A Novel Approach on Learning Material Retrieval in E-Learning System

Ekta Jalan¹, Chitra R², Dhivya P S³, Manoj Kumar D⁴, Ramesh P N⁵
¹²³ Final Year B.E (CSE), ⁴Assistant 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 systems which help students to retrieve the study materials required by them. In most of the e-learning systems 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 methods 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 , , ...., 

\[ D = \{ m_1, m_2, ..., m_n \} \]

Retrieve K study materials \{ m_1, m_2, ..., m_k \} which contains detail about the keywords \{ k_1, k_2, ..., k_n \}.

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

This paper is prepared exclusively for International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016 [ICSSCCET 2016] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr. T Ramachandran and Editors Dr. Daniel James, Dr. Kokula Krishna Hari Kanasekar and Dr. Sakikbeere Elangoavan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

2016 © Reserved by Association of Scientists, Developers and Faculties [www.ASDF.international]

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

<table>
<thead>
<tr>
<th>Study material name</th>
<th>Total number of word in study material</th>
<th>Count of term 1</th>
<th>Count of term 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1800</td>
<td>360</td>
<td>540</td>
</tr>
<tr>
<td>B</td>
<td>1900</td>
<td>300</td>
<td>250</td>
</tr>
</tbody>
</table>

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

\[ tf = \frac{tc}{D} \]

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.

### Table 3.2: Study materials & Term’s frequency

<table>
<thead>
<tr>
<th>Study material name</th>
<th>Term 1 frequency ( tf_1 )</th>
<th>Term 2 frequency ( tf_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>B</td>
<td>0.157</td>
<td>0.13</td>
</tr>
</tbody>
</table>

To compute similarity between these study materials, all study materials are represented as vectors in two dimensional spaces. The vector \( \mathbf{x}_X = (dx_1, dx_2) \).

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 \( \mathbf{t}_a, \mathbf{t}_b \) respectively.

\[ \mathbf{t}_a = (da_1, da_2) = (0.034, 0.027) \]
\[ \mathbf{t}_b = (db_1, db_2) = (0.157, 0.13) \]

First we present cosine similarity measure with the above scenario for better understanding.

#### A. Cosine Similarity

\[ \text{sim}_G(\mathbf{t}_a, \mathbf{t}_b) = \frac{\mathbf{t}_a \cdot \mathbf{t}_b}{\|\mathbf{t}_a\| \|\mathbf{t}_b\|} \]

This is measured as a cosine angle of the two vectors. The cosine similarity for \( \mathbf{t}_a, \mathbf{t}_b \) 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.

---

References