International Journal of Computer Science and Engineering Communications,
Volume.4, Issue.2 (2016): Page. 1343-1350
www.scientistlink.com/ijcsec

Adapting Keyword Recommendation Approach to Folksonomies in Edu-AREA Using Hierarchical Classification

R.Akil Sindhu[1], Dr.R.Manicka Chezian[2]
Research Scholar, Associate Professor,
Department of Computer Science,
Nallamuthu Gounder Mahalingam College, Pollachi, India
akilsindhur@gmail.com[1], chezian_r@yahoo.co.in[2]

Abstract: *Information and Communications Technology in e-learning domain is immensely growing which creates a new path to the development of Web applications supporting the reuse of Open Educational Resources in a collaborative environment. Current web offers various options to keep content and even generate it. The enhancement of Web 2.0 to Semantic Web brings innumerable applications and solutions to society. Edu-AREA is such a kind of Web 2.0 application focusing to provide teaching innovatively. Edu-AREA serves as referatory allowing users to register various resources, containing metadata and refers to them which are available in the external system. At the present stage of Edu-AREA enlargement, problem in managing the organization and classification of information, contributed by users is faced. To resolve this problem a Keyword Recommendation approach to folksonomies is proposed. As Folksonomy is a flat system, to provide a better implementation of folksonomy in Edu-AREA, this paper proposes a non-flat systematic and ontologically semantic structured folksonomy approach that can be processed with an efficient recommendation method and the recommended keywords are generated which are further applied with hierarchical classification technique applying C5.0 algorithm and highly relevant keywords are recommend for tagging the folksonomies in Edu-AREA.*

Keywords: *C5.0, Edu-AREA, Folksonomy, Hierarchical Classification, Ontology, Keyword Recommendation Method.*

I. INTRODUCTION

Web 2.0 is the People-centric and participative Web that facilities reading and writing on the Web making Web transaction bi-directional. It is the Second Generation Web encouraging Participation, Collaboration and Information Sharing. Web 2.0 services in educational domain supports teachers to enhance teaching innovation. This promotes teachers as authors of educational resources and teaching proposals. It mainly aims to make teachers register, create or adapt the existing materials and create their own materials by facilitating them to copy, use, adapt, share, Classify and organize the elements available in boards and annotate them: comments, tags, ratings, etc... This new way of openness and social contribution is applied in Wikipedia, YouTube and resulted in developing Social Bookmarking Systems such as Diigo, de-li-cious etc…This approach of users made them not only to contribute with their data and information, but also with their knowledge to organize the information. This change enhanced the movement of Web 2.0 to Web 3.0 or Semantic Web which is the third generation of Web that can be stated as "executable Web", providing a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

Edu-AREA is a Web 2.0 application that stands for Open Experiences, Resources and Activities that provides a platform to support teaching and create lesson plans innovatively. It is developed to support and facilitates the adoption of open licensing policies. It also focuses to serve as a repository to the lesson planes and guides created by teachers for supporting innovative teaching. Edu-AREA also adopts the idea of iTEC project featuring "Future Classroom". Issues in managing repositories in the e-learning domain gives path to various development processes to promote reusability of lesson plans and OER (Open Educational Resources), pedagogical activities and promotes ICT in e-learning domain enhancing innovative teaching and collaborative learning. This paper presents the functionalities of Edu-AREA, proposed system architecture to support Folksonomies in Edu-AREA and discusses the Keyword Recommendation and Extraction services that can be applied to folksonomies for resolving the problem of managing the classification and organization of information in Edu-AREA.

