Journal of Information Technology and Application in Education Vol. 3 Iss. 1, March 2014 www.jitae.org
doi: 10.14355/jitae.2014.0301.02

Similarity Based Convergence of Learning Knowledge Objects and Delivery Using Agglomerative Clustering

A.Sai Sabitha[1], Deepti Mehrotra[2], Abhay Bansal[3]

ASET, Amity University[1], Amity School of Computer Sciences[2], Amity University[3], INDIA, INDIA

[1]saisabitha@gmail.com; [2]mehdeepti@gmail.com; [3]abhaybansal@hotmail.com

Abstract

E-learning today shows an exponential growth and there is a need to develop more flexible delivery processes, which will add value to the learning experiences of a student during his/her learning process. The atomic unit of any e-learning environment is a Learning Object, a reusable digital entity. These Learning Objects are stored in repositories and managed by Learning Management Systems in order to provide a better coverage of concepts to the learner. A close relative of Learning Object is the Knowledge Object which is an essential unit of a Knowledge Management System. In this research paper, a way is proposed to converge these objects together and most similar Learning Knowledge Objects delivered to a student using hierarchical clustering techniques. The learning experience is more valuable for a student especially for those with higher order thinking skills.

Keywords

Learning Objects; Knowledge Objects; Clustering; Learning Knowledge Object; Hierarchical Clustering; Agglomerative Clustering

Introduction

One of the biggest impacts of technology on the education domain is the development of the Learning Management System (LMS) which is capable of storing, sharing and imparting knowledge to the learner. Almost all research and academic institutions own and archive a great number of documents like lesson plans, studying material and research related resources. They are stored and used for a longer period of time by lecturers, researchers and students to enhance their academic knowledge. The potential educational value of these resources is very high. Some of these resources have already been converted into Learning Objects (LO), are stored and structured in a meaningful way in an LMS, thus enriching classical teaching. Experienced faculty has in depth knowledge in a particular domain. This tactic knowledge is collected by knowledge management tools and can be stored in a knowledge base of a KnowledgeManagementSystem (KMS). These valuable experience captured helps to retain the knowledge of an expert, even when he leaves the organisation. This unstructured tacit knowledge can be converted into structured knowledge, by adding an objective, and metadata to it. This Knowledge Object (KO) and the Learning Object forms a digital entity. These can be converged and made available while teaching a module in a curriculum. Thus, the learning material defined for a particular subject and the in depth knowledge of experts in the same subject can be combined and delivered as one whole unit to the students while learning. This can be considered as Learning Knowledge Object (LKO). The objects are delivered based on various attributes like publisher, domain, title, education level, type etc. Text similarity matching of LO & KO can be very useful for identifying the different LKO having similar content. In this paper, data mining techniques are used to group objects based on similarity measures. Also hierarchical clustering algorithm is used, so that a reduced set of relevant objects are delivered.

Literature Survey

A LO, according to the experts is a reusable digital entity used as one of the core elements of the LMS. These LOs are structured in a meaningful way and have a Learning Objective (IEEE LTSC, 2002; Barron 2000; Wiley 2000; Gallenson 2002). A LO refers to any digital educational resource and lesson provided during a technology-supported learning. Many structures of the LOs have been proposed and its key features are: - a) Learning objective b) Metadata c) Assessment d) Performance goal (Wagner 2002; SCORM, 2005; Mortimer, L.2002; Gallenson, 2002; Metros, 2002-03). Basically LO comprises an asset (image, text, video, web page) and an information object that teaches a single concept. The smallest level of granularity of an object can be a picture or a text and at the largest level can be a set of courses (Wiley & Gibbons, 2000).

A metadata is a structured information that helps to fetch and manage a LO, it provides the description that facilitates and administrates LMS. The various metadata standards for Learning Objects are Dublin Core Metadata (Dublin core, 2012), IEEE Learning Object Metadata (IEEE LTSC), IMS Global Learning Consortium (IMS, 2006).

Knowledge Object is defined as a record of information that serves as a building block in a KMS. It has content, knowledge base, rules to identify and categorise knowledge components (Merrill, 2006). Horton (2001) said, "A KO is an electronic content, that should consist of a goal, content, metadata and security information". A KO is a tightly integrated bundle of ideas, related information and experiences. KOs are closely related to LOs or a part of a LO (Ruffner & Nina, 2008).

An online delivery of complex LOs, results in high quality outcome for students (Ellis 2006). A new knowledge-based model for representing LO instances can shift the reuse dimension from component-based to generative reuse in LO domain (Štuikys, 2007).

A Knowledge Puzzle Content Model which creates pertinent LKOs that can be adapted to fit a particular instructional theory and a particular learner model was proposed by (Zouaq 2007). Ontology of learning object repository (LOR) can be used for pedagogical knowledge sharing (Wang, 2008). By exploiting KMSs functionalities in LORs would bare the potential to support the organization and sharing of educational communities' explicit knowledge (Demetrios 2013).

The e-Learning system can be used to enhance knowledge management in an organization and provide the benefits of both (Walid Qassim 2011).

