**MOOCRec – MOOCs Recommender Based on Learning Style**

Project Id: 18-036

Final Report (Draft)

Hasith M. G. S.

IT14140280

Bachelor of Science Special (honors) in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

**MOOCRec – MOOCs Recommender Based on Learning Style**

Project Id: 18-036

Hasith M. G. S.

IT14140280

Dissertation submitted in partial fulfillment of the requirements for the degree Bachelor of Science Special (Honors) Degree in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

**DECLARATION**

I hereby declare that the project work entitled “MOOCs Recommender Based on Learning Styles” (MOOC Scraping, Crawler and Monitoring and Schedule the process) submitted to the Sri Lanka Institute of Information Technology, is a record of original work done by our group under the guidance of Mr. Nuwan Kodagoda (Supervisor) and Ms. Kushnara Suriyawansa (Co- Supervisor), and this project work is submitted in the fulfillment for the award of the Bachelor of Science (Special Honors) in Information technology Specialization in Software Engineering. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma. The diagrams, research results and all other documented components were developed by myself and I have cited clearly any references I have made.

**Authors:**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Registration Number | Signature | Date |
| Hasith M. G. S. | IT14140280 |  |  |

The above candidates has carried out research for the B.Sc. dissertation under my supervision.

**Supervisor:**

|  |  |  |
| --- | --- | --- |
| Name | Signature | Date |
| Mr. Nuwan Kodagoda |  |  |

**ABSTRACT**

Over the years, Massive Open Online Courses (MOOCs) has evolved as a learning phenomenon providing large number of courses in different domains to a wide range of users. However, learners mostly explore through different platforms and resources before they find the material that best fits their needs. It is a known and proven fact that different learners have different learning styles. As of yet, no system has been implemented which takes the learner’s style into consideration and recommends the best available MOOCs to the learner. The proposal aims at describing processes of building a learner-style centric MOOCs recommender that takes his/her learning style as a valuable input. Different video lecture styles (like talking head, presentation slide with voice-over, animation etc.), available quizzes, resources etc. are considered as significant factors when mapping learning style to MOOCs. We will be focusing only on Computer Science Programming courses on edX, Coursera and Futurelearn platforms. Among various available models, Felder-Silverman Learning Style Model (FSLSM) is selected based on literature review, to identify the learning style of users. Apart from looking for preferred courses, the user can also use the system as an advanced search engine to find any course of choice. The main objective of the system is to present most suitable resources to a learner so that learning and mastering a subject becomes effortless.

# **ACKNOWLEDGEMENT**

The success of the work described in this document was done as our fourth-year research for the subject Comprehensive Design Analysis Project. This project is the result of all the dedicated work of the group members and the encouragement, support and guidance given by many others. Therefore, we would like to express our appreciation to all who gave us the support to complete this significant task.

Authors would like to declare their special thanks for their research supervisor, Mr. Nuwan Kodagoda for her continuous guidance, support and insightful comments which motivated us to move further. Authors would also like to thank their co-supervisor Ms. Kushnara Suriyawansa for constantly providing guidance and assistance to implement the research successfully.

Also, authors wish to thank all colleagues and friends for all their help, support, interest and valuable advices. Finally, we would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged us to make this a success. We will forever be thankful to my former lecturers for supporting us during these past four years.

# **INTRODUCTION**

* 1. **Background**

Massive Open Online Courses (MOOCs) have taken a giant leap in the field of e-Learning, specifically open education. Over the years, various platforms such as edX , Coursera , Udacity , Futurelearn etc. have emerged with the intent of providing massive educational resources to its users. One of the major and fundamental component in MOOCs is the video lectures and its production style. There are standard styles, like talking head, presentation slides with voice-over, animation, screencast etc. that are most commonly used. Along with the video lectures, quizzes and learning resources are also made available to help in the learning process.

In the recent years, there has been increasing attention towards the characteristics of learners such as learning styles. Different learners have the ability to learn in different ways and hence poses their own style of learning. Because of this behavior, the learner mostly explores through different MOOCs platform to find the most optimal learning resources that best fit their needs, preferences and learning style. Until now, no system has been developed that takes learner’s learning style into consideration when recommending the best available MOOCs across different platforms.

Hence, we will be focusing on two different aspects: learning styles and MOOCs. Our sole purpose throughout the research project is to integrate both the factors together and develop an educational platform where learners from any domain can benefit from personalized learning services.

