ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SEMANTIC APPROACH TO ANNOTATION OF LEARNING OBJECTS

[1]E.A.VIMAL, [2]Dr S..CHANDRAMATHI

[1]Assistant Professor, Department of Information Technology,
Kumaraguru College of Technology,Tamil Nadu, India
[2]Dean, Department of Electrical sciences in Hindustan College of
Engineering & Technology, Coimbatore-TamilNadu
E-mail: [1]eavimal0781@gmail.com , [2]chandrasrajan@gmail.com

ABSTRACT

The primary motto of this paper is to design and develop the semantic annotation model for e-learning document and to find the presence of the concepts in the document. The awareness of the semantic web needs the widespread availability of semantic annotations for the obtainable and new documents on the web. Semantic annotations are to label ontology class instance data and plot it into ontology classes. Here, we are first applying the stop word removal technique and considering other contents of the documents to create the concept matrix and also we are considering the index terms to create a separate concept matrix. The concept matrix which are developed using the contents and the index terms are then combined to find the level of the presence of specific concepts in a particular document. This technique is implemented in Java and we have identified the percentage level of the presence of the concepts in the sample documents.

Keywords: *Semantic Annotation, Learning Objects, Concept Matrix, Ontology.*

1. INTRODUCTION

A deliberate use of networked information and communication technology in teaching and learning is generally denotes as e-learning. This type of teaching and learning is also explained by some other terms like online learning, virtual learning, distributed learning, network and web based learning. Basically, they all denote the educational process which uses the information and communication technique to mediate the synchronous and asynchronous learning and teaching activities. On closer examination, it is clear that these labels denote a slight difference in educational process and those labels cannot be used synonymously with the term e-learning.

The term e-learning contains more features than online learning, virtual learning, distributed learning, networked or web based learning. The letter "e" in the e-learning denotes "electronic" which would include all the educational activities. The different methods of e-learning activities are as follows: the *individualized self paced e-learning online* denotes a circumstance that an individual learner is accessing the learning resource like database through internet. A good example for this technique is conducting research on the internet. The *individualized self paced e-learning offline* denotes a circumstance which the individual learner is using the learning resource such as CD or DVD i.e. not connected to the internet. The *group based e-learning synchronously* denotes the circumstance that a group of learners working together through the internet. A good example for this technique is real time chat. The *group based e-learning asynchronously* denotes that a group of learners working over the internet which is not in real time. An example for this technique is on-line discussion through e-mail.

The ontology based modelling was adopted in the e-learning field not only for the learning resources but also for the user profile. In [9] they have provided a solution where the user profile is focused on the user competencies which are indicated in terms of same ontology adopted in LOs annotation. The ontology based annotation evenness of different sort of e-learning resources could be used in many ways to provide the users with personalized functionalities like: for choosing the suitable users for being co-opted into a certain project whose topic are also explained in ontology concept; for selecting the materials appropriate for certain users in certain condition when accessing a particular course site section or when solving a certain course homework as a student or when developing a certain course material as a teacher;

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for redirecting the student to all the materials which provides explanations, in case he gives a wrong answer; for providing the student with the suitable materials which are endorsed or published by his collaborators in different projects or interest groups. These facilities can be exported from an e-learning technique to another if they are implemented in the form of web services. Therefore, the user mobility across the distributed e-learning communities is facilitated by the user competencies profile which is recognized by other systems as well as by the e-learning materials which are evenly modeled.

The whole implementation of semantic web demands widespread availability of semantic annotations for accessible and new documents on the web. Manual annotation is easily skilled today using the authoring tools like semantic word [10] which gives an integrated environment for concurrently authoring and annotating the text. The use of the human annotators is habitually fraught with errors because of the factors like annotator familiarity with the domain, amount of training, personal motivation and complex schemas [11]. The problem in the manual annotation is that the volume of the accessible documents on the web must be annotated to make it valuable for the semantic web.