Data mining techniques facilitate organisation, extraction and delivery of the LO. A cluster analysis is a collection of statistical methods that identifies a group of samples that behave similarly or show a similar characteristic. The clustering algorithms are broadly classified as hierarchical and non-hierarchical algorithms. In a non-hierarchical clustering method, the clusters are constructed according to the distance measured from a central point. K-mean, K-medoid are some examples of it. In a hierarchical approach, clusters are generated by iteratively merging the sub clusters and generate a tree like structure that reflects the relationship between entities. Agglomerative Nesting (AGNES) and Divisive Analysis (DIANA) are the two basic types of hierarchical clustering (Han, 2008).

An agglomerative algorithm works by grouping the data one by one on the basis of nearest distance measure of all pairwise distances between the data point. Unlike k-mean, we need not specify the number of clusters in advance. The hierarchical agglomerative approach of clustering plays a vital role in domains such as medical, bioinformatics, information retrieval taxonomy etc.

Clusters are visualised as a dendogram, as shown in Fig. 1. A dendogram decomposes data objects into several levels of nested partitioning. A desired level is chosen along the y-axis of the dendogram. If we cut across the dendogram, the number of lines we cross represents the number of groups that are identified when objects are clustered. Each merge is shown by a horizontal line. We can view the data at various levels of granularity.

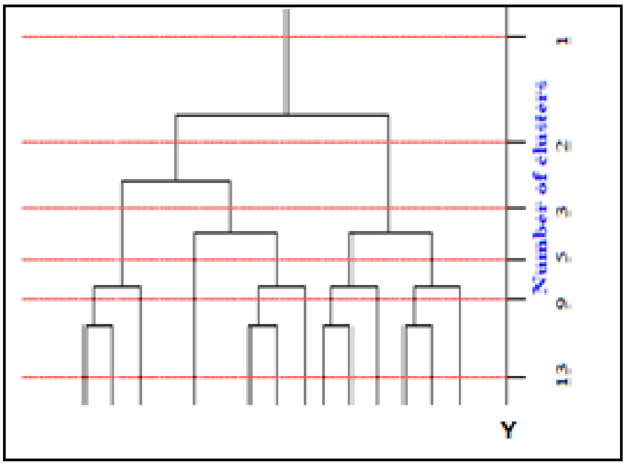

FIG. 1 A SIMPLE DENDOGRAM

Need for a LKO

LMSs mainly helps in centralise and automate training, assemble and deliver learning content, and enables reuse of learning modules. The system delivers the content framed by a teacher or a contributor. Today, a more knowledge enriched learning is needed by the user. Ideally, the material that the learner receives for a topic in a subject must be blended with enhanced knowledge. It can be saved in the form of KO in the knowledge base of KMS. There is a need to converge the LO & KO so that enhanced LKO can be delivered to the user, thus making an effective use of various distributed knowledge. LKO helps the student to get an extra edge of learning and prepare for various forms of assessment thereby improving their cognitive skills. The LKO engages the learner to actively explore, reflect and produce knowledge rather than recall and regurgitate. Appending KO to LO, knowledge-pull occurs in a learning system, thereby enriching the learning experience of a learner.

Method

Creation of LKO

Step 1. Creation of LMS, inclusion of LOs and metadata for it.

Step 2. Creation of KMS, inclusion of KOs objects and metadata for it.

Step 3. Both the objects have metadata. Cleaning of data is required, before the convergence of LO & KO.

Step 4. A set of common attributes can be considered for both LO & KO. Based on the user query with respect to a domain and topic, a set of LO & KO are retrieved from LMS & KMS respectively and can be classified using classification techniques like a decision tree (Sabitha, Mehotra, 2013) or a naïve Bayes algorithm etc., as in Fig. 2.

