A NOVEL SCHEME FOR INFORMATION RETRIEVAL FROM E-LEARNING REPOSITORY

Aasim Zafar[1], and Syed Hamid Hasan[2]

[1,2] Faculty of Computing and Information Technology,
King Abdulaziz University, P.O. Box. 80221, Jeddah-21589 (KSA)

Email: [1]azahmad@kau.edu.sa, [2] shhasan@kau.edu.sa

ABSTRACT

The repository of any learning management system (LMS) keeps growing and becomes a rich source of learning materials with the passage of time. This learning resource may serve subject experts by allowing them to reuse the existing material while preparing online instructional materials. At the same time it may help the learners by allowing them to retrieve the relevant documents for efficiently achieving their learning goals. We have proposed a novel scheme for searching documents relevant to concept knowledge to be imparted to students, which assists subject experts in synthesizing the course material, by facilitating them to reuse existing learning objects available in e-learning repository. It also helps students in finding relevant learning resources efficiently for interactive e-learning. This paper presents an efficient way of retrieving information related to the teaching domain from a vast reservoir of documents. We have employed fuzzy clustering, fuzzy relation along with information retrieval techniques to discover the underlying structure of knowledge and identify knowledge based relationship between learning material and retrieving the relevant documents. The experiments conducted to judge the suitability of fuzzy clustering for discovering good document relationships and to evaluate the performance of the proposed information retrieval system, show encouraging results. A practical implementation of this technique has also been demonstrated in the implementation of eLGuide, a framework for an adaptive e-learning system.

Keywords: ***e-Learning, Information retrieval, Domain knowledge representation, Fuzzy clustering***

1.0 INTRODUCTION

E-Learning applications generally involve preparation of material to be used for online instructions by subject experts (authors) and its use by the students for the purpose of learning & achieving their goals. So, the e-learning application can be termed as sucecssful in its goal by mapping the effort put into its development and judging how efficiently the information can be retrieved by the students. The study materials or learning objects are created, stored in the learning repository and made accessible to students via Internet and telecommunication technologies. Over the period of time, these learning repositories in itself become a rich resource of learning and may be used to provide relevant documents to students. E-Learning is such a learning environment where information or material is provided to the student through the internet. One extreme approach to achieve the objective of e-learning is to provide certain links based on particular sequence. However, this approach seems to be rigid as the student has no chance to explore by itself. Another extreme is to provide the entire pool of knowledge to the student but it may stray the students' attention from the original topic. So there must be a path in between. The aim of adaptive e-learning is to provide students the documents that are the most suitable to a specific concept [4]. The proposed work addresses this situation by proposing a method for searching documents in the learning reservoirs and retrieving documents that are related to a concept that the student is currently reading. This also enables the reuse of learning objects in preparing the instructional material by the subject experts and thereby saves a great deal of time and effort of the subject experts. In this paper, a novel approach is proposed to retrieve information related to the teaching domain from a vast reser-

voir of documents using fuzzy clustering and information retrieval techniques which dynamically discover document relationships and which may also be helpul for the teachers in designing the knowledge domain more efficiently.

The proposed solution aims to cluster similar type of documents contained in a reservoir of knowledge using fuzzy-c means clustering [29]. This results in a relevance matrix showing the association of documents in different clusters. The concepts to be imparted to the learners would be identified and specified by the experts. Another relevance matrix is generated specifying the relationship between concepts and documents. Then composing the above two matrices, we obtain the relationship between the concepts and clusters i.e. we obtain a matrix showing the membership degree of various concepts in different clusters. Here the idea of *optimal cluster* is introduced i.e. optimal cluster is the best representative of a particular concept and the probability of finding documents which are the best match to the concept lies with that cluster. The proposed method helps in the efficient retrieval of documents from knowledge reservoir. In addition to providing pedagogically relevant information, the presented work also aims to assist teachers in the most crucial part of e-learning process that is e-content development.

2.0 RELATED WORKS

There are a large number of computer-based educational systems, especially intelligent systems, which are built to support students, interact with courseware and collaborate with the teachers and peers. However, the issue of supporting teachers is rarely considered. This might be explained in the light of student-centered approaches to learning and teaching. In the era of e-learning, teaching strategies are simulated inside the software systems and often the teachers' role is mainly limited to preparing course material (e-content), which is the most important and time consuming activity.

