Efficient Search over Massive Open Online Course

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

[1]Professor, [2,3,4]Students
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 exapansion. 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.

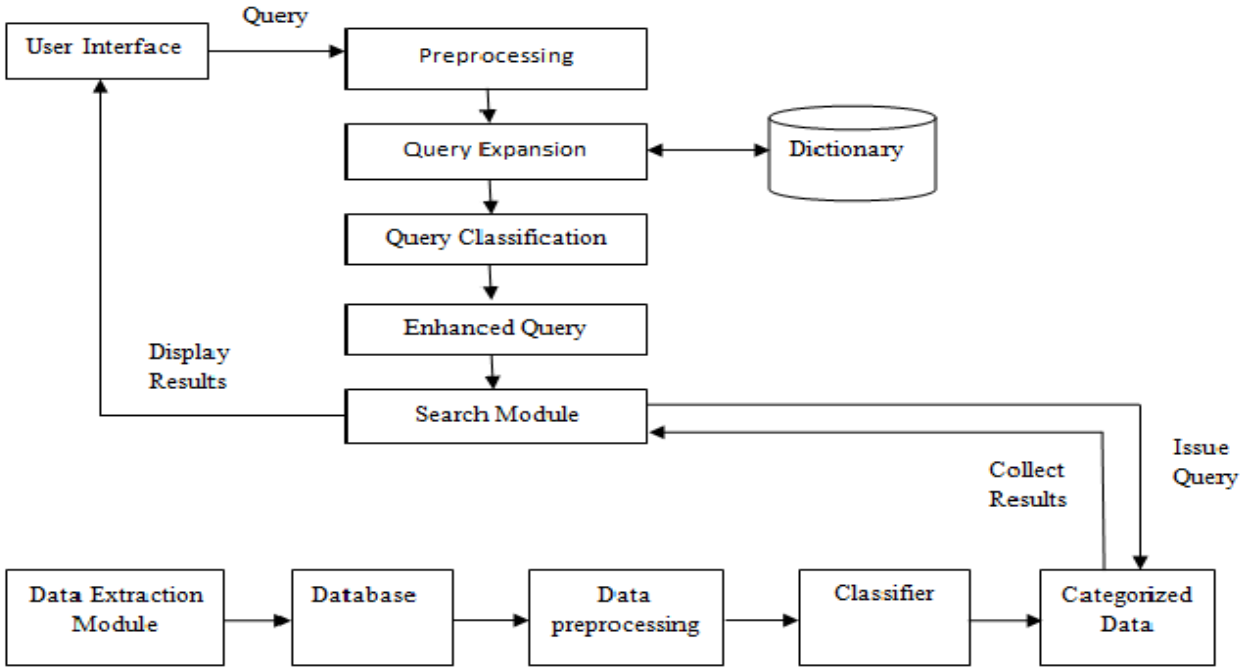


Fig. Detailed architecture

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 :