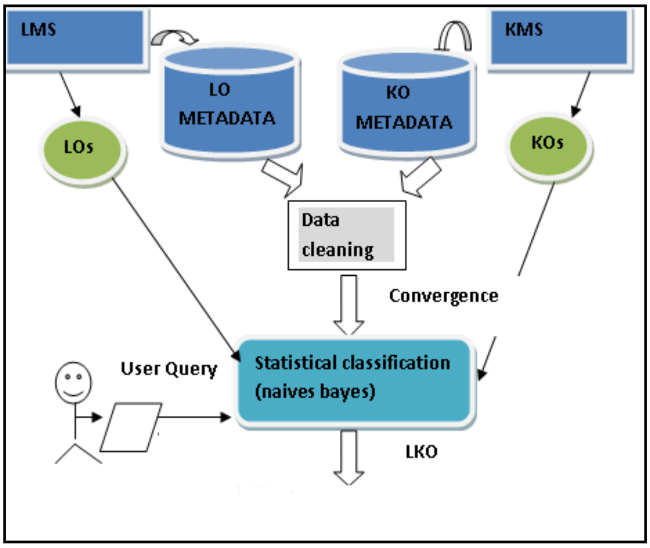

FIG. 2 CLASSIFCATION OF OBJECTS

Convergence of LO & KO as LKO

The prime concern to obtain the complete knowledge is by converging LO & KO. The metadata of LO & KO facilitates convergence. The Dublin Core metadata standard for LO is considered in the LMS. The LO of LMS possesses 15 attributes and some of these attributes are title, subject, contributor, date created, type etc., Although there are no well defined structure of metadata for a KO, attributes like title, domain, author, date created, time, knowledge source, patent, types of knowledge are used in the KMS. The LO & KO can be combined as an object called LKO. Many ways are proposed for the formation of a LKO. Extracting document content (LO) to construct a concept map for each document and by identifying the important concepts and relations between them, results in a LKO (Zouaq1, 2007). An ontology model was proposed to generate LKOs based on instructional theories (Wang, 2008). For the formation of LKO, it is proposed that KO can be classified with LO through classification technique (Sabitha, Mehotra 2013). These LKOs are a self-contained instructional unit and can be delivered to learners (who have the basic prerequisite) to improve their pedagogical learning experience. According to Hodgins (2002), a relevant LO delivered to the user is one of the key strategies in learning. A knowledge based generative object model can be generated by factoring and aggregating knowledge units within a LO (Vytautas 2007).

Proposed Models for Delivery of LKO

The model of delivery is given in Fig. 3. The classified objects as shown in Fig. 2 is considered by hierarchical clustering engine. The clusters are formed for the LKOs based on the cosine similarity coefficient. Validity measures like BSS, F-Score can be used to evaluate the best clusters (Vipin 2007). The advance learners have some prior knowledge and look for additional or enhanced knowledge. The clusters with fewer LKOs are delivered to these learners. Thus, the objective of delivering is reduced and more relevant contents are given to advance learners.

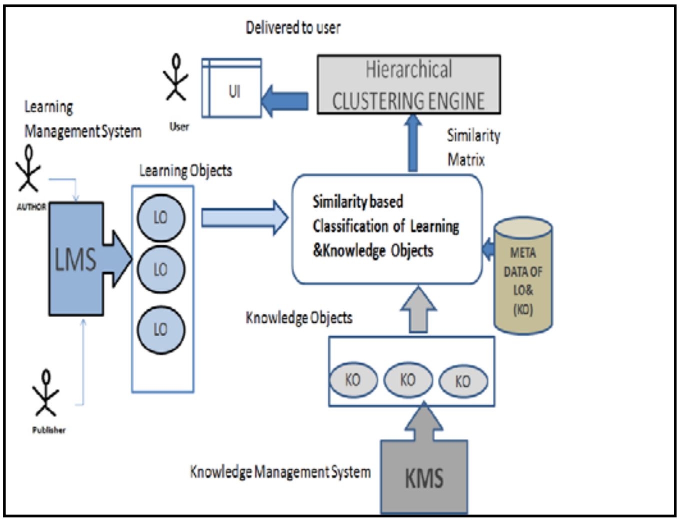

FIG. 3 DELIVERY OF LKO

Agglomerative Algorithm

Distance is calculated based on: -

1) Single link (minimum): The distance between two

clusters are defined as the smallest dissimilarity between the points.

2) Complete link: The distance between two clusters are defined as largest dissimilarity (smallest similarity) between the points.

3) Average link (average): The distance between clusters are calculated as the weighted average dissimilarity (or similarity) value.

4) Ward Criteria: It is an alternative technique assuming that a cluster is represented by its centroid. Proximity is measured in terms of sum of squared errors (SSE), so as to minimise the sum of squared distances of the points from the cluster centroid.

The agglomerative algorithm is as follows:-

Compute the proximity matrix (similarity)

Repeat

Merge the 2 clusters which are closer.

Update the proximity matrix to show proximity between the new cluster and original cluster.

Until one cluster exists.

Cluster Validity Measures

The cosine similarity was used to measure the object content instead of their metadata.

- Keyword analysis for the clustered objects are performed.
- Statistical evaluation measures for cluster validity like compactness and separation are used.

Compactness: The objects of each cluster should be as close to each other as possible and the measure is variance.

Separation: The clusters themselves should be widely spaced. The measures are within group sum of squares (WSS), between groups sum of squares (BSS) and total sum of squares (TSS).BSS and gap ratio are considered.

Experimental Set-up

Step 1: A set of 60 LOs & 15 KOs are considered and these are classified objects based on user query for domain "data mining" (refer Fig. 2). The dataset of classified objects are shown in Fig. 4.

Step 2: To calculate the proximity (similarity) between two objects measures like euclidean distance for two dimensional points are used. Sparse data like documents, uses Jacquard and cosine similarity measure. Here the content of objects are considered to find the proximity instead of their metadata, so that the cosine measure is taken.

KEY	OBJECT	TOPIC	SUBTOPIC	CATEGORY	CONTENT
1	LO1	INTRODUCTION	DATA MINING	DEFINITION	Data mining (knowledge discovery from data) Extraction of interesting (non trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
2	LO2	INTRODUCTION	KNOWLEDGE DISCOVERY	DEFINITION	data mining is a step in Knowledge discovery process
3	LO3	INTRODUCTION	DATA WAREHOUSE	DEFINITION	data warehouse is a repository which stores data from various soucres in an unified format
4	LO4	KNOWLEDGE DISCOVERY	PATTERN EVALUATION	DEFINITION	A pattern is interesting if it is easily understood by humans, valid on new or test data with some degree of certainty, potentially useful, novel, or validates
5	LO5	PRE PROCESSING	DATA CLEANING	DEFINITION	Data in the real world is dirty ,inconsistent and noisy. So cleaning is required
6	LO6	PRE PROCESSING	DATA INTEGRATIO N	DEFINITION	Integration of multiple databases, data cubes, or files Combines data from multiple sources into a coherent store