IRIS (IRakaste-Ikaste Sistima; Teaching-Learning System) is a tool that is developed for helping the instructor in building an smart teaching & learning system for various domain [1]. Elorriaga et al. [9] integrate a Lesson Planner Manager to the IRIS that allows the teacher to create specialized lesson plans for students, monitor their results in these lessons and accordingly, take the appropriate instructional decision. Merceron and Yacef [16] developed Logic-ITA, an intelligent system, which is web based, it is to be used as a Teaching Assistant that will facilitate the process of teaching & learning. Like Elorriaga et al. [9], Merceron and Yacef also focus on supporting teachers by providing only pedagogically relevant information [16]. A software tool called Pépite was developed to help instructors diagnose their students' competencies [8]. Mazza and Dimitrova developed an approach to support the facilitators in the WCMS environments [14]. In their system CourseVis, they explored the graphical presentation of the information stored in WCMS through various visualizing techniques.

Supporting the teachers in quantitative evaluation of their distance courses is discussed by Chang [6], who reports that most WCMS aren't integrated with a mechanism to evaluate distance learning course-ware quantatitively. As there are so many students in Web based courses, Santos et al. emphasized the need for helping the teachers in correctly designing and manageing the concerted activity of the learning communities [28]. The work of Santos et al is directed to helping teachers in managing solely collaborative activities whereas other activities, e.g. re-use of existing material in e-content development, knowledge domain design, are not considered.

In the proposed work, we employ fuzzy clustering techniques to discover the knowledge structure which is underlying and identify knowledge-based relationship between the learning material. For clustering of documents the most popular algorithm is Agglomerative-hierarchical-clustering [30]. The document collection is provided in a hierarchical organization by this method however K-Means algorithm, which is also used for document clustering, views the time complexity of this approach as challenging [13]. In these types of algorithms each document is allocated to only one single cluster, thus generating hard clusters. With fuzzy clustering [3] multiple memberships in different clusters are allowed. Nikravesh and Takagi [19] identify the use of fuzzy or neuro-fuzzy clusters to classify without any supervision based upon conventional learning tech-

nique and Genetic & Reinforcement learning. Such algorithms have not been widely explored for clustering of documents although a research at [11] indicates that the Fuzzy c-Means algorithm [5] performs at-par with the traditional agglomerative-hierarchical-clustering methods. Improvment in search & retrieval efficiencies is achieved in information retrieval by application of clustering techniques [26]. Cluster hypothesis supports the use of clustering for information retrival [25]. It is based on the assumption that document which are resultant of a specific query would be more similar to one another in comparison to other unrelated document thus these relevant document are more likely to be bunched with each other. To browse a large collection of documents, clustering is proposed as a tool [19] which can also be used as a tool for organization of search results on the web and meaningful groups after retrieval [32].

The search of material both for authors and learners may be facilitated by the the information provided by the ontology [17]. But we may foresee two problems in this approach, first, there would be a disagreement bytween the experts of a field about the correct ontology; Second, in field of engineering or telematics the correct ontology is dynamic. Hence, the deployment and maintenance efforts are costly. Similar limitations have been pointed out in the context of the Semantic Web [7].

Knowledge domain may be treated as an abstract area that can be developed in various manners. An approach which is emerging popularly is developing the domain ontology i.e. to define the concept and the relation between these concepts [23]. Also, in Cooperative and Network Distributed Learning Environment project (CANDLE), each piece of course material is tagged with metadata according to an extended version of IEEE LOM (Learning Objects Metadata) model [15]. This approach is a major feature of the more general enterprise of the Semantic Web [18].

In the following sections, we describe representation of teaching domain, the proposed approach of information retrieval which integrates various techniques like, soft computing, information retrieval and information extraction. Finally, the experimental results are discussed.

3.0 DOMAIN KNOWLEDGE REPRESENTATION

The Domain Knowledge Base (DKB) contains pre-stored course materials, mostly unstructured document. Domain Meta-Knowledge (DMK) contains information describing the course material along with its interrelation. This section describes the DMK or the data that should be kept to describe the contents of DKB. The entire knowledge domain also called as teaching domain (Fig. 1) is modeled as:

(i) **Domain:** Represnts the teaching domain knowledge.
(ii) **Concepts:** Logical partioning of the Domain Knowledge is done into smaller element or concept [23].
(iii) **Documents:** Pool of documents related to various concepts, which would serve as learning resource for the users.

