An Approach to Discovery and Re-ranking of Educational content from the World Wide Web using Latent Dirichlet Allocation

Jagadish V
DCSE, CEG,Anna University

Hariharan G
DCSE, CEG,Anna University

Geetha T V
DCSE, CEG,Anna University

ABSTRACT

With tremendous increase in the amount of digital data available educators are forced to author content for learning and teaching for use in their classes. With that there has emerged a need to facilitate automatic discovery of learning resources from the World Wide Web. In this work, we present a novel approach for discovering content from the web for e-learning. We argue that for an e-learning scenario, retrieval of the redundant content from the web is a serious problem to be addressed as it does not satisfy the requirements of a typical learner. Furthermore, the content retrieved should cover all topics as in his syllabus. Sense-disambiguation should be performed during information retrieval from the web so that it corresponds to the learner's actual domain of interest. This work presents a domain ontology based re-querying approach for query expansion to discover content from open corpus sources. We use the Latent Dirichlet Allocation Model for unsupervised classification of document segments to aid students and educators. Having identified the topics at the granularity of document segments in an unsupervised fashion, we state that internal topic transitions in a resource retrieved from the web can be exploited for providing relevant and personalized content. In addition to this, we propose a re-ranking scheme for ordering results from search engines to maximize topic coverage and minimize redundancy among retrieved results. We also evaluate the effectiveness of our proposed method for information retrieval and show that our work results in greater coverage of topics from the web without redundancy.

General Terms

E-learning; Information retrieval.

Keywords

Latent Dirichlet Allocation; Re-ranking; Topic coverage; Reudundancy.

1. INTRODUCTION

The World Wide Web is a huge repository of freely available content which can be used for a learning environment. The web houses several learning objects. Such content suitable for presentation to a learner can be obtained from tutorials, community edited documents, reports, learning repositories, research papers which are available for free on the web. However, such a vast pervasive resource has not yet been fully exploited towards development of personalized e-learning systems for educators and students [8].

Boyle [2] describes the development of learning objects for educators as an arduous task. The fact that proprietary learning resources are always primarily restricted to the learning environment in which they are created compounds this problem further. In addition to this, the instructor spends a lot of time and development effort in discovery and presentation of content to the learner. Gary Marchionini et al. [3] address the costs involved in the authoring of learning resources by educators. Thus, the development, reuse and presentation of suitable content from the World Wide Web pose challenges in the arena of educational data mining.

Brusilovsky et al. [1] argue that with the emergence of advanced search engines and tools, discovery materials for e-learning may not be a problem. However, the resources that one discovers might have varying styles and audience. They might lack complete coverage of topics which the instructor actually requires for content authoring. Furthermore, many resources which are retrieved are highly redundant. Hence, we state that a re-ranking of results could help in minimizing redundancy among retrieved content. In addition to this, a learning resource can be composed of several document segments. We argue that determining the topic of each of the underlying document segments in a document can enable personalized content retrieval.

The main motivation of this research is to facilitate unsupervised discovery and classification of relevant content from the web for e-learning. The work aims to help the educators in development of content from the web and reduces the overhead incurred by the educators as a part of content authoring. Hence, instructors can focus on development and design of course curricula instead of authoring content for e-learning.

The reminder of this paper is structured as follows. Section II discusses the related previous work done in the area of content retrieval and annotation for e-learning. This is followed by a listing of the different phases in the proposed methodology. The paper also describes the challenges involved in content retrieval for e-learning. The paper describes the basics of the Latent Dirichlet Allocation Model proposed by Blei et al [7] and its relevance for usage in e-learning. The paper justifies the rationale behind segment level topic identification in a document to provide personalized content. This is followed by the argument that the results from the LDA generative model could be used for a quantitative measurement of the topic coverage of the retrieved results for specific topics available in the syllabus.

The paper then goes on to discuss the proposed re-ranking scheme for ordering results based on LDA from search engines. Finally, a performance evaluation of the proposed methods for information retrieval of e-learning content is presented to justify that the method reduces redundancy and increases coverage of topics in the syllabus.

2. REVIEW OF RELATED LITERATURE

Previous research in the area of information retrieval for e-learning reviews the relevance of open corpus content and the content available in knowledge repositories for use in education.

