International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016 [ICSSCCET 2016]

ISBN	978-81-929866-6-1	**VOL**	02
Website	icssccet.org	**eMail**	icssccet@asdf.res.in
Received	25 – February – 2016	**Accepted**	10 - March – 2016
Article ID	ICSSCCET070	**eAID**	ICSSCCET.2016.070

A Novel Approach on Learning Material Retreival in E-Learing System

Ekta Jalan[1], Chitra R[2], Dhivya P S[3], Manoj Kumar D[4], Ramesh P N[5]
[1,2,3,4]Final Year B.E (CSE), [5]Assistantt Professor (CSE), Karpagam Institute of Technology, Coimbatore, Tamilnadu, India

Abstract--*Nowadays retrieving books, articles and tutorials for learning any topics or technology is a challenging task. Learners struggle to find relevant study materials to their interested topic or technology. Our work assists learners to find their relevant study materials in an e-learning system.*

Keywords: *Similarity measure, Cluster, Search, Filter*

I. INTRODUCTION

Institutions and universities are developing e-learning system which helps students to retrieve the study materials required by them. In most of e-learning system students can retrieve their books by giving book name or an author name as keyword. In other systems students can find their books by navigating through the index of the e-learning system. Both of the above method to find the study materials is inadequate to learners when they want to refer about specified topic.

For example, if a learner wants to know about deadlock recovery technique and learner does not know the book name which contains detail about deadlock recovery. In this case searching book or other study materials with book name and author name is not suitable. Our novel approach will retrieve the books or other study materials which contains the details about the specified topic.

II. Problem Definition

E-learning system database D contains study materials , ,….,

D = { , ,…., }

Retrieve K study materials { , ,…., B } which contains detail about the keywords { , ,…., }.

III. Similarity Measures in Material Retrieval

Most of the e-learning systems are providing results in the form of text study materials, that materials might be in different file formats like html, pdf , doc and etc., while retrieving study materials e-learning system has to evaluate similarity between the those study materials for the following reasons,

1. Eliminating nearest duplicates
2. Ranking study materials based on relevance to search keyword
3. Grouping similar study materials together

Cite this article as: Ekta Jalan, Chitra R, Dhivya P S, Manoj Kumar D, Ramesh P N. "A Novel Approach on Learning Material Retreival in E-Learing System". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 370-372. Print.

This section illustrates how similarity measure is evaluated between two study materials with the scenario tabulated in table 3.1.In study material comparison, frequency of interested terms or words has been taken as dimensions.

Table 3.1: Study materials & Term's count

Study material name	Total number of word in study material	Count of term 1	Count of term 2
A	1800	360	540
B	1900	300	250

With the above table, term frequency tf_1 and tf_2 can be calculated for both the two study materials as follows.

tf= tc/D

Where,

tf is the term frequency of a term in a study material
tc is the count of a term in a study material
D is the total no. of words in study material
Term frequency tf_1 and tf_2 for study materials a and b are tabulated in table 3.2

Table3.2: Study materials & Term's frequency

Study material name	Term 1 frequency tf_1	Term 2 frequency tf_2
A	0.2	0.3
B	0.157	0.13

To compute similarity between these study materials, all study materials are represented as vectors in two dimensional spaces. The vector $t_x=(dx_1, dx_2)$.

Where,

X is name of the study material
dx_1 (dimension 1) is term 1 frequency in study material X.
dx_2 (dimension 2) is term 2 frequency in study material X.
Hence study materials a, b are represented as vectors t_a, t_b respectively.
$t_a = (da_1, da_2) = (0.034, 0.027)$
$t_b = (db_1, db_2) = (0.157, 0.13)$
First we present cosine similarity measure with the above scenario for better understanding.

A. Cosine Similarity

$$sim_c(t_a, t_b) = (t_a * t_b)/(|t_a| * |t_b|)$$

This is measured as a cosine angle of the two vectors. The cosine similarity for (t_a, t_b) in our scenario is 0.87.

With this cosine similarity we can cluster the study materials which contain the specified topic by the learners.

IV. Experiment

We have implemented the e-learning system using ASP.NET. For maintaining master data we use MS SQL server 2008. We have implemented the search module by computing keyword frequency in each study material by linear accessing. We cluster the study materials which are similar and contain specified keyword.

The e-learning system produces the books containing those specified keywords and filter the dissimilar books using similarity measure. It produces the results 95% accurately.

V. Conclusion

Similarity measures hire on major task of information retrieval process. Searching, filtering the results and clustering data's are required similarity measures. Nowadays various advanced algorithms introduced for search and clustering study materials. We are going to use cosine similarity measure and evaluate the accuracy of this measure. This system will be more useful for research students.

Cite this article as: Ekta Jalan, Chitra R, Dhivya P S, Manoj Kumar D, Ramesh P N. "A Novel Approach on Learning Material Retreival in E-Learing System". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 370-372. Print.

References

1. E. Losif and A. Potamianos, "Unsupervised Semantic Similarity Computation between Terms Using Web Documents," IEEE Trans. Knowledge and Data Engineering, vol. 22, no.11, November 2010.
2. A.Huang, "Similarity Measures for Text Document Clustering," Proc. New Zealand Computer Science Research Student Conference, April 2008.
3. J. Han and M. Kamber, "Data Mining Concepts and Techniques second edition," Elsevier publications 2006.
4. T. Zhang, Y. Y. Tang, B. Fang and Y. Xiang , "Document Clustering in Correlation Similarity Measure Space," IEEE Trans. Knowledge and Data Engineering, vol. 24, no.6,June 2012.
5. Y. S. Mudhasir, J. Deepika, S. Sendhilkumar and G. S. Mahalakshmi , " Near-Duplicates Detection and Elimination Based on Web Provenance for Effective Web Search," International Journal on Internet and Distributed Computing Systems,vol.1 no.1,2011.
6. E. Petrakis, G. Varelas, A. Hilaoutakis and P. Raftopoulou, "X-Similarity: Computing Semantic Similarity Concept from Different Ontologies," J. Digital Information Management vol.4, no.4, pp: 233-238, 2006.
7. J. Jiang and D. Conrath, "Semantic Similarity Based on Corpus Statics and Lexical Taxonomy," Proc. Int'l Conf. Research Computational Linguistics, 1997.
8. Prasanna Kumar J and Govindarajulu P, "Duplicate and Near-Duplicate Documents Detection," European Journal of Scientific Research, vol.32, no. 4, pp: 514-527, 2009.

Cite this article as: Ekta Jalan, Chitra R, Dhivya P S, Manoj Kumar D, Ramesh P N. "A Novel Approach on Learning Material Retreival in E-Learing System". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 370-372. Print.