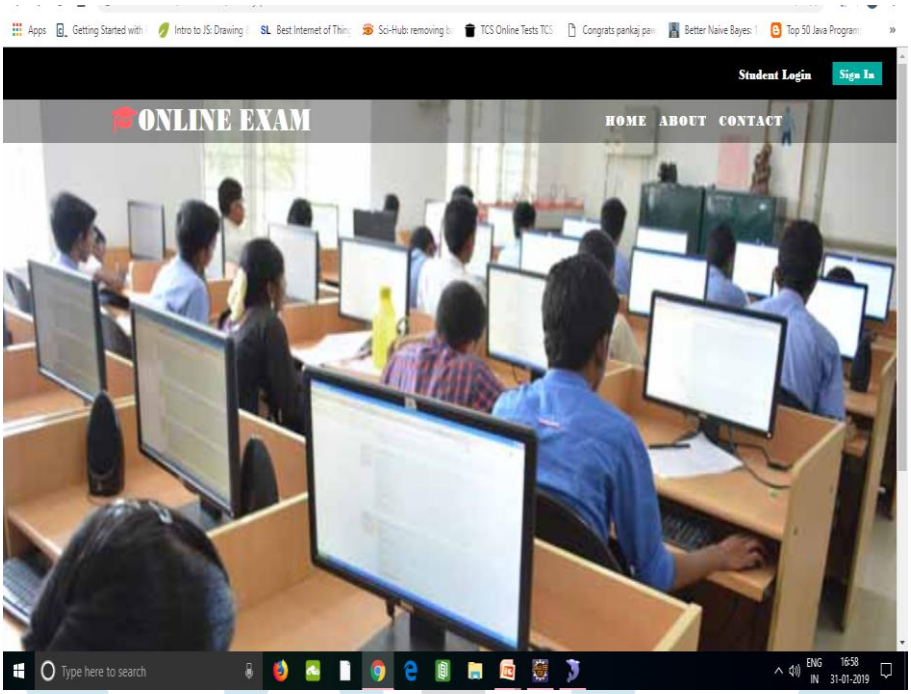


Figure 1:Student Login page

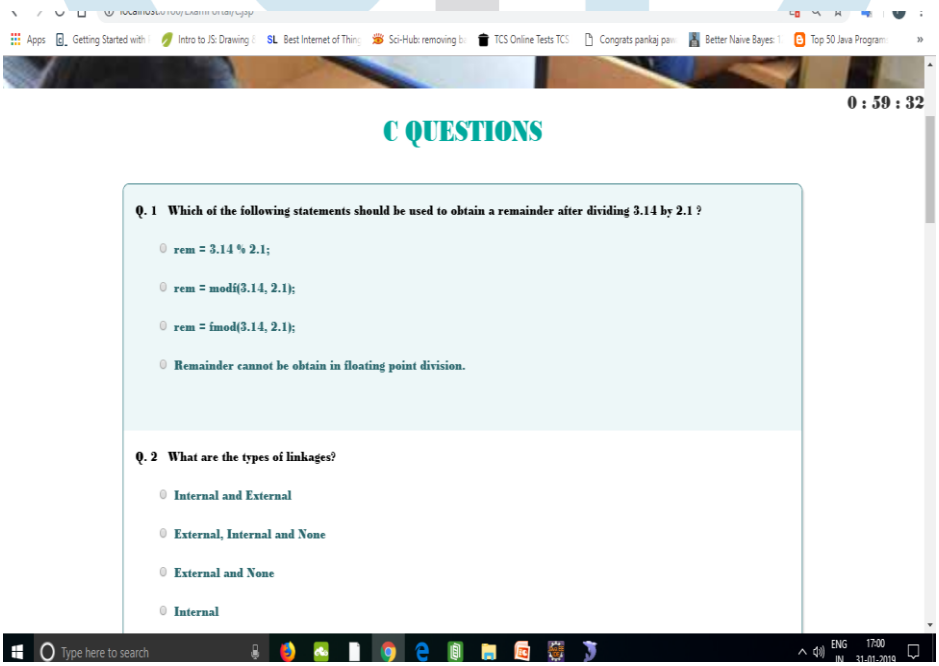


Figure 2: Mock Test

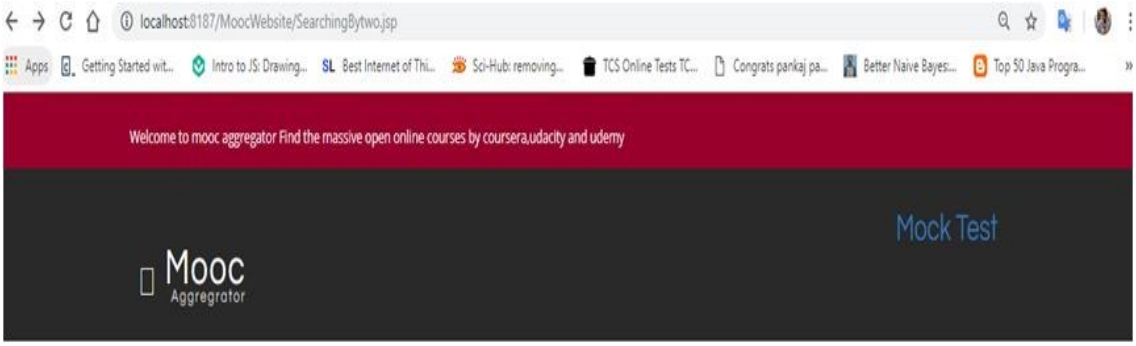
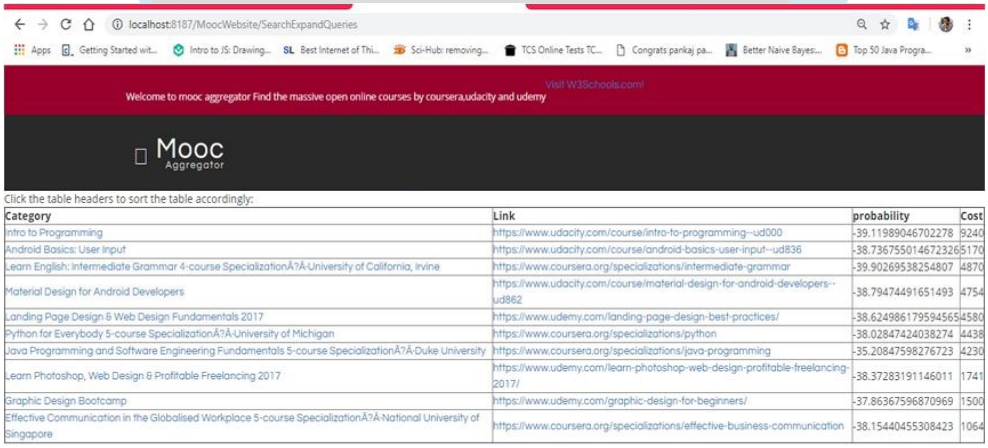