Related works [1],[2],[4] describes various challenges involved in re-use of content from web for e-learning. Steichen et al [4] state that the problem of accurate identification of content from open corpus sources has led to the development of many open source implementations of web based information retrieval systems eg: Lucene[5], Nutch [6]. But, such IR systems suffer a significant drawback that they retrieve web pages based on the relative relevance of the retrieved page to the query of the learner and therefore, cannot provide a description of the web page retrieved [4]. Steichen [4] also explains that an IR system along with an automated indexer and an annotator can assist in supporting dynamic hyper text generation.

Challenges to development of a suitable information retrieval system for e-learning include content caching for further data analysis [4]. Furthermore, meta-data must be generated to tag the obtained data for use by the personalization modules. Other challenges include ensuring document cohesion in the learning object/resource presented to the learner. Cohesion ensures that the learning resource explains about a single topic. The research by Seamus Lawless [8] addresses the problems and requirements of an e-learning system that employs content from open sources. Discoverability of suitable and relevant content from the web may be difficult because information required for e-learning requires an organization among the content and the vast quantity of content from WWW is seldom organized and exhibit variations in purpose, topic and format [8].

Several strategies such as unsupervised text segmentation, topic detection, query expansion have been identified to deal with these issues in isolation in the area of text mining. These must be integrated to produce a semantic aware content discovery system which can facilitate presentation of educational content relevant to the learner by unsupervised classification and annotation of learning resources .

When dealing with content retrieval from the web for educational domains, the usage of simple TF-IDF (Jones et al.) [9] for determining terms for query expansion is insufficient and some kind of supervisory structure is needed for discovery of associated concepts. Furthermore, a simple TF-IDF based re-querying scheme can miss some potentially relevant sub-concepts associated with the query term. We argue that, for an e-learning system, such a case is not admissible as it results in incomplete coverage of topics to the learner. Usage of an ontology for guided querying can mitigate this effect.

Any web search works on the principle of ranking the pages in www. This is one area which has drawn much attention recently due to the proliferation of search engines and many algorithms for this purpose have been developed. PageRank algorithm[10] is one such which is content independent and focuses on the position of the page in the graph of www alone. Yaltaghian et al. [11] proposed a set of 21 measures based on network analysis and showed significant improvement over google[12] in the relevance of results. But such generic measures can only aid to a certain extent in such a content-dependent area as e-learning.

Classification of the retrieved documents and annotation follows information retrieval in building an e-learning system. Document classification is a well researched area. Simple algorithms such as TF-IDF [9] use the frequency of terms alone. Latent semantic indexing [13] is a technique that utilizes singular vector decomposition and considers not only the frequency of the terms but also the associations between them. But algorithms involving probabilistic analysis increase the effectiveness of classification. In this research, we employ the Latent Dirichlet allocation (LDA) [7] which is a recently developed generative model that is based on the bag of words approach and uses Dirichlet distributions to determine the context/topic of a given word in a given document and thus ultimately the topic mixture of each document. Biro [14] details an approach for using LDA for document classification.

3. MOTIVATION

In this section we discuss certain issues addressed by the proposed system for e-learning. Seamus Lawless et al. [15] discuss an implementation of an e-learning system using a focused web crawler and an indexer. Their approach to discovery of educational content from the web is driven by a set of keywords in a file. However, such an approach which is entirely key-word driven shall fail to identify associated words. Hence, an ontology based query expansion scheme is needed because it incorporates some additional knowledge about relationships existing between the terms in the syllabus. We claim that such relationships play a pivotal role in an e-learning kind of a scenario where a learner is expected to derive maximum knowledge by assimilating all the concepts in his topic of interest. For example, when querying the web for learning objects pertaining to operating system memory the keyword driven search [15] might not identify "operating system paging" as an associated word because it is oblivious to the fact that a part-of relationship exists between a pager and an operating system.

Furthermore, incorporating domain knowledge in the form of an ontology can perform sense-disambiguation in the retrieved content. For instance, the query 'process management' when given to a search engine returns more results corresponding to business process management and industrial process management compared to the actual domain of interest which is 'computing'. Hence, an ontology or a knowledge base which relates operating system and process management by a suitable 'function of' relationship can result in retrieval of relevant content.