II. RELATED WORK

O'Reilly Media et al. [1] presents an initiative idea to define and its growth for improving the future generation of software. **Manuel Caeiro Rodriguez** [2] discussed about Edu-AREA a Web 2.0 application, which provides a platform to support teaching innovation and adopt the Open Education Resources. **Frederic Font et al.** [3] presented a general scheme for tag recommendation systems based on the tag co-occurrence in Folksonomies and deals with the issues of Collaborative Tagging such as tag scarcity or ambiguous labelling. **Leandro Balby Marinhoet et al.** [4] presented a method to overcome the uncontrolled vocabulary problem in folksonomies, which automatically enriches a folksonomy with domain expert knowledge. A new Collabulary Filtering Algorithm based on frequent item set mining technique to learn ontology from folksonomies is proposed and proved with better results.
Ivan Cantador et al. [5] presented a mechanism to analyse and classify tags in to set of purpose-oriented categories by analysing the underlying meaning of tags and automatically processing and mapping social tags to semantic concepts that are described in external structured knowledge base. Flickr dataset is used to categorise tags as content-based, context-based, subjective and organisational and shown that content- and context-based tags are superior to subjective and organisational tags by executing RWR recommendation algorithm. **Stefan Siersdorfer et al.** [6] discussed a design framework for Recommender System in Web 2.0 folksonomies. A Vector based representation is introduced in a recommender system that captures the dependencies among Users, items, annotations and social aspects such as contacts, comments, favourites that are represented in form of an IR-like Vector Space model.
K. Schoefeggera et al. [7] presented a Personalized Recommendation System for effective personalization in WIL (Work-Integrated Learning) that includes unobtrusive user modelling approach for representing the user's context in terms of their topics of interest, skills and knowledge levels based on their work process and organizational memory for learning. **J. Trant** [8] discussed a framework for the study of folksonomy, tagging and social tagging systems and various models and methods followed for constructing those three broad approaches. **Enrique Estelles and Esther del Moral** [9] analysed and compared different Social Bookmarking Systems and specifically discussed Diigo's features and benefits for research and education. **Christian GLAHN et al.** [10] analysed the open standards for supporting the reusing OER in different knowledge domains. They discussed about the problem in reusing the available resources and presented the OER projects, MACE, OpenScout and Share.TEC. The principal structure of these projects is similar, but, its focus on OER varies.
Andreas Hotho et al. [11] evaluated the Adapted PageRank on the del.ico.us dataset and proposed a FolkRank algorithm and proved with better recommendation experimenting in a large scale dataset. **Zarli Htun and Phyu Phyu Tar** [12] proposed a Resource Recommender System tested by using the de-li-cious and LastFM by exracting Latent topics from tagging data by LDA and proved showing better results than Collaborative Filtering.

III. FUNCTIONALITIES OF EDU-AREA

Edu-AREA provides a collaborative platform for Registering Resources that can act as a referatory to applications, devices, websites present in the external system and enables creation of activities, lesson plans, guides using the registered resources [2]. The facility of Edu-AREA provides browsing the existing resources, registering several information records for a particular object and creation of activities providing pedagogical information and Learning goals and documenting the experiences. Table.I provides the functionalities and features of Edu-AREA in detail.

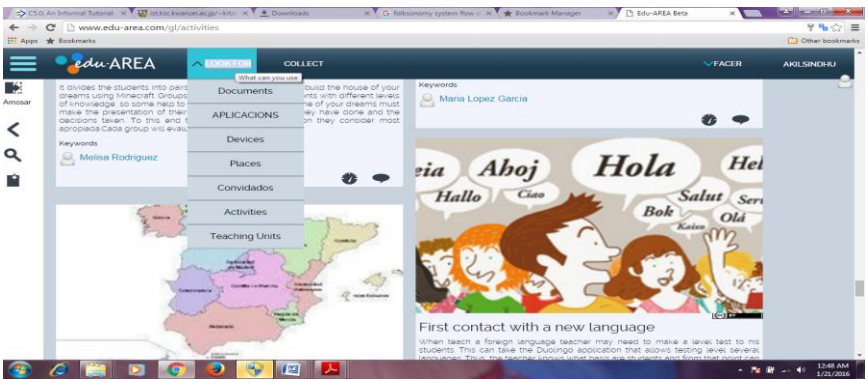

Fig.1 Edu-AREA Home Page

IV. PROPOSED SYSTEM MODEL

Edu-AREA as a typical Web system allows users to access the application through a Web browser. The browser sends HTTP queries to a Web Server (Apache), and it is connected to a relational database (MySQL) for information storage and retrieval. Fig.2 presents the architecture model of the system with its features.