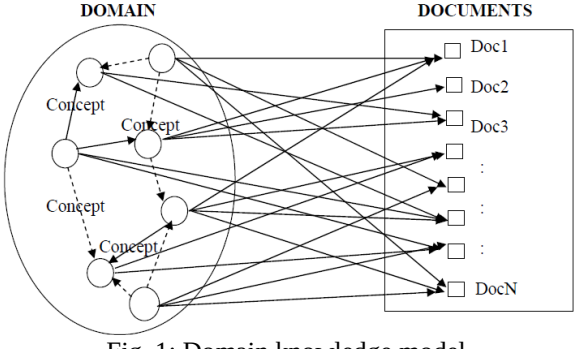

Fig. 1: Domain knowledge model

Domain is the representative of the entire teaching domain and is composed of a number of different con-

cepts (represented by *circles* in Fig. 1). An important feature about concepts in a domain is that they are not isolated but are related to one another in various ways. The concepts within the domain may be interrelated in two ways [10]:

- If it is necessary to learn a particular concept before learning the other concept, then it is said that former concept is pre-requisite for later concept and this relation is shown with full arrow.
- Sometimes the essential pre-requisite relation exists between two concepts only to some extent and this type of relationship between concepts is shown as dashed arrow.

Each concept has some degree of association to a number of documents (represented by small squares in Fig. 1). The concept is considered learned when the user has successfully completed learning of the associated documents. The knowledge of a particular domain is thought to be complete once the entire concepts in that particular domain are learned. . In this sense, domain may be treated as a super set of user model which keeps track of concepts learned by user. Fig. 2 depicts this mapping between user knowledge on domain knowledge.

Brusilovsky and Milla proposed the strict overlay model over the teaching domain for user model representation [21]. It is proposed to start with mapping initial user knowledge (as shown in the grey zone in Fig. 2) with domain knowledge. For each domain concept learnt, the grey zone expands to cover the learnt concepts. Thus, the grey zone of the user model reflects the real-time understanding of the user of that domain. With the progress of the user's performance, the grey zone grows in size to cover larger part of the domain knowledge. Greater the size of the grey zone, better is the understanding of the user of the subject. Once all the concepts of the particular domain are covered with grey zone, the learning of the domain is said to be completed.

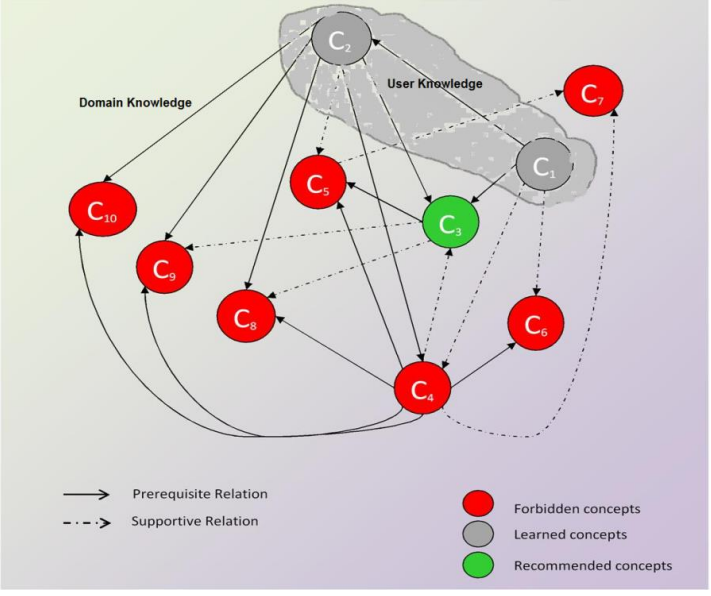

Fig. 2: Mapping of user knowledge on domain knowledge

3.1 Knowledge Representation

Knowledge in a domain is represented by logically partitioning different documents based on certain properties so as to place similar documents in same cluster. Initially there was a hard clustering approach in which the documents were associated with one cluster exclusively but gradually it was found that one document

may have some degree of association to one cluster while some degree of association to other clusters and hence fuzzy c-means clustering is used here to logically partition documents into various clusters, so that similar documents logically belongs to the same cluster.