The work by Steichen [4] requires a content authoring phase and requires manual intervention to tag and annotate content. It assumes content authoring to be done by means of crowd sourcing to annotate documents. However, such crowd-sourcing annotations are inherently subjective and vary depending upon the user's perception. There is no guarantee that users tag content accurately. Furthermore, this approach requires time on the part of the person who annotates the learning material. We make use of the Latent Dirichlet Allocation model for unsupervised segment level document classification.

Conventional e-learning systems do not take into account topic transitions between the segments of a single document which considerably affects the cohesion in the learning object presented to the learner. Our approach operates on the granularity of document segments as compared to other approaches which operate at the document level. Considering gradual topic transitions across segments in a document shall be useful in providing more appropriate content to the learner. For example, if the system has apriori knowledge that a learner has studied 'semaphore' then documents which make a gradual transition from 'semaphore' to a topic unknown to the learner might be presented to him.

In addition to this, the main problem with discovery of content from the web is the redundancy of the retrieved results. We use the LDA model to build a vector space of topics corresponding to the domain of interest. Then, retrieved set of documents are chosen so as to minimize redundancy with the existing corpus. A cosine-distance similarity measure is considered for re-ranking.

Also, in the case of information retrieval for e-learning one must consider coverage of topics beyond a threshold. We assume the length of a document (after the removal of its stop words) to be indicative of the document's coverage of the topic.

4. ARCHITECTURE

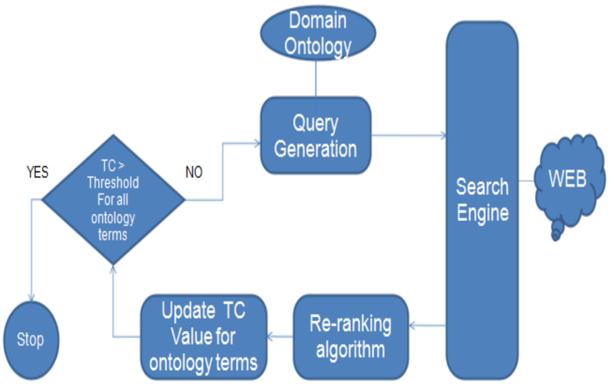


Figure 1- Content Retrieval for e-learning

Figure-1 describes the proposed architecture. The input is an ontology corresponding to the domain of interest. The algorithm then performs query expansion and classification of results obtained from the web at the level of segments. The LDA generative model to give the probability of association of a word and a topic.This is followed by re-ranking of results in a topic space where the dimensions are formed by the terms in the LDA model.

The proposed methodology for retrieval of content from web for e-learning consists of 2 phases

4.1 Query generation and expansion :

The query expansion phase involves a TF-IDF based re-querying to the search engines and also an ontology based query expansion for better content coverage from the web. The top 'k' terms ordered by their TF-IDF values are obtained from each search. And, the obtained terms are concatenated with the original query term to generate a new query. For example, our search on operating systems generated 'kernel' following the previous method. Hence, the new query would be generated as 'operating systems kernel' to the search engine.

As web pages are retrieved they are saved and an LDA based classification is done on the web page. The web pages retrieved are divided into several segments. (By segment, we mean a paragraph separated by a newline character). From the LDA model (which is generative in nature) we obtain the confidence level that the segment is associated with a given topic. The topic with the maximum probability is chosen as the topic of the segment and the segment is annotated suitably. Cohesion among learning material is important for a learner. Hence, consecutive segments which have the same topic are suitably coalesced so that they are cohesive. The topic coverage of the corresponding topic (initially set to zero for all topics) is incremented by the length of the segment. Once the topic coverage of the topic queried exceeds a threshold, the TF-IDF based query expansion is stopped and the next term in the ontology is used to expand the query.

For example , if a part-of relationship exists between operating system and memory management then, the next query proceeds as " operating system + memory management ". The same process is repeated for every concept in the syllabus.

The rationale behind paragraph level annotation is to exploit topic transitions in a document for providing relevant topic to the learner. From the topic of the individual segments in the document, the dominant topic of the entire document is inferred.

Algorithm1 : (Content Retrieval from Web)

Input:
A domain Ontology, a topic 'T' to begin the query,t_hold: threshold parameter for topic coverage.