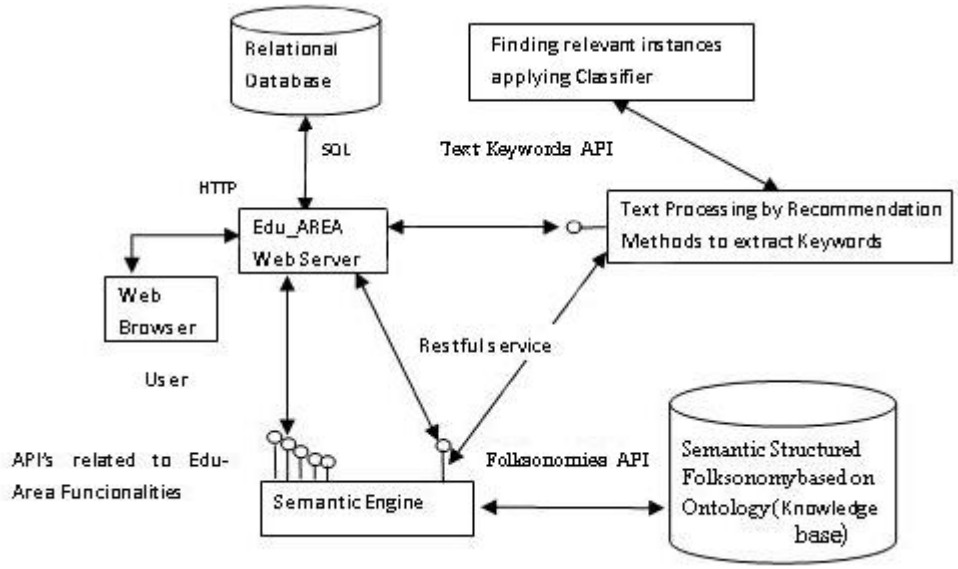

Fig.2. Proposed System Architecture

1) Folksonomy: A folksonomy is the process of tagging data and creating conceptual structures by the user. It is a classification system where the classifying elements are defined by the users. Its structure is:

- A tuple ***F: = (U, T, R, Y)*** where U denotes the finite set of users, T denotes a finite set of Tags; R denotes a finite set of Resources.

- **Y ⊆ U × T × R** is a ternary relation between users, tags, and resources. Representing the fact that user U has assigned tag T to resource R [4].

2) RDF Triplets: It is a simple modelling language providing a standard model for data interchange on the Web. It uses URI to identify the Web resources and forms a graph model describing the relation between the resources. The RDF data model's approach is in the form of Subject-Predicate-Object relation, which is known as "triples" in RDF terminology.

3) Web Ontology Language (OWL): The key factor of Semantic Web is the Ontology language, which focuses on the study of existence, inference or behaviour. The Structure of ontology [13] is presented in Fig.3.

5-tuple O: = (C, HC, R, HR, I).

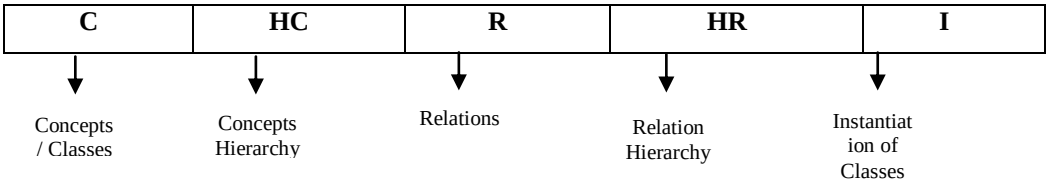

Fig.3. The Structure of Ontology

In Edu-AREA, the objects and information records will be maintained in a Knowledge Base, which is processed by a Semantic engine to manage and relate the several folksonomies and their items. The Semantic Engine includes Application Interfaces (APIs) of the proposed system and uses RESTful approach. A particular API will be focused for identifying similarities among information records, as Edu-AREA allows registering different records for the same object. The information provided by the Web Server using API methods are stored in the Knowledge Base as RDF triples in accordance to a pre-defined Ontology. The folksonomies in Edu-AREA are managed based on Ontologies, along with a set of logical rules defined with Semantic Web Rule Language (SWRL), adapting the principle "whenever the conditions specified in the antecedent holds, then the conditions specified in the consequent should also hold" builds the knowledge of the system. This framework helps to perform the inference process in the available information records to filter records effectively. In Edu-AREA to overcome the problem of managing and organizing information records, a Keyword Recommendation experiment is applied to folksonomies, in a hierarchical classification approach.