FIG. 4 DATASET OF LKO

Cosine similarity is calculated as follows:-

Given two vectors of attributes, A and B, the cosine similarity, $cos\ (\theta)$, is represented using a dot product and magnitude:-

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Step3: The similarity between the content of the objects (refer Fig. 4) are generated and are shown in table 1:-

TABLE 1 SIMILARITY MEASURES

Object1	Object2	Similarity Measure
1	2	0.2877408
1	3	0.04241506
1	4	0.12719396
1	5	0.01491909
1	6	0.06201117
1	7	0
1	8	0.03396251
1	9	0.12719396
1	10	0.03076241
1	11	0.08882356
1	12	0.10597134
1	13	0.07352384
1	14	0.02658609
1	15	0.2022802
1	16	0.01837801
1	17	0.02214211
1	18	0.03992946
1	19	0.05474195
1	20	0.07119731

Step 4: Based on the similarity values given in Table 1, a 75 X 75 matrix is generated and is shown in Fig. 5.

	1	2	3	4	5	6	7	8	9	10
1	0	0.2877	0.0424	0.1272	0.0149	0.062	0	0.03396251	0.12719396	0.03076241
2	0.2877	0	0.0466	0.0297	0.0448	0.0229	0.0093	0.05454494	0.02967865	0.07815726
3	0.0424	0.0466	0	0.022	0.0332	0.0392	0.0059	0.01825525	0.02202368	0.04826957
4	0.1272	0.0297	0.022	0	0.0218	0.0416	0.0087	0.01766519	1	0.03930328
5	0.0149	0.0448	0.0332	0.0218	0	0.0103	0.1297	0.01978079	0.02182096	0.03857769
6	0.062	0.0229	0.0392	0.0416	0.0103	0	0.0061	0.04134843	0.04155542	0.03295555
7	0	0.0093	0.0059	0.0087	0.1297	0.0061	0	0.01986865	0.00866214	0.0194336
8	0.034	0.0545	0.0183	0.0177	0.0198	0.0413	0.0199	0	0.01766519	0.02105181
9	0.1272	0.0297	0.022	1	0.0218	0.0416	0.0087	0.01766519	0	0.03930328
10	0.0308	0.0782	0.0483	0.0393	0.0386	0.033	0.0194	0.02105181	0.03930328	0
11	0.0888	0.0164	0.015	0.0562	0.0177	0.0134	0	0.02581227	0.05619985	0.01534859
12	0.106	0.0054	0.0034	0.0646	0.014	0.0183	0.0093	0.01123449	0.06458372	0.03461639
13	0.0735	0	0	0.0296	0.0072	0.021	0	0.0104538	0.02959397	0.01240587
14	0.0266	0.0151	0.0191	0.0584	0.0177	0.0234	0.0079	0.02053739	0.058447	0.01331287
15	0.2023	0.2055	0	0.0373	0	0.0053	0	0.00944218	0.03732981	0.00689354
16	0.0184	0.017	0.0354	0.0393	0.0286	0.0156	0.0102	0.01290971	0.03927531	0.08944917

FIG. 5 MATRIX OF LKO

Step 5: To generate clusters, data mining tool is used. The similarity matrix is loaded as shown in Fig. 6 and hierarchical clustering operators are used.

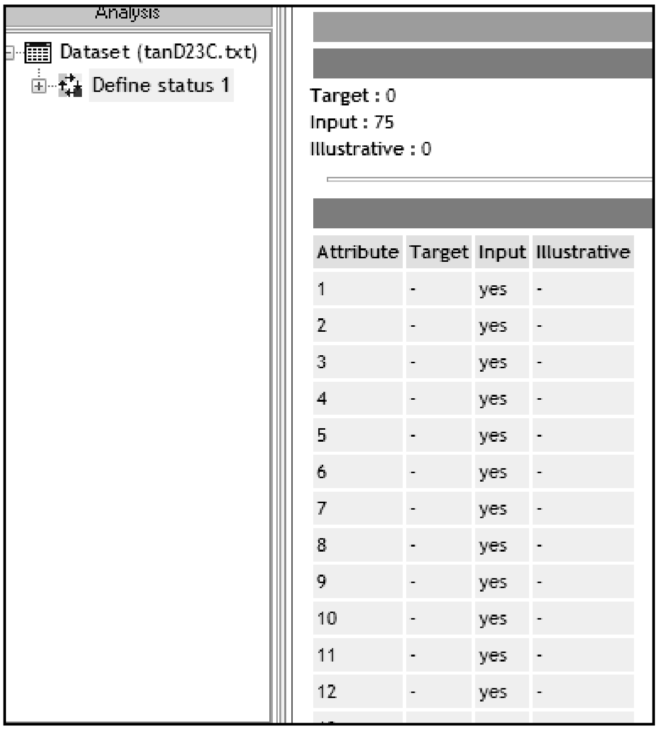

FIG. 6 MATRIX LOADED IN TOOL

Output

Clustering Results

A total of seven clusters is formed as shown in Fig. 7. Ward's method is used. It also has the same objective function as K-mean clustering. The initial cluster distances in ward's minimum variance method are defined to be the squared euclidean distance between points.

$$dist(X,Y)=\sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2}$$

Clustering results

Clusters	7	
Cluster	Description	Size
cluster n°1	c_hac_1	5
cluster n°2	c_hac_2	7
cluster n°3	c_hac_3	30
cluster n°4	c_hac_4	9
cluster n°5	c_hac_5	12
cluster n°6	c_hac_6	2
cluster n°7	c_hac_7	10

FIG. 7 SIZE OF CLUSTERS

Objects and Clusters

The object id's and their corresponding clusterid's are shown below in Fig. 8a.