Output:
A corpus of documents of all concepts associated with 'T', topic coverage vector Di for each document .

Process:

```
While(terms remain in ontology)
Do
$T_0$ := next term from domain ontology
Use $T_0$ to query the web
$L_1$: = Re-ranked list of the top n results by
        using algorithm 3.
<Document set> D = top k documents from L1.
For each Document $D_i$ in D
        For each Segment 's' in $D_i$
            Use LDA to annotate document segments.
             len = Segment.length
             Di ($T_0$) += len.
         end For
end For
Obtain TF-IDF for Di.
List L2 = dominant terms of Di
 For each term 'ti' in L2
             If (topicCoverage Di ($T_0$)) < t_hold)
              Requery using ( Topic + ti )
     end For
end Do
```

For our research we employ the LDA model [7] proposed by Blei et al. The LDA is a generative probabilistic model for modeling a corpus of text. It is a hierarchical Bayesian model where in an item is modeled as a mixture over an underlying topic set each topic in which is inturn viewed as an infinite mixture over an underlying text of topic probabilities [7]. Interested readers are advised to read through [7] for a comprehensive overview of LDA.

Algorithm 2 : (Segment Level Topic annotation using LDA)

We use a Score_matrix as a data structure for the algorithm. Score_matrix[i,j] is a measure of the confidence level for which the I th segment is associated with the J th topic. The function segment_coalesce(I,J) concatenates the segment I and J.

Input:
The result of running the LDA model on a training set of documents (the form of a triple <topic_id,word,probability>) along with the corpus retrieved by algorithm 1.

Output:
Segment level annotations of the document and the dominant topic of the entire document.

Process:

```
For each Document Di in D
    For each Segment 's' in D
        For each Tuple <Wi, Pi,'Ti'> in the LDA Model,
          if (s.contains(Wi))
          Increase Score_matrix[si,Ti] by score_val = Pi.
        end For
Order each topic by its weight in the Score_matrix.
ts = maximum scored topic for current segment 's'.
prev_topic = maximum scored topic for segment 'si-1'
if (prev_topic == ts )
  Segment_coalesce( s, prevs )
else
  Indicate topic transition from  si to si-1
end For
end For
```

4.2 Re-ranking and Document Modelling

Most resources retrieved from the web might be redundant. Presentation of such redundant content to the learner might result in decrease of learner productivity. For re ranking, we use a topic vector space model [12] by Jorg Becker and Dominik Kuropka.

A document is represented as a point in a 'k' dimensional vector space of which each dimension is a topic. The weights associated with the individual topics in the document are obtained using the LDA generative classification (Algorithm 2). Thus every document is represented as a vector of topics with weights corresponding to individual topics. The new set of documents which are retrieved are ordered so as to maximize their cosine distance measure from the existing set of documents. The Cosine similarity measure is defined in literature as,

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

For example, consider a document which contains a line – "Paging is a function of Operating System". The output of the LDA model (Figure2) indicates that paging (page after stemming) has a high probability to be generated from the topic 'memory management.' Let the probability that the term 'paging' is generated from the topic memory management be p. Hence, the dimension in the document corresponding to 'memory management' is increased by p. The same process is repeated to obtain the weight contributed by the entire document to each topic dimension. And we model a single document in the vector space as

<Topic 1> weight 1 </topic1> <topic 2> weight 2 </topic 2> <topic 3> weight 3 </topic3> and so on.

Algorithm 3: Re-ranking

Input:
Unranked set of documents D, a set of stored documents S.

Output:
Re-ranked set of documents D' for e-learning.

Process:

```
For each document 'd' in D,
Model a document point 'd' as a vector in k-dimensional topic space.
For each document 's' in S,
   Calculate the sum of distances(μ) of 'd' from 's'.
    Sort all documents in D by μ in descending order
D' = Set of top 'm' obtained results .
Return D'
```

5. Evaluation

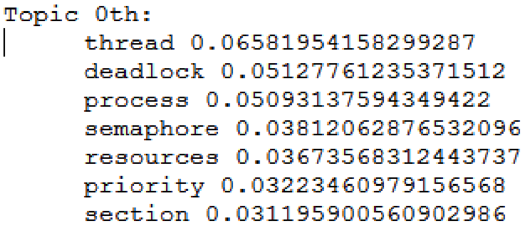
5.1. Experimental definition and Setup

Our ontology driven approach to query expansion and the re-ranking algorithm were both tested through retrieval of 500 documents from the World Wide Web. Our work to study the effectiveness of our algorithm can be divided into 2 major interleaved phases:

I) Document retrieval from the web.
II) Classification of the retrieved documents.

