Efficient Search over Massive Open Online Course

1Prof. Anuradha Dandwate, 2Pankaj Pawar, 3Vishal Nimbalkar, 4Pooja Dhenge

1Professor, 2,3,4Students
Department of Information Technology
Engineering Saraswati College of Engineering, Kharghar, Navi Mumbai, India

Abstract: Due to the widespread of internet, online education is gaining popularity. In the field of education, Massive Open Online Courses (MOOC) are used in delivering learning content to any person who wants to take the course with no constraint on attendance. The courses by different providers may differ in session timing, price, difficulty level etc. Hence a user has to visit every MOOC providers site and go through the course details. To make this task user-friendly, a Information aggregator is used which can aggregate online courses from multiple course providers. Before aggregating this courses from different MOOC’s, data preprocessing is performed. And to combat the limitations of stemming, we are using lemmatization. In Information Retrieval, one of the important task is retrieving relevant information. However an important issue for retrieval effectiveness is the mismatch problem where in the indexers and users do not often use the same words. User query is often too short and may not contain relevant terms. This issues are handled by query expansion. User query terms are enriched with additional semantically related terms like synonyms using a dictionary.

We present a web portal called MOOCLink that utilizes the data to discover and compare courseware. Also providing filtering which include MOCK test based and budget based for user to choose appropriate courses.

Keywords: Massive Open Online Courses (MOOCs); K-nearest neighbor algorithm(KNN), Courseware.

I.INTRODUCTION

Internet is an important technology of the information age. It serves as a large reservoir of data from which one can retrieve required information. However, information available on Internet is not stable. At any time this information may be altered, moved or deleted which leads to a problem of finding relevant information on internet. One of the problem web is facing today is information overload. There are a large number of information sources over the web which provide similar or related information for a particular topic. It is the users job to go to each of these sources and get the required information. To effectively use this data from multiple sources it needs to be aggregated at one place

Information aggregators help to solve this problem. Aggregator is a web site or computer software that aggregates a specific type of information from multiple online sources. Due to the widespread of internet, online education is gaining popularity. In the field of education, Massive Open Online Courses (MOOC) are used in delivering learning content to any person who wants to take the course with no constraint on attendance.

Benefits of MOOC are as follows:

1. It helps students to find a right course.
2. Courses are offered for free.
3. Courses are available to large and diverse audience across the globe.
4. It provides easy access to global resources and promotes sharing of ideas and knowledge.
5. It enhances active learning.
6. MOOC provides the opportunity to learn from world class universities and from renowned instructor.
7. It promotes knowledge sharing in discussion forum.

One of the important feature of MOOC is that it provides open access to most of the courses thus making available to all. In the domain of education, there are a large number of MOOC providers such as Coursera, Udacity, Udemy etc. MOOCs are Massive Open Online Courses. They act as a medium for collaborative sharing of knowledge and unlimited participation via web. Each of these course providers may be offering similar courses at the same time. Therefore, if a user wants to take up a particular course, he has multiple choices. The courses by different providers may differ in session timing, price, difficulty level etc. Hence a user has
to visit every MOOC providers site and go through the course details. To make this task user-friendly, a Information aggregator is used which can aggregate online courses from multiple course providers.

II. SYSTEM DETAILS

Hardware Specification:
1. Processor: Pentium 4
2. RAM: 1 GB or more
3. Hard disk: 16 GB or less

Software Specification:
1. Windows Operating System.
2. Eclipse
3. MySQL
4. Java (JDK)

III. PROPOSED SYSTEM

It consists of two main components: Information Aggregator and Efficient Search Module Using Query Expansion.

1. Information aggregator collects information from different MOOC websites and utilizes the data to discover and compare online courseware.

2. Efficient Search Module using Query Expansion to include more meaningful terms and fetch the results. The MOOCs considered are: Udacity, Coursera and Udemy.
   - Udacity: In a Stanford University experiment, two of the professors Sebastian Thrun and Peter Norvig offered their "Introduction to Artificial Intelligence" course online. More than 160,000 students in more than 190 countries enrolled and this way Udacity was born.
   - Coursera: Coursera was founded in 2012 by two Stanford Computer Science professors, Daphne Koller and Andrew Ng. It is platform where anyone can learn from worlds best universities and education providers.
   - Udemy: Udemy was founded by Eren Bali of Turkey. This is a global marketplace for providing online education.

Detailed Architecture Components consists of the following modules:

- Data Extraction Module It is responsible for extracting the data from MOOC websites and delivering the extracted data to a MySQL database. It consists of 3 sub modules: 1.1: Coursera Data Extraction: Data will be fetched from Coursera using Coursera API. 1.2: Udacity Data Extraction: Data will be fetched from Udacity using Udacity API. 1.3: Udemy Data Extraction: Data will be extracted from Udemy website using web crawler.
Data Preprocessing The data needs to be cleansed before the classifier is applied. Stopwords: Stopwords are removed from data to reduce noise. Stop words are the common words that carry no information (eg. prepositions, pronouns etc). Tokenization: It is the process of breaking a stream of text into words, phrases, symbols, or other meaningful elements called tokens. Lemmatization: It uses vocabulary and morphological analysis of word and tries to remove inflectional endings. Advantages of lemmatization over stemming: 1. It returns words to their dictionary form. 2. It analyzes if query words are used as verbs or noun. 3. It also helps to match synonyms.

IV. RESULT:

![Figure 1: Student Login page](image1)

![Figure 2: Mock Test](image2)
V. ADVANTAGES

1. It helps students to find a right course.
2. Courses are offered for free.
3. Courses are available to large and diverse audience across the globe.
4. It provides easy access to global resources and promotes sharing of ideas and knowledge.
5. It enhances active learning.
6. MOOC provides the opportunity to learn from world class universities and from renowned instructor.
7. It promotes knowledge sharing in discussion forum.
VI. FUTURE SCOPE
1. Extracting data from MOOCs: Textual data about the courses from MOOC websites can be extracted using APIs and web crawlers.
2. Categorization of data collected: Manually categorizing courses into different categories is time consuming and error prone. Therefore, this process can be automated by using machine learning algorithms such as Nave Bayes classification.
3. Semantic search by query expansion: Users query may not always contain relevant words to fetch correct courses. Hence users query needs to be enriched with semantically related terms like synonyms.
4. Query Classification: Task of assigning the query to one of the predefined categories based on content of the query.

VII. CONCLUSION
Information aggregator aggregates online education courses from different MOOC providers. Therefore, it becomes easy for the users to find the relevant information at one place. But before aggregating this information, data pre-processing is performed. So to overcome the limitations of stemming, lemmatization is used. However, retrieving this relevant information is one of the important tasks in order to handle the problem of vocabulary mismatch. So we have proposed a technique called query expansion to handle this problem. It enriches query with additional semantically related terms like synonyms using a dictionary or wordnet.

References