Each of the documents is denoted by an indexing term in the form a k-dimensional vector:

$$xi = \{wi1, wi2, wi3, ..., wik\} \quad \textbf{(1)}$$

Here k represents the number of terms and w_{ij} represent the *weight* of term j in the documents i and is calculated as:

$$wij = \frac{fij.\log(N/nj)}{\sqrt{\sum_{i=1}^{i=k}(fij.\log(N/nj))^2}} \quad \textbf{(2)}$$

Where, f_{ij} denotes *frequency of j in* i,
n_j - number of document containing term j, and
N - Total number of document.

Here the *weight* reflects the importance of the terms appearing in the entire document set. For every document in the document set (N), the weight for each index term needs to be calculated. Every document contains a document vector which has the weight for every index term that appears in that document. For example, if a word appears in all documents it will not have any value across the document's set. For calculating term weights, a method of combining the importance of a word in a specific document in the whole document set is required. After we obtain the required values, the weights can be calculated [23].

3.2 Clustering Algorithm

The concept behind fuzzy clustering is that of fuzzy logic and the fuzzy set theory [31], which provide the mathematical methods to tackle uncertainty [20]. Fuzzy c-means clustering is used here to logically partition documents into various clusters, so that similar documents logically belong to the same cluster. Every document in a domain contains different association with the given concepts. i.e. a document may have close association to one concept while with other concept it may have lower association. Here our main aim is to categorize document according to their association with different concepts.

The fuzzy c-means algorithm takes as input (N*k) matrix X = [x*i*] for a data set with N elements each represented by k-dimensional feature vector. The number of clusters, (let it be c), and fuzzification parameter m (m=2). Both the clusters centers and the partisan matrix are computed to minimize the objective function given below:

$$Jm(U,V) = \sum_{i=1}^{i=N}\sum_{\alpha=1}^{\alpha=c} u\alpha i^m \|xi - v\alpha\|^2 \quad \textbf{(3)}$$

where, $u_{\alpha i} \in [0,1] \ \forall \ \alpha \in (1,2,\ldots\ldots c) \ \forall \ i \in (1,2,\ldots\ldots N)$

$$\sum_{\alpha=1}^{\alpha=C} u_{\alpha i} = 1 \quad \forall \ i=\{1,2,\ldots.N\}$$

$$0 < \sum_{i=1}^{N} u_{\alpha i} < N, \ \forall \alpha \in \{1,2,\ldots..c\}$$

At each iteration, the grades of membership and the cluster centers are updated according to the equation (4) and (5) as given below:

Each element of matrix U,

$$v_\alpha = \frac{\sum_{i=1}^{N} u_{\alpha i}^{m} .x_i}{\sum_{i=1}^{N} u_{\alpha i}^{m}} \quad \textbf{(4)}$$

$$u_{\alpha i} = \left[\sum_{\beta=1}^{C} \left(\frac{\|xi - v\alpha\|^2}{\|xi - v\beta\|^2} \right)^{\frac{1}{(m-1)}} \right]^{-1} \quad \textbf{(5)}$$

The algorithm ends when the termination criteria is met, i.e., $\left(\left\| U^{(t+1)} - U^{t} \right\| < \varepsilon \right)$ or the maximum number of iteration is achieved ($t+1 > t_{max}$). As a result, we get the matrix representing relationship between document and cluster:

$$[U]_{M\times N} = \begin{array}{c} \begin{matrix} D_1 & D_2 & \cdots & D_N \end{matrix} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix} \\ M\times N \end{array} \begin{matrix} C_1 \\ C_2 \\ \\ \\ C_M \end{matrix} \quad \textbf{(6)}$$

w*here*,

M= number of clusters (taken as rows).

N= number of documents (taken as columns).

a_{11} shows association of document D_1 in cluster C_1, while a_{MN} shows association of document D_N in cluster C_M.

A teaching domain consists of various concepts, which is defined by subject experts (Fig. 2). Suppose there are N document and k concepts in a domain model. The relationship between the concepts (as defined by experts) and documents is presented in the form of N*k matrix.

$$[R]_{N\times k} = \begin{array}{c} \begin{matrix} O_1 & O_2 & \cdots & O_K \end{matrix} \\ \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ b_{N1} & b_{N2} & \cdots & b_{Nk} \end{bmatrix} \\ N\times k \end{array} \begin{matrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ \cdot \\ D_N \end{matrix} \quad \textbf{(7)}$$

where,

N - number of documents (taken as rows).

k - number of concepts in the domain (taken as columns).

b_{11} shows association of concept O_1 in document D_1, while b_{MN} shows association of concept O_N in document D_M.