V. HIERARCHICAL CLASSIFICATION APPROACH

A. Hierarchical Classification for Proposed System

Classification is the process of determining labels to undiscovered instances when it is newly evolving. It is a supervised learning approach. Researches in the fields such as Data Mining, Machine Learning, Statistical Pattern Recognition and related areas have focused on flat classification problems. The huge variety of the classifiers deals with a flat class structure where a single class is assigned to an example data and hence no hierarchical relationship between classes is found.

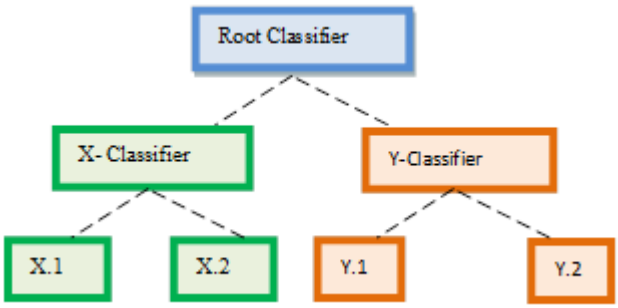

Fig.4. Hierarchy reflection in a tree of classifiers

The classes in hierarchical classification are arranged in a hierarchical structure. The most general level of instance is assigned to the root of the tree and specialized levels are assigned at the branches of the tree and the tree grows on more specific exploration of instances. When two classes X and Y are considered which are different from each other, Hierarchical classification explores the elements belonging to each class (i.e.) X.1 is belonging to X and Y.1 belonging to Y based on its similarities with X and Y, so it denotes, X.1 and X.2 are more alike and belonging to X class and similarly evaluating the Y class. Fig.4 illustrates the representation of hierarchy reflection. In hierarchical classification approach the classes to be analyzed are organized into a class hierarchy which typically reflects a tree or a DAG, which is a Directed Acyclic Graph containing no cycles. A rooted tree is a special kind of DAG and a DAG is a special kind of directed graph. The advantage of constructing hierarchical framework of folksonomies allows Abstract elements also to register in the system as it can be associated with its concrete classes and defined, analyzing its hierarchy and a semantic structured folksonomy framework will be generated resulting in a (sub) graph and each instance of the (sub) graph is assigned a relevance value which is obtained through our classifier algorithm to recommend highly relevant keywords to organize the information records.

B. C 5.0 Algorithm for Classification

Decision trees remains as an effective method in supervised learning. Its input is a set of classified data, and produces output in a tree structure. Its each leaf node is a decision (a class) to which the instance belongs to after analyzing all test path from root where, each non-leaf internal node denotes a test. Better results are produced by a balanced tree. ID3 (Iterative DiChaudomiser 3) algorithm was developed by Ross Quinlan. This Concept Learning Algorithm builds a decision tree from a fixed set of elements and resulting tree classifies future samples. C4.5 is a statistical classifier

generating decision tree for classification, handling categorical and continuous attributes is an extension of Quinlan's ID3 algorithm. Quinlan continued creating an improved C4.5 algorithm called C5.0 to prove improvement in major features of performances which is presented in Table II. C5.0 constructs a classifier to predict case's class from the values of the other attributes and express as decision trees. Every case's belongs to one of a divided number of mutually exclusive classes (e.g.: Positive, Negative). Properties of those cases which are relevant to its class are provided where some cases may hold unknown values of some attributes. C5.0 classifier can produce effective results for our proposed system as it is highly performable compared to ID3 and C4.5.