For example:-

- The contents of the objects of cluster number: "4" in Fig. 8b are considered for validation and almost all objects belong to topic "association" for the domain "data mining".
- The contents of the objects of cluster number: "1" in Fig. 8c have all object belong to the topic "data integration" for the domain "data mining".

id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no	id	Cluster no
6	c_hac_1	54	c_hac_2	18	c_hac_3	42	c_hac_3	61	c_hac_3	53	c_hac_4	15	c_hac_5	75	c_hac_6	72	c_hac_7
27	c_hac_1	55	c_hac_2	19	c_hac_3	45	c_hac_3	64	c_hac_3	56	c_hac_4	20	c_hac_5	12	c_hac_7	74	c_hac_7
38	c_hac_1	67	c_hac_2	26	c_hac_3	46	c_hac_3	65	c_hac_3	58	c_hac_4	21	c_hac_5	30	c_hac_7		
49	c_hac_1	3	c_hac_3	28	c_hac_3	47	c_hac_3	69	c_hac_3	66	c_hac_4	22	c_hac_5	39	c_hac_7		
62	c_hac_1	5	c_hac_3	29	c_hac_3	48	c_hac_3	70	c_hac_3	71	c_hac_4	23	c_hac_5	43	c_hac_7		
31	c_hac_2	7	c_hac_3	34	c_hac_3	50	c_hac_3	10	c_hac_4	1	c_hac_5	24	c_hac_5	44	c_hac_7		
32	c_hac_2	11	c_hac_3	37	c_hac_3	51	c_hac_3	16	c_hac_4	2	c_hac_5	25	c_hac_5	59	c_hac_7		
33	c_hac_2	13	c_hac_3	40	c_hac_3	57	c_hac_3	17	c_hac_4	4	c_hac_5	36	c_hac_5	63	c_hac_7		
35	c_hac_2	14	c_hac_3	41	c_hac_3	60	c_hac_3	52	c_hac_4	9	c_hac_5	73	c_hac_6	68	c_hac_7		

FIG.8a OBJECT IDS AND CLUSTER NUMBERS

16	LO16	OBJECTIVE M	SUPPORT	DEFINITIO	SAI SABITH	An itemset satis es minimum support if the occurrence frequency of the itemset is greater than or equal to
10	LO10	DATA MININ	ASSOIATIO	DEFINITIO	NIDHI	Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on
17	LO17	OBJECTIVE M	CONFIDEN	DEFINITIO	SUNITA	Confidence can be interpreted as an estimate of the probability P(Y|X), the probability of finding the RHS of the rule in transactions under the condition that these transactions also
62	LO62	ASSOCIATIO	hashing	concept	SEEMA	hashing is a technique in association mining.A hash-based technique can be used to reduce the size of the candidate k-itemsets, Ck, for k > 1. For example,
63	LO63	ASSOCIATIO	APRIORI	technique	SEEMA	Apriori is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to count the support of itemsets and uses a candidate
68	LO68	ASSOCIATIO	FPTREE	technique	SEEMA	FP stands for frequent pattern. In the first pass, the algorithm counts occurrence of items (attribute-value pairs) in the
66	KO6	ASSOCIATIO	APRIORI	CONCEPT	SAI SABITH	it is an algorithm used in market analysis. It uses two measures support and confidence
71	KO11	ASSOCIATIO	FPTREE	technique	EE	FP stands for frequent pattern.

FIG. 8b CONTENTS OF CLUSTER NUMBER: 4

L06	PRE PROCESSING	DATA INTEGRATION	DEFINITION	SUNITA	Integration of multiple databases, data cubes, or files Combines data from multiple sources into a coherent store
L038	DATA INTEGRATION	SCHEMA INTEGRATION	CONCEPT	SUNITA	Entity identification problem: identify real world entities from multiple data sources, e.g., A.cust-id B.cust-#
L049	DATA INTEGRATION	SCHEMA INTEGRATION	technique	SEEMA	Schema integration: e.g., A.custid B.custno.Integrate metadata from different sources Entity identification problem: Identify real world entities from multiple data sources, e.g., Bill Clinton –
K02	DATA INTEGRATION	SCHEMA INTEGRATION	CONCEPT	LEENU	Data integration: Combines data from multiple sources into a coherent store Schema integration A.cust-id B.cusno
L027	PRE PROCESSING	DATA INTEGRATION	TYPES	SUNITA	coherent data store. Metadata, correlation analysis, data conflict detection