Composing equation (6) that gives us the degree of association of particular document within different clusters with equation (7) which gives us the association between concepts and documents using U o R = O, we obtain a relation (i.e. degree of association) of different concepts within different clusters. The relationship is represented in the form of the following matrix:

$$[O]_{k\times M} = \begin{array}{c} \begin{matrix} C_1 & C_2 & \cdots & C_M \end{matrix} \\ \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1M} \\ e_{21} & e_{22} & \cdots & e_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ e_{k1} & e_{k2} & \cdots & e_{kM} \end{bmatrix} \\ k\times M \end{array} \begin{matrix} O_1 \\ O_2 \\ \cdot \\ \cdot \\ \cdot \\ O_K \end{matrix} \quad \textbf{(8)}$$

where,

M- number of clusters (taken as columns).

k- number of concepts(taken as rows).

e_{11} shows association of concept O_1 in cluster C_1, while e_{MN} shows association of concept O_M in cluster C_N.

Using the equation (8) derived above, we introduce the idea of "*optimal cluster*", which may be defined as, "*The optimal cluster is a cluster which has the highest association for that particular concept i.e. it acts as the best representative for that concept".*

Suppose the user wants to learn a particular concept related to a particular domain. Then the user will be linked to the most optimal cluster i.e. the cluster which has the highest membership for that particular concept and thus the search for documents related to that concept would be limited to the documents belonging to the optimal cluster rather the entire pool of documents. This ensures an efficient searching of document related to a particular concept. The process of document retrieval is discussed in the following section.

4.0 DOCUMENT RETRIEVAL

In the proposed model for domain knowledge, we generated a matrix defining relationship between concepts and clusters i.e. we may find out clusters that are the best representative of individual concepts. The concept vector is compared to every document vector within respective optimal cluster. This identifies the document coming closest to the query and determines the rank of closeness of fit to the others. Similarity function between the concept vector and the document vector [24] in the optimal cluster is used for retrieving the documents.

Suppose the document Di is represented as a vector of dimension t, $Di = \{wi1, wi2, wi3, \ldots\ldots\ldots, wit\}$ and concept C_j as a t-dimension vector of the form $C_j = \{w_{j1}, w_{j2}, \ldots\ldots\ldots\ldots, w_{jt}\}$. The similarity between the two items can be obtained as the inner product between corresponding weighted term vector as follows [23]:

$$sim(d_j, c) = \frac{\vec{d_j} \cdot \vec{c}}{|\vec{d_j}| * |\vec{c}|}$$

$$= \frac{\sum_{i=1}^{t} w_{ij} \times w_{ic}}{\sqrt{\sum_{i=1}^{t} w_{ij}^2} \times \sqrt{\sum_{i=1}^{t} w_{ic}^2}} \quad \textbf{(9)}$$

Where, t is the number of keywords appearing in a concept.

The vectors are compared and similarity level computed for obtaining the degree of matching amongst a concept and the document. A graded collection of "best- match" document according to cosine simililarity measure is outputed to the user for each concept.

4.1 Performance Measure of Clustering Algorithms

Internal performance measure is used for evaluation of the performance of a clustering algorithm. These measures are dependent only on the algorithm and don't contain data set's structural information. Various validity indexes for the FCM-algorithm use this approach. Algorithm independent, external performance measures are used for comparing the clustering results with the benchmarks, where knowledge about cluster formation is available. The following section covers these two types of evaluation measures.

Internal Performance Measures: Validity Indices for the FCM

The document clusters produced by fuzzy c-means algorithm should be fuzzy so that uncertainty and imprecision in the knowledge space can be handled. However, we need to achieve a balance amongst the level of fuzziness, the capacity to get good clusters & the meaningfulness of document relationships. Increasing the value of the fuzzification parameter m, is known to result in fuzzier partition matrix. Thus, this parameter can be adjusted to manage this compromise. In our data set, experimentally the value of fuzzification parameter (m>1) is set to m=1.03577 to obtain pretty good clusters and meaningful document relationships. However, establishing the appropriate values for m requires the use of a validity index.