These classification algorithms use the concept of information entropy. It is the average value of the information contained in each message. Its general form is:

$$Entropy\ (X) = \sum_{i=1}^{k} -p_i \log_2 p_i$$

Where, if X is the training dataset, containing the set of classified samples (x_1, $x_2 \ldots x_2$) and each samples contains the p-dimensional vector (p_1, $p_2 \ldots p_n$), then p_n denotes the attributes of features of the sample.

$$Gain\ (X, A) = Entropy\ (X) - \sum_{v \in values\ (A)} \frac{|X_v|}{|X|} Entropy\ (X_v)$$

Information Gain is the expected reduction in entropy resulted by splitting the samples in to subsets according to a given attribute. Highly normalized information gain (difference or change in entropy) is chosen to make the decision. Where, values (A) are the possible set of values of attribute A and X_V is the subset of X for which attribute A has value V. The information records of Edu-AREA can be experimented implementing C5.0 algorithm and improved to retrieve and organize the relevant information records based on the relevancy values generated with our algorithm.

TABLE.II ADVANTAGES OF C5.0 OVER ID3 and C4.5

Factors	Improvements Compared	Performance Compared
Speed	More faster	Several orders of Magnitude
Memory	More memory efficient	Highly Scalable, adopted for computers with multiple CPU's.
Smaller decision trees	Considerably small decision trees resulted	Better Pruning of trees.
Boosting	Improved trees	Better Accuracy
Weighting	Weight Different cases	Handles Misclassified cases and different weighted attributes
Winnowing	Winnows the unwanted and unhelpful attributes	Better Winnowing of attributes
Missing data	Unknown values are handled	Missing data can be handled more in number

VI. KEYWORD RECOMMENDATION SERVICES AND METHODS

Recommendation systems make effective filtering of information and helps user for better decision making. .Natural Language Processing and Machine Learning along with Information Extraction technique achieves coherent text processing of information adapting a suitable recommendation method to support extracting set of keywords from the information record. Fig.5 illustrates the various recommendation methods.

A. Proposed Keyword Extraction Service

Keyword Extraction Services extracts effective keywords that help in relevant retrieval of information. The proposed Keyword Extraction Service derives the (sub) graph of instances based on the semantic structure of the ontology for text

annotation of information records. The relevant value assigned to the instances through our algorithm is used to sort and predict highly relevant instances and they are considered to recommend as keywords to annotate the topics of Edu-AREA. The curation information contributed by the users and the description parts can be considered to implement our proposed system for recommending keywords to tag folksonomies.

B. Proposed Recommendation Methods

Content Based Filtering method can be used to filter Edu-AREA element's description part and analysis on users comments based on likes and dislikes can be collected from boards processed based on the User Profile, which helps to derive folksonomy structure for classification, organization and retrieval of information records in Edu-AREA.

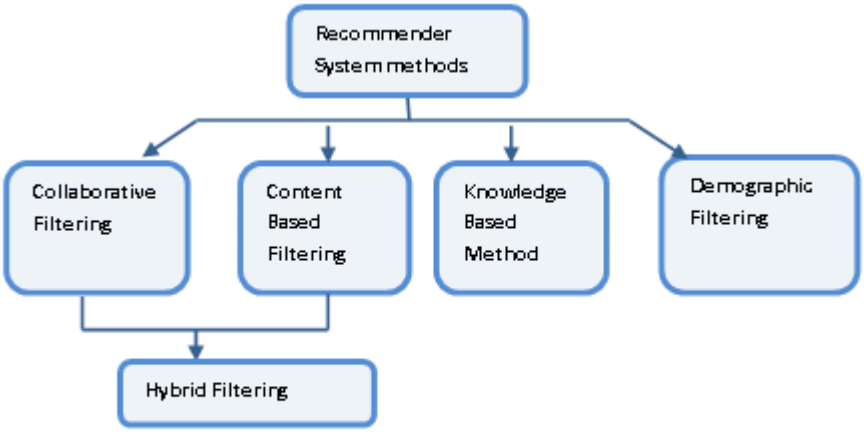

Fig.5. Basic Recommendation Methods

Copious candidate elements are compared with other elements previously rated by the user and the best-matching elements can be recommended based on its similarities and history of user likes and nature of comments can be also analyzed to recommend elements that are similar to the previously liked. So Content Based Filtering can be combined with other recommendation methods and a Hybrid Recommendation method can be generated which is applied on the ontologically preprocessed folksonomy and relevant keywords are extracted and recommended to tag the folksonomies in Edu-AREA. The root of this recommendation method is from the research of information retrieval and information filtering.