FIG. 8c CONTENTS OF CLUSTER NUMBER: 1

Cluster Levels

66	1	-	-
67	1	-	-
68	1	-	-
69	1	-	-
70	1	-	-
71	1	-	-
72	1	-	-
73	1	-	-
74	1	-	-
75	1	-	-
76	2	(29,51)	0.0824
77	2	(3,61)	0.1455
78	3	(14,77)	0.1519
79	2	(18,19)	0.2280
80	4	(5,78)	0.2639
81	2	(20,23)	0.2754
82	2	(21,25)	0.2757
83	2	(33,35)	0.2978
84	2	(13,47)	0.3034

FIG. 9 CLUSTER LEVELS

Fig. 9, shows

- At node '76', the child with object id '29' and object id '51' are clustered and belong to cluster '3'.

  As shown in Fig. 9a, the node '76' contains the objects under the topic "data transformation" for the domain "data mining"
- At node '77', the child with object id '3' and object id '61' are merged and belong to cluster '3;.
- At node '78', child with object id '14' and node '77' are merged.

KEY	OBJECT	TOPIC	SUBTOPIC	CATEGORY	AUTHOR	CONTENT
29	L029	PRE PROCESSING	DATA TRANSFORMATION	TYPES	SAI SABITHA	Transformation of data.Data transformation techniques are Normalization and Aggregation,Smoothing,generalisation
51	L051	DATA TRANSFORMATION	NORMALISATION	TECHNIQUES	SEEMA	min-max,z-score and decimal scaling are the tech

Fig. 9a NODE 76

Best Cluster Selection

Since hierarchical clustering algorithms discover clusters which are not known prior, the final partitions of a data set requires some evaluation in most applications. Finding the quality of the clustering is necessary and it is measured under three approaches namely external criteria, internal criteria (proximity matrix) and relative criteria. Cluster cohesion, measures how closely are objects related in a cluster (SSE) and a cluster separation measures, how distinct or well separated a cluster is from other clusters (squared error). Separation is measured by the between sum of squares (BSS). The formula of BSS is appended below:-

$$\text{BSS} = \sum_{i=1}^{K} |C_i| \; (m - m_i)^2$$

where |Ci| is the size of a cluster.

The compactness of the data is measured by the gap value while higher value of gap means good clustering. A good clustering also has high BSS ratio. The dissimilarity between clusters are considered by the BSS ratio. A high BSS and gap values are at cluster 7 as shown in Fig. 10.

Best cluster selection

Clusters	BSS ratio	Gap
1	0.0000	0.0000
2	0.0668	1.4711
3	0.1139	0.1090
4	0.1596	0.1314
5	0.2036	0.3781
6	0.2425	0.2859
7	0.2776	0.5362
8	0.3055	0.0332
9	0.3330	0.1326
10	0.3587	0.0080
11	0.3844	0.0949
12	0.4087	0.1933
13	0.4305	0.1936
14	0.4497	0.0192
15	0.4687	0.0125
16	0.4875	0.0750

FIG. 10 BEST CLUSTER SELECTION

Dendogram

The clustering of data objects is obtained by cutting the dendogram at a desired point along the Y-axis. The point can be taken for BSS as 0,0.2776 and for gap ratio it is 0,0.5362. The line that cut across in the dendogram is shown in Fig. 11 and crossing seven lines represents seven groups that are identified when objects are clustered.

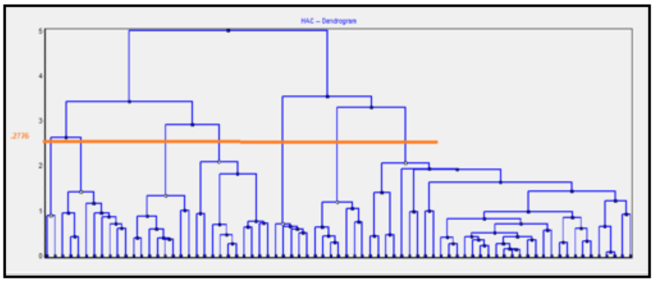

FIG. 11 DENDOGRAM

Validation

Keyword Search

The following figure shows the objects retrieved for a word –"support". Objects 10, 16, 52,53 and 66 belong to cluster 4.

53	Apriori is the best-known algorithm to mine association	objects	ob_153.t	C:\sabith	Wed Aug	53	0.262521
66	it is an algorithm used in market analysis. It uses two	objects	ok_166.tx	C:\sabith	Wed Aug	66	0.243722
30	Objective: based on statistics and structures of patterns,	objects	ob_130.t	C:\sabith	Wed Aug	30	0.15729
16	An itemset satis es minimum support if the occurrence	objects	ob_116.t	C:\sabith	Wed Aug	16	0.146817
10	Association analysis is the discovery of what are	objects	ob_110.t	C:\sabith	Fri Aug	10	0.087297
52	ashing is a technique in association mining.A hash-based	objects	ob_152.t	C:\sabith	Wed Aug	52	0.054204

FIG. 12 KEY WORD SEARCH

Validity Based on Compactness

The measure used is variance. The total variance of the entire objects is 0.292994. The total variance of objects of cluster number 4 is 0.211. Variance and the scatter plot of cluster number 4 are shown in Fig. 13a, Fig. 13b respectively.