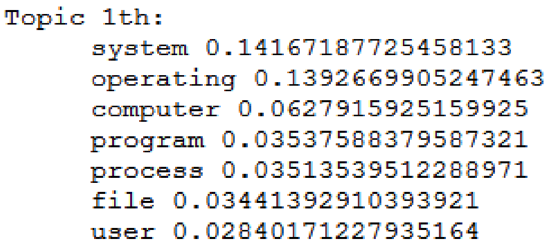
Initially we constructed an LDA model, using the Jgibb LDA([17]) package, from a training set of 350 manually downloaded documents from the web. This model was developed to be later utilized during the inference of documents retrieved by an automated content discovery system.

The following figures 2a,2b,2c show a sample snapshot of word-topic probabilities from LDA model:



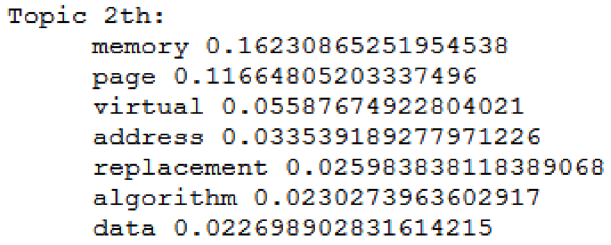
```
Topic 0th:
        thread 0.06581954158299287
        deadlock 0.05127761235371512
        process 0.05093137594349422
        semaphore 0.03812062876532096
        resources 0.03673568312443737
        priority 0.03223460979156568
        section 0.031195900560902986
```

(2a)



```
Topic 1th:
        system 0.14167187725458133
        operating 0.1392669905247463
        computer 0.0627915925159925
        program 0.03537588379587321
        process 0.03513539512288971
        file 0.03441392910393921
        user 0.02840171227935164
```

(2b)

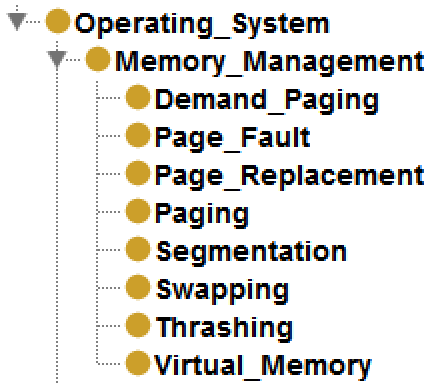


```
Topic 2th:
     memory 0.16230865251954538
     page 0.11664805203337496
     virtual 0.05587674922804021
     address 0.033539189277971226
     replacement 0.025983838118389068
     algorithm 0.0230273963602917
     data 0.022698902831614215
```

(2c)

Figures 2a,2b,2c-Topic word probabilities from the LDA model

In the first phase, we constructed an ontology, using the protege ontology editor[18], for the domain of 'Operating Systems'.



Operating_System
Memory_Management
Demand_Paging
Page_Fault
Page_Replacement
Paging
Segmentation
Swapping
Thrashing
Virtual_Memory

Figure 3-Snapshot of Ontology

This ontology was used to identify sub-concepts which can be used to perform guided expansion of queries in addition to the terms determined by TF-IDF. The expanded query was utilized to search the web through google.

For each query, from the results obtained from the search, we first retrieved the top 10 documents. We arrived at the value 10 after extensive observation led us to the conclusion that the relevance of the results, for searches from an educational point of view, was found to reduce significantly when more results were considered.

Each document in the retrieved set was then preprocessed using stemmer and stop words are removed. Then it is classified according to the algorithm 2 described above . This classification occurs at the granularity of segments and the position of the document in the term vector space is identified. For each document its degree of coverage of the terms was stored in xml format so as to enable easy retrieval of that information.