The semantic annotation platforms (SAPs) can be divided based on the sort of annotation technique which we have used. There are two primary divisions, they are pattern based and machine learning based. Moreover, some platforms would use the techniques from both the classes which are called multi-strategy in order to take the benefit of strengths and to compensate the weakness of the techniques in each category. The pattern based SAPs can execute the pattern discovery or have patterns manually explained. Most pattern discovery techniques follow the basic method outlined by Brin [12]. The preliminary set of entities is explained and the corpus is scanned to discover the patterns in which the entities exist. Fresh entities were identified along with the new patterns. This process continues repeatedly until no more entities are identified or the user stops the process. The annotations can also be developed using the manual system to find the entities in the text.

In this paper, we have selected some documents to find the presence of concepts in those documents. The master concepts are the concepts which we are checking in the documents for the level of presence in those documents. The master concepts are then applied in the word net to find the derived concepts. The derived concepts and the master concepts are then used to find the concept matrix based on contents and the concept matrix based on index terms. After finding the concept matrixes, we find the importance measure and apply the matrix reduction technique. The final resultant matrix is found by combining the concept matrix based on contents and the concept matrix based on index terms after applying the matrix reduction technique.

This paper is structured as follows: The second section of this paper shows the descriptions of some related works and the third section explains our proposed technique and the fourth section details the result of our proposed technique and the fifth section concludes our technique.

2. RELATED WORKS

A number of researches have been performed based on the semantic annotation. Some of the contemporary works regarding to the semantic annotation is as follows:

Hong Cui [2] has proposed a technique named charaParser for fine-grained semantic gloss of organism morphological descriptions. His article defines the improvement and estimation of charaParser, a software application for semantic gloss for morphological descriptions. CharaParser annotates the semistructured morphological depiction in a detailed manner that all the stated morphological characters of an organ are marked up in extensible markup language format. Using an unconfirmed machine learning algorithm and a general purpose syntactic parser as its key annotation tools, CharaParser needs minimal additional knowledge engineering work and seems to perform well across diverse description collections and taxon groups. The system has been officially evaluated on over 1,000 sentences arbitrarily selected from Volume 19 of *Flora of North American* and Part H of *Treatise on Invertebrate Paleontology*. CharaParser achieves and exceeds 90% in sentence-wise recall and accuracy, exceeding other comparable methods reported in the literature. It also considerably outperforms a heuristic rule-based system we developed earlier. Early evidence that enriching the lexicon of a syntactic parser with domain terms alone may be adequate to adapt the parser for the biodiversity domain is also observed and may have significant implications.

Tong Zhen Zhang *et al.* [3] has proposed a Learning Objects routine Semantic Annotation by Learner significance Feedback. In this technique,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

they have introduced an associated semantic network as the semantic depiction model; use semantic keywords, a linguistic ontology in semantic resemblance calculation and use learner significance feedback to complete automatic semantic annotation. After a number of iterations of learner significance feedback, semantic network is enriched mechanically. In addition, semantic seeds and semantic loners are employed predominantly to rapid up the growth of semantic network and to get a balance annotation.

Rim Faiz *et al.* [4] proposed a relevant knowledge objects extraction based on semantic annotation of documents. This technique is about to automatically gloss the texts with semantic metadata: learning sort of textual segments. These metadata would permit us to explore and extract knowledge information from texts indexed in that way. This model is build up from two units: the first unit consists on a semantic annotation of learning objects according to their semantic classes (*definition, example, exercise, etc*.). The second unit exploits automatic semantic annotation which is generated by the first part to create a semantic inverted index which is able to find relevant learning objects for queries associated with semantic classes. To class the results based on their significance, we apply the Rocchio's categorization system on the learning objects. They have applied a system called SRIDoP, on the basis of the proposed model and they have demonstrated its effectiveness.