VII. CONCLUSION

The need for movement of Web from Web 2.0 to Web 3.0 and its impact on e-learning domain is rapidly developing. It causes teachers to evince their contribution effectively in reusing open educational resources. Edu-AREA stands for Open Experiences, Resources and Activities creating a platform to support teaching and creating lesson plans innovatively. As numerous users are being part of Edu-AREA and contributing information records to the system for openness and social learning, the problem in managing the classification and organization of information records is faced. To resolve this issue, this paper proposes to implement a keyword recommendation approach to folksonomies in Edu-AREA based on semantic technologies, an efficient Recommendation method and hierarchical classification technique which can effectively solve the problem of managing the classification and organization of data in Edu-AREA.

REFERENCES

[1] Tim O'Reilly," What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software.",Communications & Strategies, No. 1, pp.17-35, First Quarter 2007. Available at SSRN.

[2] Manuel Caeiro-Rodriguez, "Opening Lesson Plans to Support Teaching Innovation and Open Educational Resources Adoption", Proceeding OpenSym '14 Proceedings of The Symposium on Open Collaboration Internationa,l ACM New York, NY, USA , pp.47, 2014.

[3] Frederic Font, Joan Serra, "Folksonomy-based tag recommendation for collaborative tagging systems", International Journal on Semantic Web and Information Systems, vol.9, Issue 2, pp.1-30, April 2013.

[4] Leandro Balby Marinho, Krisztian Buza and Lars Schmidt-Thieme, "Folksonomy-based Collabulary Learning ",Proceeding ISWC '08 Proceedings of the 7th International Conference on The Semantic Web, Springer-Verlag Berlin, pp.261-276, October 2008.

[5] Ivan Cantador, Ioannis Konstas, Joemon M. Jose, "Categorising Social Tags to Improve Folksonomy-based Recommendations", Web Semantics: Science, Services and Agents on the World Wide Web, vol. 9, Issue 1, pp.1-15, March 2011,

[6] Stefan Siersdorfer, Sergej Sizov , "Social Recommender Systems for Web 2.0 Folksonomies", HT '09 Proceedings of the 20th ACM conference on Hypertext and hypermedia, ACM New York, NY, USA , pp.261-270, June 2009.

[7] K. Schoefegger, P. Seitlinger, T. Ley, "Towards a User Model for Personalized Recommendations in Work-Integrated Learning: A Report on an Experimental Study with a Collaborative Tagging System", 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), pp.1-10,2010.

[8] J.Trant, "Studying Social Tagging and Folksonomy:A Review and Framework", University of Toronto / Archives & Museum Informatics, Journal of Digital Information, vol.10, No 1, 2009.

[9] Enrique Estelles, Esther del Moral, "Social Bookmarking Tools as Facilitators of Learning and Research Collaborative Processes: The Diigo Case", Interdisciplinary Journal of E-Learning and Learning Objects,vol.6, Issue 1, pp.175-193 ,2006.

[10] Christian GLAHN, Marco KALZ, Marion GRUBER & Marcus SPECHT, "Supporting the Reuse of Open Educational Resources through Open Standards", pp.308-315, 2010.

[11] Andreas Hotho, Robert Jaschke, Christoph Schmitz, Gerd Stumme, "FolkRank: A ranking algorithm for folksonomies", UNIVERSITY OF HILDESHEIM, INSTITUTE OF COMPUTER SCIENCE, vol.1, pp.111-114, 2006.