These top documents in each set were then re-ranked according to the sum of the distances, in the topic vector space, from the currently existing corpus of documents in the topic vector space. The top 5 results were chosen from the re-ranked list and the corresponding documents are added to the corpus with the coverage being updated accordingly. This process was repeated for complete coverage of topics.

To verify the effectiveness of our approach, we then performed the same steps, but without the re-ranking and the usage of ontology, and taking the top results from the google search as such and performing query expansion with the usage TF-IDF.

Then we compared the two approaches on their diversity of content and their degree of coverage of the various terms from the LDA model.

5.2 Results

The two measures which we chose to compare our approach with the baseline approach were:

I) Coverage of concepts: The degree to which each approach covers the top keywords that were found to belong to each topic during the LDA model construction run(done using manually accumulated documents).

$$\text{Coverage} = \frac{No.of\ concepts\ covered}{Total\ no.of\ concepts\ in\ the\ pedagogy}$$

A concept is said to be covered if along the associated dimension on the topic vector space the weight is non-zero.

II) NR (Non- Redundancy) measure: which we define to be the sum of the cosine distance values of the set of documents retrieved from the currently existing corpus of documents. We consider this to be a optimal measure of the dissimilarity of content in the document corpus.

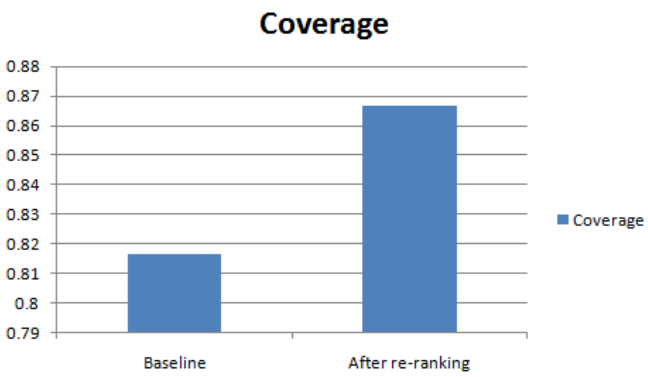


Figure 4-Comparison of topic coverage

The usage of ontology shows significantly improved coverage of topics over the simple TF-IDF approach as can be clearly seen from the graph above. The benefit brought about by the usage of ontology-driven query expansion can be clearly noticed.

The following graph signifies the degree of improvement that the re-ranking scheme brings to the measure of diversity of the content that we retrieve and store in the document corpus. Re-ranking is shown to bring about a 20% enhancement to the degree of non-redundancy of content.

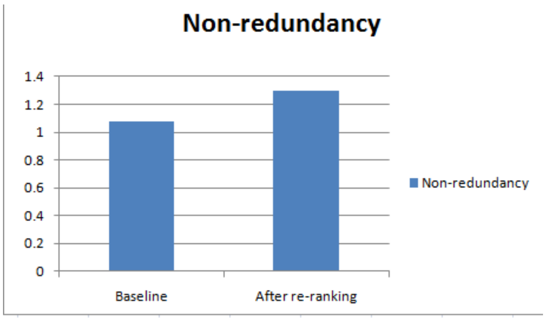


Figure 5- Comparison of NR measure

6. CONCLUSION AND FUTURE WORK

In this paper we have presented a novel ontology-driven approach for the discovery of educational content from the web. Also, we have proposed an algorithm for re-ranking the results obtained from web search based on the redundancy of content, so as to enhance the diversity of the content retrieved. Our implementation utilizes the latest improvements in the field of content retrieval from the web and combines them with our proposed techniques to develop an effective system for retrieving web content and classifying them by building a term vector space. The degree of improvement that our proposed approach brings to content retrieval is significant as can be noticed from the evaluation above. Also, the classification was done at the level of document segments and we succeeded in obtaining the sequence of topics in each document.

In our work we have restricted ourselves to the domain of 'Operating Systems' and have proved the level of improvement brought about by our algorithm. Construction of ontologies in sufficient detail can enable the system to be used for other domains also. Further work can be done in this area.

Also we have identified for each document the sequence of topics in it. This set of topic transitions of each document can be utilized to improve the presentation of content to the learner by enhancing personalization. The coverage of various topics by the documents has been stored in an xml file format of our own creation. Standardized formats such as SKOS or SCORM so as to enable easy exchange of this metadata among e-learning systems.