Jain.S *et al.* [5] proposed a keyphrase extraction tool for semantic metadata footnote of learning materials. Keyphrases play a significant role in depicting a document. In learning management systems they lead to enhanced information retrieval. On the other hand, comparatively few learning documents have key terms assigned and therefore finding techniques to computerize the extraction is desirable. The goal of their proposed technique is to depict the generation of list of key phrases from a document using part of speech tagging and ranking them by means of formatting features implemented on them by the author rather than trusting only on statistics (such as term frequency).

Jain.S *et al.* [6] proposed automatic topic recognition from learning material. The capability to judge the significance of topics and associated sources in information-rich environments is a key to victory when scanning online learning environments. A Learner may be searching for learning materials defining given topic or exercises on the topic. Any learning material may wrap various topics related to multiple subject domains. Jain.S *et al.* proposed an ontological approach for identifying major topics, covered in the learning material. Along with the topics, the subject and discipline to which those topics belongs to and relevance of the topic in the learning material as contrasted to other topics present in the same document are also exposed. Domain ontology is created to retrieve the topics enclosed in the document. They have presented estimation against a manually classified topics as well as author's judgment of relevance of the topics discovered by our system. Evaluation outcome showed that the technique presented by them is effective in recognizing topics and subtopics enclosed in a single learning document.

Mark J. Weal *et al.* [7] proposed a semantic annotation of omnipresent learning environments. Skills-based learning environments are utilized to endorse the attainment of practical skills, decision making, communication, and problem solving. It is significant to provide feedback to the students from these sessions and remarks of their actions may inform the estimation process and help researchers to better comprehend the learning process. After a series of prototype demonstrators, they have explored the use of semantic gloss in the recording and consequent understanding of such simulation environments. Their Semantic Web approach is outlined and conclusions drawn as to the appropriateness of diverse annotation techniques and their combination with ubiquitous calculating techniques to provide new mechanisms for both student feedback and increased understanding of the learning environment.

N.Dovrolis *et al.* [8] proposed a semantic footnote and linking of medical educational resources. Educational content can be reused in diverse contexts because it is habitually shared among different educators and is enriched, adapted and in general repurposed. They suggested the metamorphosis background for publishing, sharing and repurposing instructive content in medical education. The motto is to enable more appropriate searching and retrieval of medical educational resources and linking other related resources in the medical domain together with scientific publications and clinical data.

Mihaela M. Brut *et al.* [1] proposed a semantic oriented approach for organizing and developing gloss for e-learning. They proposed a solution to widen the IEEE LOM standard with ontology based semantic annotation for well-organized use of learning management systems. The technique

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which they proposed extends and combines two consecrated alternative methods for structure based indexing of textual resources.

3. CONTRIBUTIONS OF THE PAPER

The important techniques which we have included in this paper are as follows:

- We have considered the index terms of the documents to identify the concepts present in that document with more accuracy.
- We have calculated the concept matrix based on contents and the index terms of the documents by applying the contents and index terms in the word net and we have used the techniques like matrix reduction and final resultant matrix to combine both the concept matrix.
- We have calculated the importance measure for both the concept matrixes based on the contents of the document and the index terms to identify the concepts in a document with good precision.
- We have added certain weight values for the concept matrix based on contents and the concept matrix based on index terms in the final resultant matrix technique to improve the accuracy.

4. PROPOSED TECHNIQUE TO SEMANTIC ANNOTATION OF LEARNING OBJECTS

The complete process which we have proposed in this paper is explained in Fig.1 that is shown as block diagram.