	10	16	17	52	53	56	58	66	71	
10	0	0.0452	0.067	0.121433	0.1166	0.062217	0.112204	0.086795	0.112204	
16	0.045	0	0.086	0.057845	0.0275	0.016035	0.058761	0.024116	0.058761	
17	0.067	0.086	0	0.060008	0.0166	0.007582	0.073818	0.034557	0.073818	
52	0.121	0.0578	0.06	0	0.1024	0.114554	0.153114	0.021285	0.153114	
53	0.117	0.0275	0.017	0.102447	0	0.072368	0.035846	0.148419	0.035846	
56	0.062	0.016	0.008	0.114554	0.0724	0	0.04407	0.065791	0.04407	
58	0.112	0.0588	0.074	0.153114	0.0358	0.04407	0	0.035078	1	
66	0.087	0.0241	0.035	0.021285	0.1484	0.065791	0.035078	0	0.035078	
71	0.112	0.0588	0.074	0.153114	0.0358	0.04407	1	0.035078	0	
VARIANC	0.002	0.0007	0.001	0.003049	0.0026	0.001316	0.099385	0.001996	0.099385	0.21113

FIG. 13 a VARIANCE OF CLUSTER 4

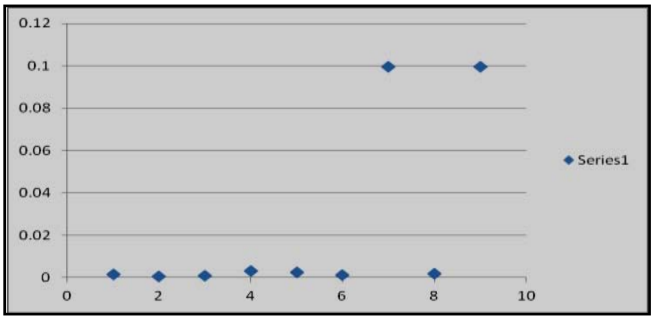

FIG. 13b SCATTER PLOT

Validity Based on Separation

OBJECTS	BSS	WSS	TSS
10	0.0523	0.0531	0.1054
16	0.0346	0.0329	0.0675
17	0.0311	0.047	0.0781
52	0.0588	0.0511	0.1099
53	0.0357	0.0727	0.1084
56	0.0244	0.1134	0.1378
58	0.1929	0.7903	0.9832
66	0.0548	0.0947	0.1495
71	0.1929	0.7903	0.9832

FIG. 14a WSS, BSS OF CLUSTER4

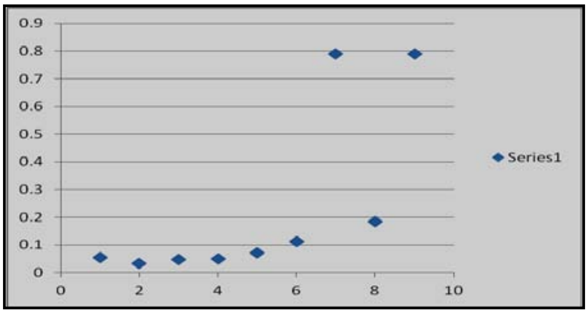

FIG. 14b SCATTER PLOT OF WSS

Conclusions

The constant evaluation of a LO helps in high quality of web-based education. Many features are considered in the creation and delivery of LO like presentation, content accuracy, learning goals, motivation, interaction, reusability and accessibility. The quality of learning content delivered to a user also plays an important role. Determination of the quality of the Learning Objects is a challenge in an e-Learning environment. Though there are thousands of repositories for Learning Objects, the quality of retrieved resources varies and the processes involved are not an easy task. A way to improve the content of LOs is proposed by converging it with KOs. Granularity in the context a LO is often used to refer to the size of an object. Considering the size of granularity a few closely related objects could be delivered. The hierarchical clustering (ward's) produces smaller size, reasonable clusters. It minimises the total within the cluster variance. The idea of using the above technique is to deliver the closely related LKOs. To reduce the computational cost, we can firstly use K-mean clustering and then perform hierarchical clustering on these small clusters. We can rank the objects (Xavier Ochoa & Duval, 2008; Sabitha & Mehotra, 2012) and the best ranked objects can be classified. Thus a reduced set of objects are considered for clustering. The hierarchical clustering provides an advantage by letting the learner to select a number of cluster wanted in the output, rather than fixed as a case of K-mean clustering.

Future Scope

The actual LMS has a mixed type of metadata i.e., Metadata may have attribute values that are numeric or categorical. To consider both parameters as a criteria for hierarchical agglomerative clustering method, we can use the spread (data distribution) of clusters as the merging criteria. For numeric attributes, variance and for categorical attributes, entropy has been used to measure the data distribution (spread) in a cluster. The current method does not take care of usage of the objects, but the quality of the new method can be rated with the help of user feedback and the rank of LKO can be generated and used as a quality metric for clustering the objects. Further, clustering based on a similarity index can be used in converging two LMS and developing a suitable semantic LOR.