Also, on the side of performance evaluation our 'Coverage' parameter can be replaced by more sophisticated content richness measures.

7. REFERENCES

[1] Brusilovsky, P. & Henze, N., 2007, "Open Corpus Adaptive Educational Hypermedia". In The Adaptive Web: Methods and Strategies of Web Personalisation, Lecture Notes in Computer Science, vol. 4321, Berlin: Springer Verlag, pp. 671-696.

[2] Boyle T., 2003, "Design Principles for Authoring Dynamic, Reusable Learning Objects". In the Australian Journal of Educational Technology, vol. 19(1), pp. 46-58.

[3] Gary Marchionini, 1995, "The Costs of Educational Technology: A Framework for Assessing Change". Invited paper at ED- MEDIA 95 in Graz, Austria.

[4] Steichen, B., Lawless, S., O'Connor, A. & Wade, V. 2009, "Dynamic Hypertext Generation for Reusing Open Corpus Content" In the proceedings of the 20th ACM Conference on Hypertext and Hypermedia, Hypertext, in Torino, Italy.

[5] Apache Lucene - a free/open source information retrieval

software library, originally created in Java. It is released under the Apache Software License.

[6] Nutch an open source search engine based on Lucene at

nutch.apache.org.

[7] David M. Blei, Andrew Y. Ng, Michael I. Jordan ,2003,"Latent Dirichlet Allocation",The Journal of Machine Learning Research,3, pp. 993-1022.

[8] Lawless, S.,2009, "Leveraging Content from Open Corpus Sources for Technology Enhanced Learning". Ph.D. Thesis, Submitted to the University of Dublin, Trinity College.

[9] Sparck-Jones, K. ,1972,. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28(1), pp. 11-21.

[10] Page, Lawrence; Brin, Sergey; Motwani, Rajeev and Winograd, Terry, 1999 , "The PageRank citation ranking: Bringing order to the Web".

[11] Behnak Yaltaghian,, Mark Chignell, 2002,"Re-ranking search results using network analysis a case study with google: a case study with Google",CASCON '02 Proceedings of the 2002 conference of the Centre for Advanced Studies on Collaborative research, pp. 14.

[12] Search engine available at www.google.com

[13] Jian-Tao Sun; Zheng Chen; Hua-Jun Zeng; Yu-Chang Lu; Chun-Yi Shi; Wei-Ying Ma; 2004 ,"Supervised latent semantic indexing for document categorization",Proceedings of IEEE International conference on Data Mining, pp. 535-538.

[14] Istvan Biro, 2009, "Document Classification with Latent Dirichlet Allocation". Ph.D Thesis, Eötvös Loránd University.

[15] Lawless, S., Hederman, L. & Wade, V.,2008, "OCCS: Enabling the Dynamic Discovery, Harvesting and Delivery of Educational Content from Open Corpus Sources" In the proceedings of the 8th IEEE International Conference on Advanced Learning Technologies, I-CALT 2008, in Santander, Cantabria, Spain.

[16] J. Becker, D. Kuropka, 2003,""Topic-based Vector Space Model" , In Proceedings of the 6th International Conference on Business Information Systems, pp. 7-12.

[17]JGibb LDA, Package available at http://jgibblda.sourceforge.net

[18]Protégé ontology editor available at http://protege.stanford.edu/

8. BIOGRAPHY

Jagadish V is a part of the Language Technologies Research Group at College of Engineering, Guindy, Anna University, Chennai. He is a senior undergraduate student at CEG collaborating with Professor TV Geetha. His areas of interest include text mining, web information retrieval and multi-core architectures. His thesis focused on developing text mining techniques for discovery of educational content from the web.

He received the best means and merit student award from Anna University for the year 2010

Dr.T.V.Geetha is a Professor in the Department of Computer Science and Engineering at College of Engineering, Guindy. Her areas of interest include Natural Language Processing, Intelligent Databases and Data mining. She has over 50 reputed conference and journal publications to her credit. She held the position of the Head of the Department of Computer Science and Engineering from 2003 to 2006. She currently serves as the Chair of the faculty of Information and Communication Engineering at Anna University.