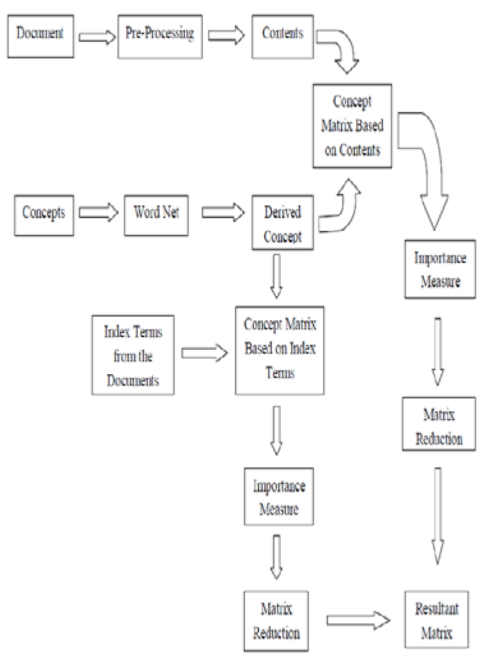

Fig.1 Block Diagram of our Proposed Technique

4.1 Preprocessing

The preprocessing is a technique which is used to make the dataset for further processing. In its original form it would be difficult to process the document. So we have to convert the dataset which is understandable to the algorithm. The methods which are used to format the documents are stop word removal and stemming. The details of the methods are as follows:

The stop word removal is the process of removing the commonly used words that has less significant meaning than the keywords. Generally the search engines remove the commonly used words or the stop words from the key word phrase to give the most pertinent result. While searching, the entire stop words, for example ‘a’ and ‘the’ will be detached from multiple word queries to increase the search performance. The common stop words are ‘it’, ‘can’, ‘an’, ‘and’, ‘by’, ‘for’, ‘from’, ‘of’, ‘the’, ‘to’, ‘with’. The stop word removal is done while parsing a document to gain the information about the content or while scoring new URLs that the page recommends.

4.2. Concept extraction using mutual information

The concept is a keyword which has some relevance with the domain and has some specific characteristics.

$$B = A_1, A_2, \ldots\ldots, A_n$$

Where,

B → Domain

A_n → Concept belongs to the domain

The main objective of this step is to find the relation between the concept to concept (word to word) and concept to domain. We are using a sentence level windowing technique which the window moves in a sliding manner. The text window formed is in the form of four term window which was enclosed in a sentence. At first, we have to find the highest frequent word and then the technique finds the dependency of this word to other and other words to this.

$$freq(x) = \frac{X_n}{N_A}, X \in A$$

Where,

X_n → Number of x present in the domain

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

x → Element subjected for frequency finding

N_A → Total number of elements in the domain

The above equation is used to find the most frequent elements. After finding the most frequent keyword, we have to explore whether it belongs to a concept in the concept map. The selection of keyword to concept is done by finding the inter relationship amid the keyword and other keywords. The bond amid the two keywords was obtained by finding the probability of occurrence of the keyword. We have used a conditional probability to discover the relation between the keywords. The mutual information is a technique which is used to extract the concept. If the keyword shows higher dependency amid others, then it is considered as concept. Examination of this technique shows that more dependency would extract the concept more from the text corpora. The calculation of the mutual information is as follows:

$$MI(x:y) = \frac{P\left(\frac{x}{y}\right)}{P(y)}, x, y \in D$$

$$P\left(\frac{x}{y}\right) = \frac{P(y \cap x)}{P(x)}$$

Where,

MI → Mutual information

x, y → Terms from the document

D → Document

The function $MI(x:y)$ is used to find the mutual information amid the terms and thus extracting the concepts which is needed for concept extraction. The function $P(\bullet)$ is the probability of each word from the document.

4.3. Generating concept matrix

The concept matrix is generated as follows; after the stop word removal, we have to create the concept from the documents. The concepts are created by taking two words together from each document. Thereafter we have to check the master concept and the derived master concept with the concepts (i.e. contents) of each document. The master concepts are the concepts which we are giving as keywords. The derived master concepts are the concepts which we have derived from the master concept using the word net. For instance, we are giving the master concept as "Error Analysis". Thereafter we have to apply this master concept words separately in word net i.e. "Error" a single word and "Analysis" a single word. The word net will then give the meanings of both words separately. The word net gave the meaning of error as mistake, error and fault and the word net gave the meaning of analysis as analysis. After we got the synonyms of both the words from the word net, we have to combine the synonym of one word with the synonym of another. This is called derived master concept. The derived master concepts which we got here are mistake analysis, error analysis, fault analysis and error analysis. In this derived master concepts, the error analysis is repeated because we have to include the master concept with the derived concepts.