Abbreviations

LO: Learning Object

KO: Knowledge Object

LMS: Learning Management System

KMS: Knowledge Management System

LKO: Learning Knowledge Object

LOR: Learning Object Repository

BSS : Between Sum of Squares

WSS: Within Sum of Squares

SSE : Sum of Squared Errors

AGNES: Agglomerative Nesting

DIANA: Divisive Analysis

REFERENCES

Amal Zouaq 1, Roger Nkambou 2 and Claude (2007). Using a Competence Model to Aggregate Learning Knowledge Objects at Seventh IEEE International Advanced Learning Technologies (ICALT 2007)0-7695-2916-X/07.

Barron (2000). Learning Object Pioneers. ASTD LEARNING CIRCUITSRetrieved31July,2005,from:http://www.Learningcircuits.org/2000/mar2000/barron.htm

David Merrill (2006). Knowledge Objects and Mental-Models. The Instructional Use of Learning Objects. Available at http://id2.usu.edu/Papers/KOMM.PDF.

Demetrios G. Sampson (2013). Learning object repositories as knowledge management systems. Knowledge Management & e-Learning, Vol.5, No.2. Jun 2013

Dublin Core Metadata Initiative (2012). Dublin Core Metadata Element Set, Version 1.1. DOI: 2012-06-14. Retrieved online 21-12-2011 from http://dublincore.org/documents/dces/.

Gallenson, Heins (2002). Macromedia MX: Creating Learning Objects [white paper]. Retrieved April 23, 2006.

Horton (2001). Designing Knowledge Objects. William Horton Consulting Inc. e-Learning by design, San Francisco.

IEEE LTSC (2002). IEEE Standard for Learning Object Metadata. 1484.12.1-2002. Available at http://ltsc.ieee.org/wg12/

IMS Global Learning Consortium (2006). IMS Metadata Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata. Version 1.3. Retrieved online 21-12-2011 from http://www.imsglobal.org/Metadata/mdv1p3/imsmd_bestv1p3.html

Jiawei Han, Micheline Kamber **(2008),** "Data Mining Concepts & Techniques" Elsevier.version3 Publisher**:** Morgan Kaufmann Publishers

John Ruffner and Nina Deibler (2008). Knowledge Objects and Learning Objects: Birds of a Feather or Different Species Altogether. The Inter service/Industry Training, Simulation & Education Conference (I/ITSEC).

Kristy de Salas and Leonie Ellis (2006). The Development and Implementation of Learning Objects in a Higher Education Setting, Interdisciplinary Journal of Knowledge and Learning Objects Volume 2, 2006.

Metros (2002-03). Learning Object ontology. NLII Lo. WorkingGroup.Availableathttp://people.cohums.ohiostate.edu/dagefoerde2/NLII_LO/ontology/ontology.htm.

Mortimer (2002). Learning objects of desire: Promise and practicality. Learning Circuits. Retrieved May12, 2006, from http://www.Learningcircuits.org/2002/apr2002/mortimer.html.

Sabitha, Mehotra, Bansal (2012). Quality Metrics quanta for Retrieving Learning Object by Clustering Techniques. The Second International Conference on Digital Information and Communication Technology and its Applications (DCTAP2012). *IEEE.* 978-1-4673-0734-5/12.

Sabitha, Mehotra, Bansal (2013). "Enhanced Learning through Learning Knowledge Object (IJETCAS), ISSN (Print): 2279-0047, ISSN (Online): 2279-0055, Issue 4, Vol. 1, 2 & 3, March-May, 2013

SCORM (2005). Sharable Content Object Reference Model Available at http://www.adlnet.org/scorm/index.cfm.

Shouhong Wang (2008) of University of Massachusetts Dartmouth, MA, USA. Ontology of Learning Objects Repository for Pedagogical Knowledge Sharing in Interdisciplinary Journal of e-Learning and Learning Objects Volume 4.

Vipin, Pang-Ning Tan, Michael (2007) Introduction to Data Mining by Pearson Education India. ISBN: 978-81-3171472-0

Vytautas Štuikys(2007), Towards knowledge-based generative learning objects in ISSN 1392 – 124X information technology and control. Vol.36, No.2

Wagner (2002). Steps to creating a Content Strategy for your organization. ELearning Developers' Journal. ELearning Guild. Available at http://www.elearningguild.com/pdf/2/ 102902MGT-H. pdf

Walid Qassim Qwaider (2011). Integrated of Knowledge Management and E-Learning System. International Journal of Hybrid Information Technology Vol. 4 No. 4, October, 2011

Wayne Hodgins (2002). The Future of Learning Objects .ECI conference on e-Technologies in Engineering Education: Learning Outcomes Providing Future Possibilities, Switzerland. Vol. P01, Article 11.

Wiley (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), The Instructional Use of Learning Objects: Available at http:// reusability.org/read/chapters/ wiley.doc.

Wiley and Recker (2000). A reformulation of learning object granularity. Available at http://reusability.org/ granularity.pdf.

Xavier Ochoa and Erik Duval (2008) Relevance Ranking Metrics for learning objects IEEE transactions on learning technologies, vol. 1, no. 1, January-March.

Xing Dong and Vipin Kumar (2008). Top 10 algorithms in data mining.Knowledge Information System. Springer. DOI 10.1007/s10115-007-0114-2, PP (1–37)