* 1. **Literature Review**

MOOCs have evolved as big players in the field of online learning and its unique characteristics makes it an effective Technology-Enhanced Learning (TEL) model in the modern era of education and technology. Lots of prominent research have been carried out considering the state-of-art and describing an overview of MOOCs. According to the survey from Class Central, until 2017, around 81 million students are registered worldwide, where 23 million were new learners that signed up only in 2017. Similarly, the total number of MOOCs surpassed 9400 contributed from over 800 universities in 2017. This exponential growth rate is promising to believe that MOOCs are a new revolution that can help people distributed all over the world to gain access to diversified quality education in a more convenient way.

Video lectures are the significant component of MOOCs and researches have been keen in improving the effectiveness of these videos. However, limited research has been carried out and more work needs to be carried out. Guo et al. explores the relationship between video production style and student engagement. Videos were classified into 6 types of production styles. Another promising research by Hansch et al. presented pros and cons of different video styles. One of the most extensive work is presented by Reutemann which shows the comparison between different video styles in different platforms. Thorough statistics of video styles used in edX, Coursera, Futurelearn and Iversity is reported. The paper also raises an interesting argument to correlate the video styles with drop out rates.

Learning style refers to the way a learner receives and processes information. Therefore, different learners have different learning styles. Considering learners learning style when designing a course has been found effective and shown positive results. It has been stated that providing learners with learning materials and activities that suit their preferences and learning style makes learning easier for them. More evidence to support the statement is provided by studies which showed students can achieve better learning outcomes and higher scores, and can also master the course in less time.

Several studies have proposed to integrate learning style into the open learning environment (MOOCs), to provide adaptive and personalized support for learning. Other studies reveal the use of data mining and machine learning algorithms to automatically identify the learners’ learning styles. However, there has been no significant research to support the direct mapping of a specific learning style with MOOC based on characteristics such as, video lecture production style, available quizzes, resources etc.

Researches regarding MOOCs search engine and recommenders are also being carried out at a rapid rate. A recommender system using Case Based Reasoning (CBR) approach is proposed in. User’s query is described by five attributes where each attribute is assigned a weight value based on user’s preference. “Courducate” is another system proposed for personalized search engine with two functionalities: multi-site search and multi-filed search. Besides using BM25 ranking function, a novel ranking function is used to rank the sites upon query. A different approach is taken in where the authors proposes to associate MOOCs with learning outcomes. Hence, allowing learners to discover the most suitable MOOCs for their learning objectives. proposes two contributions: Using attribute and attribute value weight of resources to get specific user preferences; A new algorithm to overcome the shortcomings of the Collaborative Filtering (CF) and provide more accurate personalized recommendations on MOOCs. In the similar context, Content based, and collaborative filtering recommendation approaches are used to accommodate several undergraduate characteristics when recommending MOOCs.

Finally, we can conclude that, there has been no system which addresses our research problem and hence a system that recommends the most suitable resources to the learner based on their learning style while at the same time, acting as an advanced search engine for MOOCs needs to be implemented.

Extracting information from three different MOOC platforms: edX, Coursera, Futurelearn and implementing a crawler to periodically look for new data is another research topic in our system. A similar research is carried out in which designs a set of configurable news collection system based on web crawler, which can crawl news from target news website. It can crawl a variety of multi-source data and the crawler is customized highly. In addition, it can do corresponding processing to crawled news content in accordance with need. Another research on crawler implements an incremental web crawler on the basis of the crawler architecture named “Scrapy”. It crawls the information of news on the website successfully and implements incremental crawling. But, one disadvantage of the web crawler is that it is not a general web crawler and we need write different crawling rules for different websites. A hybrid approach is proposed in which employs the meta-search engines and a semantic-structure-based Web page analysis algorithm. Their design goal of the focused crawler is to collect web pages probably related to a specific topic as many as possible and reduce irrelevant pages as few as possible.

An intelligent web crawler is presented in making the rule settings dynamic, at the same time, TF-IDF method is used to calculate the Web document correlation, and the automatic acquisition of data extraction rules is realized, which reduces the development cost and maintenance cost and improves the development efficiency of the crawler.

* 1. **Research Problem**

With the rapid advancement of technology, learning resources available over the web are more distributed than ever. Specifically, MOOCs (Massive Open Online Courses) are a relatively recent learning phenomenon with different platforms available like, edX, Coursera, Udacity, Futurelearn, Khan Academy etc. which provides large number of courses to a massive audience in different domains. MOOCs has transformed the era of learning since its introduction and one of the main reasons behind it is openness. Learners coming from different background has access to massive resources at any point of time. It truly has brought a revolution in field of e-Learning.

As the popularity of MOOCs increases, more platforms are stepping up into the play where each of them has unique format of displaying the course information. This will bring much inconvenience to the users and there is a high demand of a unified platform where learners can explore the courses based on their preference. An advanced MOOC search engine that allows the users to filter through multiple fields and parameters unlike other current general search engine is a key necessity and of high demand.