4.4. Concept matrix based on contents

After discovering the derived concept from the master concepts, we have to create the concept matrix separately for all the master concepts which we are giving based on the contents in the documents. A sample concept matrix for the first master concept based on the contents in the documents is shown in the Fig.2. In this figure C_1 is the first master concept, $CB_1, CB_2, \ldots\ldots, CB_8$ are the derived master concepts and $D_1, D_2, \ldots\ldots, D_{10}$ are the documents which we have taken.

The method for creating the concept matrix based on the contents in the documents is as follows:

$$C_m D_n C = \frac{R}{TC}$$

Where,

$m = 1,2,3,\ldots..$ → Number of Concepts

$n = 1,2,3,\ldots..$ → Number of Documents

R → Repeated Concept

C → Concept

D → Document

TC → Total number of contents in a particular document

This formula explains that, the concept matrix is the ratio of derived concept to the total number of concepts in the document. An example of concept matrix for eight derived concepts with the master concept and ten documents is shown in the Fig.2.

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$C_m D_n C =$

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
CB_8	0.62	0.84	0.90	0.28	0.70	0.75	0.48	0.88	0.36	0.34
CB_7	0.51	0.90	0.97	0.22	0.51	0.59	0.51	0.78	0.18	0.26
CB_6	0.42	0.95	0.96	0.21	0.41	0.50	0.55	0.72	0.07	0.25
CB_5	0.38	0.94	0.98	0.20	0.42	0.48	0.53	0.79	0.04	0.22
CB_4	0.68	0.01	0.21	0.88	0.72	0.82	0.70	0.85	0.40	0.92
CB_3	0.52	0.81	0.90	0.45	0.52	0.64	0.67	0.88	0.12	0.45
CB_2	0.92	0.08	0.15	0.41	0.93	0.90	0.16	0.68	0.86	0.48
CB_1	0.75	0.72	0.91	0.33	0.74	0.84	0.51	0.90	0.46	0.38
C_1	0.38	0.94	0.98	0.20	0.42	0.48	0.53	0.79	0.04	0.22

Fig. 2 Sample Concept Matrix Based On The Contents

The above concept matrix explains that, the concept matrix value of the eighth derived concept of the first master concept in the first document is 0.62 and the concept matrix value of the eighth derived concept of the first master concept in the second document is 0.84 and so on. Similarly, the concept matrix contains the values of each derived concepts with its respective documents. Likewise, we have to calculate the concept matrix for all the master concepts which we have taken.a

4.5. Concept matrix based on index terms

Thereafter, we have to calculate the concept matrix with respect to the index terms $I_1, I_2, ..., I_n$. The index terms are the terms which are given as keywords in each document. We need to check these index terms $I_1, I_2, ..., I_n$ with each derived concepts $CB_1, CB_2, ..., CB_n$ of every master concepts. For instance, if the concept is "Error Analysis", we have to check the index terms of every document with respect to "Error" and "Analysis" separately. The formula for calculating the concept matrix based on index terms is as follows:

$$C_m D_n I = \frac{RW}{NWC}$$

Where,

RW → Related words in the index terms

NWC → Number of words in the concept

This formula explains that, it is the ratio of the related words in the index terms to the number of words in the concept which we are checking. Here, we have the concept as "Error Analysis". It has two words, so the number of words in the concept is two.

A sample concept matrix with respect to the index term is shown in the Fig.3. In this sample concept matrix, we have taken eight derived concept $CB_1, CB_2, ..., CB_8$ with the master concept and ten documents $D_1, D_2, ..., D_{10}$.