There are many validity indexes for the fuzzy c-means algorithm that are used to analyze the intrinsic quality of the clusters. Partition Entropy (PE) represents the proximity of a fuzzy partition with a hard one [3], defined as:

$$PE = -\frac{1}{N} \sum_{i=1}^{N} \sum_{\alpha=1}^{c} u_{\alpha i} \log_a(u_{\alpha i}) \quad \textbf{(10)}$$

The value of PE lies in between 0 - where U is hard - to loga(c) – where each of the data elements have same membership in each cluster ($u^{\alpha i}$ = 1/c). Dividing PE by loga(c) normalizes the value of PE to range in the [0,1] interval.

External Performance Measures: Precision, Recall, F-Measure

Two prevalent measure of evaluation of performance of information retrieval systems are Precision and recall [2]. They represent the propotion of the relevant document to the total document retrieved as a result of a query and the Propotion of retrieved documents to the number of relevant documents, respectively [12][27].

For a given cluster A and a reference cluster B, we define precision and recall as follows:

Precision = Returned relevant-documents/Total returned documents

$$P(A,B) = |A \cap B|/|A| \quad \textbf{(11)}$$

Recall = Returned relevant-documents/Total relevant-documents

$$R(A,B) = |A \cap B|/|B| \quad \textbf{(12)}$$

5.0 EXPERIMENTAL RESULTS

We performed several experiments with an objective to investigate the suitability of fuzzy clustering for discovering good document relationships and to evaluate the performance of the proposed information retrieval system. As our prime aim is to use this technique of document retrieval in e-learning applications for retrieving most relevant documents related to a particular concept, we preferred to choose data set related to teaching domain. Hence, we collected data related to Books Preview from three famous Publishers' website, namely, http://www.phindia.com, http://www.tatamcgrawhill.com, and http://www.cengageasia.com. We mainly focused on the catalogues related to Computer Science, Computer Engineering and Information Technology. From these, we extracted textual information related to books introduction and kept these as independent text files. The extracted data were from 10 different domains, namely, Operating systems, Database management system, Computer networks, Information technology, Programming Languages, Neural network, e-commerce, Bioinformatics computing, Algorithms and Software engineering. For efficiently managing the dataset, we performed the Information retrieval experiments by taking 10 documents from each domain, thereby making a total of 100 documents. The data set was pre-processed for extracting the vector space representation, before the clustering algorithm was applied. Initially the stop words were removed and subsequently stemming was done. We used the java version of Porter stemming Algorithm [22]. Frequency of each word in different documents and their respective weights were determined. This produced a very big list of keywords (or index terms). The top t maximium weighted terms were chosen from the dataset which resulted in the list of terms further being reduced to an even smaller size. The maximal weight of t_j over all 100 document is taken to obtain the maximal-weight wj of the term t_j i.e., $wj = \max_{i=1}^{100}(wij)$. Once w_j for all of these terms was obtained, the term list was sorted based on decreasing value of wj, and from this list we chose the top t. Here, for our experiment, t was kept at 100. So, every document was denoted as a vector of dimensionality of 100, after preprocessing.

Before performing the clustering, we determined the number of clusters dynamically and experimentally it was found 10 (i.e., number of cluster c=10) for the experimental data set. We used this value of c in fuzzy c-means algorithms and performed several experiments varying the value of fuzzification parameter m>1. The quality of cluster varies with the changing value of fuzzification parameter m>1 and pretty good clusters with meaningful document relationships were found with m=1.03577, while with increasing value of m the quality of clusters degrade and the clusters as a whole do not give a meaningful picture. The results were presented in the form of *document-cluster relationship matrix*.

The concepts (e.g., *Operating system*, *Computer networks*, *Database*, etc) to be taught are identified and defined by subject expert. The concept vectors were created and represented in similar manner as that of document vector of dimensionality of 100. We used cosine similarity measure for determining the degree of similarity (i.e. relevance) between document and concept. The relevance between concepts and documents were represented in the form of *concept-document relevance matrix*. The *concept-cluster relationship matrix* was obtained by composing *concept-document* and *document-cluster* relations. We used max-min composition for this purpose. The significance of concept-cluster relationship matrix lies in the fact that whenever one wants to search documents related to a particular concept, this relation tells about the most appropriate cluster (i.e. optimal cluster) with highest probability of finding the desired document related to that concept. The search is then only limited to the optimal cluster rather than to the complete pool of documents, thereby

making the process of Information retrieval more efficient.

The concept vectors are treated as query vectors for the purpose of retrieving documents matching the concept (i.e. query) vector. For a specific concept, a graded collection of "best match" document are provided to the user which is based on the similarity measures. We performed several document retrieval experiments on the experimental data set with different concept vectors and found satisfactory results in terms of *recall* and *precision* as defined in Equation (12) and (11) respectively.

