectively are compared with the base set by two criteria, the precision and the recall rates. The precision rate represents the percentage of the picked keywords that are in the relevant keyword set and the recall rate shows the percentage of relevant keywords that are correctly picked out in both methods; they are formulated as follows:

$$Precision = \frac{|\{\text{Relevant Keywords}\} \cap \{\text{Retrieved Keywords}\}|}{|\{\text{Retrieved Keywords}\}|}$$

$$Recall = \frac{|\{\text{Relevant Keywords}\} \cap \{\text{Retrieved Keywords}\}|}{|\{\text{Relevant Keywords}\}|}$$

B. Learning Path Construction

Several experiments have being conducted to evaluate the proposed approach in this section, two will be discussed here. Say for the example we taken 10 experts of Java programming language for evaluating the results in multiple dimensional views. And the system automatically chose the appropriate document over the many documents. The first step is about how to identify the course units among the collection of many learning materials to provide appropriate learning materials desired by the user. Then the second experiment is doing the evaluation on the correlation between learning objects and the construction of learning path don by the proposed methodology.

$$Precision = \frac{|\{\text{Relevant Course Units}\} \cap \{\text{Retrieved Course Units}\}|}{|\{\text{Retrieved Course Units}\}|}$$

$$Recall = \frac{|\{\text{Relevant Course Units}\} \cap \{\text{Retrieved Course Units}\}|}{|\{\text{Relevant Course Units}\}|}.$$

$$F = \frac{2*precision*recall}{precision+recall}$$

IV. CONCLUSIONS

Most research on e-learning has used a pre-constructed category or some manual information for semi-automatic learning object classification and learning path construction in specific domains. The experiments show that the proposed approaches achieve more than 50% accuracy in finding candidate courses, and at least a 66.7% suitability in constructed learning paths. Over 77% of learners agreed that the appropriate learning path, which the proposed system constructed, helped them learn the Java programming language. Moreover, 90% of learners agreed that the recommended learning materials also helped them. The proposed approach helps learners browse and read collected learning objects or documents in the correct order to understand the fragmented knowledge.

REFERENCES

[1] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In Proceedings of 20th international conference on very large data bases (pp. 487–499).Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[2] Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD conference on management of data (pp. 207–216). Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[3] Aizawa, A. (2003). An information-theoretic perspective of TF–IDF measures.Information Processing & Management, 39(1), 45–65.Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[4] Chen, C.-M. (2008). Intelligent web-based learning system with personalized learning path guidance. Computers & Education, 51, 787–814.

[5] Chen, C.-M., & Chung, C.-J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. Computers & Education, 51, 624–645.Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender

[6] Chen, C.-M., Liu, C.-Y., & Chang, M.-H. (2006). Personalized curriculum sequencing utilizing modified item response theory for web-based instruction. Expert Systems with Applications, 30(2), 378–396.

[7] Chi, Y.-L. (2007). Elicitation synergy of extracting conceptual tags and hierarchies in textual document. Expert Systems with Applications, 32(2), 349–357.