$C_m D_n I =$

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
CB_8	1	1	0	0.5	1	0.5	0	0	1	0.5
CB_7	0	1	0	0	0.5	1	0.5	0	0.5	0.5
CB_6	1	0	0.5	1	0.5	0	1	0.5	0.5	1
CB_5	0.5	1	0.5	0	1	0.5	0	1	0.5	0
CB_4	0.5	0	1	0.5	0	1	0.5	0	1	0.5
CB_3	0	0.5	0	0.5	1	0	1	0.5	0	1
CB_2	1	1	0.5	0	0.5	1	0.5	1	0.5	0
CB_1	0.5	0	1	1	0	0.5	0	0.5	1	0.5
C_1	1	0.5	0	1	0.5	1	0.5	0	0.5	1

Fig.3 Sample Concept Matrix With Respect To Index Terms

The above sample figure explains that, the concept matrix value for the eighth derived concept in the first document related to the index terms is 1 and the concept matrix value for the eighth derived concept in the second document related to the index terms is 1 and so on. Similarly, the concept matrix contains the values of each derived concepts with its respective documents related to the index terms. In the same way, we have to calculate the concept matrix for all the master concepts which we have taken with related to the index terms of the documents.

4.6. Importance Measure

After finding the concept matrix for the documents related to the contents $A_1, A_2, ..., A_n$ and index terms $I_1, I_2, ..., I_n$, we have to find the importance measure for the entire derived concept $CB_1, CB_2, ..., CB_n$ and the master concept with its respective documents $D_1, D_2, ..., D_n$. The formula for finding the importance measure is as follows:

$$I.M = \frac{CB_j(D_n)}{\sum_{i=1}^{n} D_i}; where, \ j = 1,2, ... and \ n = 1,2, ...$$

Where,

$I.M$ → Importance Measure

CB_j → Number of Derived Concept

D_n → Number Documents

The above formula describes that the importance measure which is the ratio of the concept matrix value of a derived concept in a particular document to the sum of the values of documents for the same derived concept.

4.7 Matrix Reduction

After finding the importance measure for all the derived concepts with its respective documents, we

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

have to reduce the matrix into one row in order to show the similarity between every concept and every document in the test set. To reduce the matrix into one row, the following formula is used.

$$CMR = C_1 + \frac{1}{2}\sum_{i=1}^{n} CB_i \; ; \quad where\, n = 1,2,......$$

Where,

CMR → Concept Matrix Reduction

C_1 → Master Concept

CB_i → Number of Derived Concepts

This formula is applied on each document separately. If we are considering the first document, we have to take the C_1 value of the first document and the CB_i denotes the values of CB_1, CB_2, CB_3, CB_4, CB_5, CB_6, CB_7 and CB_8 of first document. We have to reduce the concept matrixes which are based on the contents and index terms of the documents after applying the importance measure.

4.8. Final Resultant Matrix

After the concept matrix reduction, we have to find the final resultant matrix by combining the reduced concept matrix based on the contents of the documents and the reduced concept matrix based on the index terms of the documents. The final resultant matrix is found by the following formula:

$$FRM = \frac{W_C * CMR_C + W_I * CMR_I}{W_C + W_I}$$

Where,

FRM → Final Resultant Matrix

W_C → Content based weight value

CMR_C → Content based Concept Matrix Reduction

W_I → Index term based weight value

CMR_I → Index term based Concept Matrix Reduction

The final resultant matrix is the final matrix value for a master concept with respect to the documents $D_1, D_2, ..., D_n$. We have to find the final result matrix for all the master concepts after finding the concept matrix based on the contents of the documents, the concept matrix based on the index terms of the documents and after applying the concept matrix reduction technique. From the final result matrix of all the master concepts, we can find a document comes under which concept.