Figure 3: Home page



Category	Link	probability	Cost
Intro to Programming	https://www.udacity.com/course/intro-to-programming--ud000	-39.11989046702278	9240
Android Basics: User Input	https://www.udacity.com/course/android-basics-user-input--ud836	-38.736755014672326	5170
Learn English: Intermediate Grammar 4-course SpecializationÃ?Â·University of California, Irvine	https://www.coursera.org/specializations/intermediate-grammar	-39.90269538254807	4870
Material Design for Android Developers	https://www.udacity.com/course/material-design-for-android-developers--ud862	-38.79474491651493	4754
Landing Page Design & Web Design Fundamentals 2017	https://www.udemy.com/landing-page-design-best-practices/	-38.624986179594565	4580
Python for Everybody 5-course SpecializationÃ?Â·University of Michigan	https://www.coursera.org/specializations/python	-38.02847424038274	4438
Java Programming and Software Engineering Fundamentals 5-course SpecializationÃ?Â·Duke University	https://www.coursera.org/specializations/java-programming	-35.20847598276723	4230
Learn Photoshop, Web Design & Profitable Freelancing 2017	https://www.udemy.com/learn-photoshop-web-design-profitable-freelancing-2017/	-38.37283191146011	1741
Graphic Design Bootcamp	https://www.udemy.com/graphic-design-for-beginners/	-37.86367596870969	1500
Effective Communication in the Globalised Workplace 5-course SpecializationÃ?Â·National University of Singapore	https://www.coursera.org/specializations/effective-business-communication	-38.15440455308423	1064

Figure 4 : result

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. 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.
3. Query Classification: Task of assigning the query to one of the predefined categories based on content of the query. Search Module: Responsible for firing the query to the categorized courses dataset and collecting the results which will be displayed on the user interface.

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 task 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

[1] Aggregator Definition Wikipedia, https://en.wikipedia.org/wiki/Aggregator

[2] MOOCs Wikipedia, https://en.wikipedia.org/wiki/MOOCs Aggregator Definition Wikipedia, https://en.wikipedia.org/wiki/Aggregator

[3] Dhekne, Chinmay, and Srividya K. Bansal. ”Linking and Maintaining Quality of Data about MOOCs Using Semantic Computing.” Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.

[4] Saini, Chandni, and Vinay Arora. ”Information retrieval in web crawling: A survey.” Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on. IEEE, 2016.

[5] Pawar, Nisha, K. Rajeswari, and Aniruddha Joshi. ”Implementation of an efficient web crawler to search medicinal plants and relevant diseases.” Computing Communication Control and automation (ICCUBEA), 2016 International Conference on. IEEE, 2016.

[6] Lu, Meili, et al. ”Query expansion via wordnet for effective code search.” Software Analysis, Evolution and Reengineering (SANER), 2015 IEEE 22nd International Conference on. IEEE, 2015.

[7] (2017) The Udacity website. [Online]. Available: https://in.udacity.com/

[8] (2017) The Coursera website. [Online]. Available: https://www.coursera.org/

[9] (2017) The Udemy website. [Online]. Available: https://www.udemy.com/courses/ 22 Efficient Search over Information Aggregator

[10] Sarode, Shraddha, and Jayant Gadge. ”Hybrid dimensionality reduction approach for web page classification.” Communication, Information Computing Technology (ICCICT), International Conference on. IEEE, 2015.

[11] http://www.igniteengineers.com/mooc-advantages-and-disadvantages

[12] Sebastian Kagemann, Srividya Bansal, MOOCLink: Building and Utilizing Linked Data from Massive Open Online Courses, IEEE 9th International Conference on Semantic Computing, 2015.

[13] JDBC tutorial, https://www.tutorialspoint.com/jdbc/jdbc-create-database.html

[14] JSON website, http://json.org/

[15] Jsoup website, https://jsoup.org/

[16] Carpineto, Claudio, and Giovanni Romano. ”A survey of automatic query expansion in information retrieval.”ACM Computing Surveys (CSUR) 44.1,2012.