[12] Zarli Htun, Phyu Phyu Tar, "A RESOURCE RECOMMENDER SYSTEM BASED ON SOCIAL TAGGING DATA", Machine Learning and Applications: An International Journal (MLAIJ) vol.1, No.1, pp.1-11, September 2014.

[13] Mohammad Mustafa Taye," Understanding Semantic Web and Ontologies: Theory and Applications" Journal of Computing, vol.2, Issue 6, pp.182-192, June 2010.

[14] Anurag Upadhayay, Suneet Shukla,Sudsanshu Kumar , "Empirical Comparison by data mining Classification algorithms (C 4.5 & C 5.0) for thyroid cancer data set", International Journal of Computer Science & Communication Networks,vol.3, Issue 1,pp.64-68,2013.

BIOGRAPHY

R.Akil Sindhu received her M.Sc., in Software Engineering from MCET, Pollachi under Anna University, Chennai in 2012. She worked as a Software Engineer from 2013 to 2015 in HCL Technologies Ltd, Chennai. Currently she is a Research Scholar in Department of Computer Science, NGM College, Pollachi under Bharathiar University, Coimbatore. She has presented her Research Papers in National and International Conferences. Her Areas of interest includes Data Mining, Semantic Web Mining, Ontology Engineering and Big Data Analytics.

Dr. R. Manicka Chezian received his M.Sc., in Applied Science from PSG College of Technology, Coimbatore, India in 1987. He completed his M.S. degree in Software Systems from Birla Institute of Technology and Science, Pilani, Rajasthan, India and Ph.D degree in Computer Science from School of Computer Science and Engineering, Bharathiar University, Coimbatore, India. He served as a Faculty of Maths and Computer Applications at P.S.G College of Technology, Coimbatore from 1987 to 1989. Presently, he has been working as an Associate Professor of Computer Science in NGM College (Autonomous), Pollachi, India since 1989. He has published more than hundred papers in various International Journals and Conferences. He is a recipient of many awards such as Desha Mithra Award, Best paper Award, Best Research Supervisor Award, Life Time Achievement Award in Computer Science Field, and Best Computer Science Faculty of 2015. He is a member of various professional bodies like Computer Society of India and Indian Science Congress Association. His research focuses on Network Databases, Data Mining, Data Compression, Mobile Computing and Real Time Systems, Network Security, Bio-Informatics and Distributed Computing.

TABLE.I Edu-AREA Functionalities

Functions	Function Description	Features
Registering Resources	Resources can be registered by different users in Edu-AREA introducing the data: name, picture, description and URI. Optional: keywords, relations with other resources. To register different information record of the same object, the system will check and if found will provide three options:	(i) To discard the suggested object and register the record as new object. (ii) To accept the suggested object and stop the registering the information record. (iii) To accept the suggested object but also to register the information as a new record for the same object.
Creating and Copying Activities	Creation of Activities and guides as information records including different sections following the guidelines, and Copying and adapting an existing element to make a personal copy for modifications. The original element cannot be modified, and can be edited only by the owner.	Information can be provided such as: (i) Name, picture, description and a URI to the original source described outside of Edu-AREA. (ii) Pedagogical information about opportunities and goals of lesson plans. (iii) Resources that can be involved and guidelines to support the activity. AREA is a system that presents information on proposed teaching references and support teachers to develop their own lesson plans, which is called AREA guides.
Browsing Resources	Browsing of information in Edu-AREA are shown in listings as boxes that includes a name, a picture, a short description and some keywords about the element, information of author, annotations (comments, social actions).	Searching records in Edu-AREA facilitates the user to (i) Introduce a text to search and find the elements containing such text. (ii) To select a certain user and browse all his/her elements. (iii) To denote a certain item of a folksonomy to get relevant elements.
Documenting Experiences	Documenting experiences where the educational experiences of teachers can be made as record of evidences and documented.	This involves the registration of positive and negative comments, attachment of evidences, such as outcomes, images and videos to the teaching guides.
Curation	Classifying, organizing annotating elements via boards.	Facilitates users in commenting, tagging, rating, classifying, and organizing. This participation made users as curators of information.