6.0 CONCLUSION

Development of smart information retrieval systems is greatly required currently, when searching relevant information from a big information pile has become a bigger challenge. Additionally, the exponentinal growth of information technology, increase in the e-learning deployment and online services and the avalibility of WWW(world wide web) has inundated people with so much information. Thus, we have a great requirement of a strong automated information-retrieval-system. In this paper we have focused mainly e-learning applications and have presented an efficient method for information retrieval by determining the most optimal cluster so that a user's search for relevant documents is limited to a specific cluster rather than to the entire pool of documents. This considerably minimizes the searching time and makes the system more efficient. The domain knowledge representation and user model have also been discussed.

The domain mainly comprises of the subject or domain knowledge, skills, and procedures that the system intends to teach and assess the students' knowledge. The main role of a subject expert is to create domain knowledge base, which is the most crucial aspect and time consuming task of e-learning activities. We have proposed a novel scheme for searching documents related to concepts knowledge to be imparted to students. This may help students in efficient finding of documents related to their knowledge level (i.e. content adaptation) and also reduces the workload of subject-expert to a greater extent by helping them to synthesize course using existing learning objects available as e-learning repository. We have employed fuzzy clustering, fuzzy relation along with information retrieval techniques for discovering the underlying knowledge structure and identifying knowledge-based relationships between learning materials and retrieving the relevant documents and have implemented this technique in *eLGuide* prototype.

ACKNOWLEDGMENT

This paper was funded by Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (6-611-D1432). The authors, therefore, acknowledge with thanks DSR technical and financial support.

REFERENCES

[1] A. Arruarte, I. Fernández-Castro, B. Ferrero, J. Greer, "The IRIS Shell: How to Build ITSs from Pedagogical and Design Requisites", *International Journal of Artificial Intelligence in Education*, Vol. 8, No. 3/4, 1997, pp. 341-381.

[2] R. Baeza-Yates, B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley, ACM Press, New York, 1999.

[3] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[4] K. I. Ghauth, N. A. Abdullah, "An Empirical Evaluation of Learner Performance in E-Learning Rec-

ommender Systems and an Adaptive Hypermedia System", *Malaysian Journal of Computer Science*, Vol. 23, No. 3, 2010, pp 141-152.

[5] J.C. Bezdek, R.J. Hathaway, M.J. Sabin, W.T. Tucker, "Convergence theory for fuzzy c-Means: counterexamples and repairs", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 17, No. 5, 1987, pp. 873-877.

[6] F. Chang, "Quantitative Analysis of Distance Learning Courseware", *Multimedia Tools and Applications*, Kluwer Academic Publishers. Printed in the Netherlands. Vol. 20, 2003, pp. 51–65.

[7] S.M. Cherry, "Weaving a web of ideas", *IEEE Spectrum*, Vol. 39, No. 9, 2002, pp. 65 -69.

[8] E. Delozanne, B. Grugeon, D. Prévit, P. Jacoboni, "Supporting Teachers When Diagnosing Their Students in Algebra", *In É. Delozanne, & K. Stacey (Eds.), Workshop Advanced Technologies for Mathematics Education, Supplementary Proceedings of Artificial Intelligence in Education*, Sydney, IOS Press, Amsterdam, 2003, pp. 461-470.

[9] J. Elorriaga, A. Arruarte, I. Fernandeze-Castro, "Increasing Teachers' Participation in ITS Pedagogical Decisions", *Proceedings of International Conference on Educational Uses of Communication and Information Technology, ICEUT'2000*, 2000, pp. 453-460.

[10] A. Kavcic, "Fuzzy user modeling for adaptation in educational hypermedia", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 34, No. 4, Nov 2004, pp. 439 – 449.

[11] D.H. Kraft, J. Chen, A. Mikulcic, "Combining fuzzy clustering and fuzzy inference in information retrieval", *Proceedings of the 9th IEEE International Conference on Fuzzy Systems, FUZZ IEEE 2000*, Vol. 1, May 2000, pp. 375-380.

[12] V. Loia, M. Nikravesh, L.A. Zadeh (Eds.), *Fuzzy Logic and the Internet*. Springer, 2004.