Algorithm for our Proposed Technique

This section shows the sample algorithm of our proposed technique. The step by step process is as follows:

Step 1: Start the program

Step 2: Read the document

Step 3: Remove the stop words

Step 4: Read the contents $A_1, A_2, ..., A_n$ and index terms $I_1, I_2, ..., I_n$ from the document

Step 5: Apply the master concept $C_1, C_2, ..., C_n$ in word net

Step 6: Compare the derived concepts $CB_1, CB_2, ..., CB_n$ and contents $A_1, A_2, ..., A_n$ from the documents and create the concept matrix based on contents.

Step 7: Compare the derived concept $CB_1, CB_2, ..., CB_n$ and index terms $I_1, I_2, ..., I_n$ from the documents and create the concept matrix based on index terms.

Step 8: Find the importance measure for the concept matrix which we have calculated

Step 9: Apply the matrix reduction technique after finding the importance measure

Step 10 :Find the resultant matrix after applying the matrix reduction technique and from the resultant matrix we can find the presence of concept in a document

Step11: Stop the program

5. RESULT AND DISCUSSION

This section shows the result which we got for our proposed technique and the description of the dataset which we have taken for our proposed technique and the discussion of our proposed technique.

5.1. Dataset preparation

This section explains the dataset which we have chosen for the demonstration of our proposed technique. We have chosen three documents and three master concepts for the demonstration of our proposed model. These datasets are taken based on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the concepts from the AMC Taxonomy. The master concepts which we have used for our proposed technique is taken from the AMC Taxonomy under the main keyword "Mathematics of Computing". The documents which we taken for our proposed technique is selected from the IEEE transaction based on the master concept "Error Analysis" which comes under the main keyword "Mathematics of Computing" in AMC Taxonomy. The taxon id for the master concept "Error Analysis" is G.1.0.c that comes under the main id G.1 which is "Numerical Analysis". The "Numerical Analysis" is a sub-keyword which comes under the main keyword "Mathematics of Computing". The three master concepts which we chosen for our proposed technique are taken from the sub-keyword "Numerical Analysis".

5.2 Experimental Results

This section shows the experimental result of our proposed technique. The documents and the master concepts which we have chosen from the AMC Taxonomy are then implemented in Java which uses the tool NetBeans IDE 7.0. The output of our proposed technique is shown as follows in the format of Extensible Markup Language:

<Classification = Mathematics of computing>

<Sub Classification = Numerical Analysis>

<Document Name =D1 >

<Concept Name =error analysis, >

<Percentage Value =0.1668377823408624 >

<Concept Name =computer arithmetic, >

<Percentage Value =0.0668377823408624 >

</Document 1>

<Document Name =D2 >

<Concept Name =error analysis, >

<Percentage Value =3.443526170798898E-4 >

<Concept Name =computer arithmetic, >

<Percentage Value =0.1008377823408624 >

</Document 2>

<Document Name =D3 >

<Concept Name =error analysis, >

<Percentage Value =1.718213058419244E-4 >

<Concept Name =convergence stability, >

<Percentage Value =0.000377823408624 >

</Document3>

</Sub Classification>

</Classification>

The above Extensible Markup Language (XML) explains as follows: the first line of the XML is the starting point which denotes the main keyword "Mathematics of computing" which we gave the name as 'Classification'. The second line of the XML denotes the sub-keyword "Numerical Analysis" which we gave the name as 'Sub Classification'. The third line of the XML denotes the starting point of the process of first document. In the first document, we got the percentage value as 0.1668377823408624 for the master concept "Error Analysis" and 0.0668377823408624 for the master concept "Computer Arithmetic". The eighth line of the XML shows the ending process of the first document. In the second document, we got the percentage value as 3.443526170798898E-4 for the master concept "Error Analysis" and 0.1008377823408624 for the master concept "Computer Arithmetic". In the third document, we got the percentage value as 1.718213058419244E-4 for the master concept "Error Analysis" and 0.000377823408624 for the master concept "Convergence Stability". The twentieth line of the XML shows the end process of the third document and the next line shows the end process of the sub classification and the last line shows the end point of XML.

5.3. Discussion of our Proposed Technique

The experimental result of our proposed technique shows that the presence of the master concept "Error Analysis" has more value when compared to the other master concepts in the first document. In the second and the third documents also the master concept "Error Analysis" has more values compared to the other master concepts which we have chosen for our demonstration.

Our technique is useful for e-learning to find the level of presence of the concepts in a particular document. If we need to find the documents in the search engine based on a keyword or concept, it is essential to display the documents which are more related to that keyword or concept which we are giving. The search engine would display the documents which have the value of given concept or keyword as high. If the documents are classified based on the values of the concepts present in the document i.e. the document has certain value for the first concept, second concept, etc., it would be easier for the people who are doing e-learning. Our technique classifies the percentage level of presence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of concepts in a document and it is more useful for e-learning.

6. CONCLUSION

In this paper, we have proposed a technique to design and to develop the semantic annotation model for e-learning documents and to find the presence of the concepts in the documents. In our proposed technique, we have applied the stop word removal technique in the given documents to make the documents suitable for further processing and we have applied the master concepts in the word net to find the derived concepts to calculate the concept matrix based on contents in the documents and the index terms in the documents. The resultant matrix is found after applying the importance measure and matrix reduction technique in the concept matrixes. The resultant matrix is then used to find the percentage level of the concepts present in the documents.

REFERENCE

[1] Mihaela M. Brut, Florence Sedes and Stefan Daniel Dumitrescu, "A Semantic-Oriented Approach for Organizing and Developing Annotation for E-Learning," *IEEE Transactions on Learning Technologies*, vol. 4, no. 3, pp. 239-248, 2011.

[2] Hong Cui, "CharaParser for fine-grained semantic annotation of organism morphological descriptions," Journal of the American Society for Information Science and Technology, vol. 63, no. 4, pp. 738-754, 2012.

[3] Tong-Zhen Zhang, Rui-Ming Shen, "Learning objects automatic semantic annotation by learner relevance feedback," *Biomedical Engineering and Informatics*, pp. 1-4, 2009.

[4] Rim Faiz, Boutheina Smine, Jean-Pierre Descles, "Relevant learning objects extraction based on semantic annotation of documents," *2nd International conference on web intelligence, Mining and Semantics*, no.47, 2012.

[5] Jain.S, Pareek.J, "KeyPhrase extraction tool for semantic metadata annotation of learning materials," *International conference on signal processing system*, pp. 625-628, 2009.

[6] Jain.S, Pareek.J, "Automatic topic identification from learning material: an ontological approach," *Computer Engineering and Applications*, vol. 2, pp. 358-362.

[7] Mark J. Weal, Danius T. Michaelides, Kevin Page, David C. De Roure, Eloise Monger and Mary Gobbi, "Semantic annotation of ubiquitous learning environments," *IEEE Transaction on learning technologies*, vol. 5, no. 2, 2012.

[8] N. Dovrolis, T. Stefanut, S. Dietze, H.Q. Yu, C. Valentine and E. Kaldoudi, "Semantic annotation and linking of medical educational resources," *5th European IFMBE Conference*, vol. 37, pp. 1400-1403, 2011.

[9] M. Brut, F. Sedes, T. Jucan, R. Grigoras and V. Charvillat, "An ontology based modeling approach for developing a comprehencies oriented collective intelligence," *proc. ED-L2L Conf. World Computer Congress*, 2008.

[10] Tallis.M, "Semantic word processing for content authors", *second international conference on knowledge capture*, 2003.

[11] Bayerl P.S, Lungen H, Gut U and Paul K.I, "Methodology for reliable schema development and evaluation of manual annotation," *workshop on knowledge markup and semantic annotation at the second international conference on knowledge capture*, 2003.

[12] Brin.S, "Extracting Patterns and Relations from the World Wide Web," *WebDB Workshop at 6th International Conference on Extending Database Technology*, 1998.