[13] J.B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5-th Berkeley Symposium on Mathematics, Statistics and Probability*, Berkeley, University of California Press, Vol.1, 1967, pp. 281-296.

[14] R. Mazza, V. Dimitrova, "Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education", *Proceedings of 13th International Conference on World Wide Web (WWW'04)*, 2004, pp. 154-161.

[15] M.E.S. Mendes, L. Sacks, "Dynamic Knowledge Representation for E-learning Applications", *Proceedings of the 2001 BISC International Workshop on Fuzzy Logic and the Internet, FLINT'2001*, University of California Berkeley, Aug 2001, pp. 176-181.

[16] A. Merceron, K.Yacef, "A Web-Based Tutoring Tool with Mining Facilities to Improve Learning and Teaching", In U. Hoppe, F. Verdejo, & J. Kay (Eds.*), Proceedings of the 11th International Conference on Artificial Intelligence in Education*, Sydney, Australia, IOS Press, 2003, pp. 201-208.

[17] W. A. Banu, P.Sheik Abdul Khader , R. Shriram, "Mobile Ontology Design for Semantic Web: A Case Study", *Malaysian Journal of Computer Science*, Vol. 24, No. 4, 2011, pp 205-216.

[18] M.Nikravesh, **"**Concept-based Semantic Web Search and Q&A", *Studies in Computational Intelligence (SCI)*, Vol. 37, 2007, pp. 95-124.

[19] M. Nikravesh, M. T. Takagi, "Enhancing the Power of the Internet", *Studies in Fuzziness and Soft Computing, M Nikravesh, B Azvine, R Yagar, LA Zadeh (Eds)*, Physica-Verlag, Springer, 2003, pp.1-20.

[20] S. Parsons, "Current approaches to handling imperfect information in data and knowledge Bases", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 3, 1996, pp. 53-372.

[21] P. Brusilovsky, E. Milla, "User Models for Adaptive Hypermedia and Adaptive Educational Systems", *The Adaptive Web*, Springer Berlin / Heidelberg, 2007, pp. 3-53.

[22] M. Porter, "An algorithm for suffix stripping", *Program*, Vol. 14, No. 3, Jul. 1980, pp.130-137.

[23] J. Qin, N. Hernández, "Building interoperable vocabulary and structures for learning objects", *Journal of the American Society for Information Science and Technology*, Vol. 57, No. 2, 2006, pp. 280-292.

[24] M. Moohebat, R.G.. Raj, S. B. A. Kareem, and D. Thorleuchter, "Identifying ISI-indexed articles by their lexical usage: A text analysis approach", *Journal of the Association for Information Science and Technology*, Vol. 66, No. 3, pp. 501–511. doi: 10.1002/asi.23194.

[25] C.J. Rijsbergen, *Information Retrieval. 2nd Edition*, Butterworth-Heinemann Newton, MA, USA, 1979.

[26] W.L.Yeow, R. Mahmud, R.G. Raj., "An application of case-based reasoning with machine learning for forensic autopsy", *Expert Systems with Applications*, Vol 41, No. 7, 2014, pp. 3497-3505, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2013.10.054. (http://www.sciencedirect.com/science/article/pii/S0957 417413008713).

[27] L. B. Huang, V. Balakrishnan, R.G. Raj, "Improving the relevancy of document search using the multi-term adjacency keyword-order model." *Malaysian Journal of Computer Science*, Vol. 25, No. 1, 2012, pp. 1-10.

[28] O. Santos, A. Rodríguez, E. Gaudioso, J. Boticario, "Helping the Tutor to Manage a Collaborative Task in a Web-based Learning Environment", *In R. Calvo, & M. Grandbastien (Eds.), Workshop of Intelligent Management Systems, Supplementary Proceedings of Artificial Intelligence in Education*, Sydney, 2003, pp. 153-162.

[29] J.R. Timothy, *Fuzzy Logic with Engineering Applications*. McGraw-Hill, Inc, 1995.

[30] P. Willett, "Recent trends in hierarchical document clustering: a critical review", *Information Processing and Management*, Vol. 24, No. 5, 1988, pp. 577-597.

[31] L.A. Zadeh, "Fuzzy Sets", *Information and Control*, Vol. 8, 1965, pp. 338-353.

[32] O. Zamir, O. Etzioni, "Grouper: a dynamic clustering interface to Web search results", *Computer Networks*, Vol. 31, No. 11-16, May 1999, pp 1361-1374.