The concept of learning styles has acquired a great attention and influence within the educational field over the recent years. There is a thriving industry devoted to provide the best learning resources to the learner that fits his/her learning style. Many studies have also quoted that learning and mastering a subject becomes more effortless when the right resources are provided. MOOCs search engines that are available currently, do not provide the facility to search for courses depending on their learning style. Most of the users are even unaware of their style of learning and providing them insights on it, will definitely help them in the journey of learning.

Considering the above facts, we can conclude that there is a high demand of a unified platform for recommending MOOCs based on the learning style of user and at the same time, serve as a search engine allowing the users to filter through more advanced fields.

# **OBJECTIVES**

## **Main objectives**

• Develop a unified platform for searching MOOCs distributed on various platforms: edX, Coursera and Futurelearn using advanced filters.

• Recommend the most suitable courses from different platforms based on learners’ learning style and domain knowledge.

• Develop user-friendly interfaces so that users can easily get what they need.

• Help users identify their learning style based on Felder and Silver Learning Style Model.

• Explore and understand the application of deep learning techniques in various scenarios.

# **METHODOLOGY**

Web scraping is a software program that extracts information form websites. These programs are able to simulate human web surfing behavior by implementing either low level Hypertext Transfer Protocol (HTTP) or embedding a web browser. Web scraping concentrates on transforming unstructured data found online, into structured data that can then be stored and analyzed in a database or internal storage. Our web crawler executes five different functions. First, we undergo the task of web interaction. This will happen when we detect a new course in our one of MOOC sites. Secondly, programs known as wrappers need support for generation and execution. This point we collect overall information about download content and size. Scheduling is the third function that allows wrappers to be applied repeatedly to their respective target pages. Next, data transformation occurs which involves the filtering, mapping, refining, and integrating of data from one or more sources. The result is then structured to create the desired output format. Finally, we download the content from course

First, We create a json object for one course. It include all resources we have to download and API endpoint for that resources. Before create the json object read the all content of the web pages using library call “beautifulsoup”. After create the json object we download courses according to json objects.

**3.1 Beautiful Soup**

Beautiful Soup is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping. It is available for Python 2.6+ and Python 3.

**3.2 Feasibility Study**

**3.3 Requirement Analysis and Specification**

**3.4 Design**

**3.6 Testing & Implementation**

**3.6.1 Implementation**

In this step, project design should have to implement by using a programming language. In our case mainly we are using python and anguler as the main programming. we are using MongoDB as a database and mLab fully managed cloud database service use to host the database.

**3.6.2 Testing**

After integrating all the components, we have to test the products. Since we are using github version control repository, we don’t have to sacrifice time to integrate the final project. Because it integrates by itself when a developer makes changes to the code and checking in. Therefore, we have to focus on testing part toughly. There are three types of testing.

**3.6.2.1 Unit Testing**

The unit testing is carried out by dividing the designing process into smaller parts known as units. Each and every individual unit should be tested under unit testing. When the system design documents are received, the work is divided into modules/units and actual coding is started. The system is first developed in small modules called units, which are integrated into the later phase. Each unit is developed and tested for its functionality. Unit testing mainly verifies if the modules/units meet their specifications. Since we developed our application in functionality wise, the unit test was also done as functionality wise.

**3.6.2.2 Integration Testing**

When comes to the integration testing, the individual modules are combined and tested. Objective of this testing is to make sure that the interaction of two or more components produces results that satisfy the functional requirement.

**3.6.2.3 System Testing**

Testing the whole system against the specifications is called as system testing. Once the unit and the integration testing is done this system testing is started to make sure the compatibility and the integration of the units are done properly. Waterfall model in testing can only be done by dividing coded program into various manageable units. After that in integration phase those units are integrated into complete system. After that we have to test to verify if every modules are coordinate with each other and the system as a whole behaves as per the specification.

**3.6.2.4 Functionality Testing**

All functions will test according to their requirement.

**3.6.2.5 Performance Testing**

This testing will give effective and efficiency of the product. Using this testing developer can identify where the defects are.

**3.6.2.6 Security and Portability Testing**

This testing is done when the product is run on different platforms as well as accessed by different users.

We had to analyze the project outcome with quantitative and qualitative measurements. Quantitative measurements will increase the performance of the system. To implement the high-performance system we had to have a proper testing plan. Here are the some of the test cases we created for system.

--- Test Cases

* 1. **Research Findings**

1. **Results and Discussion**

**4.1. Results**

**4.2 Discussion**

1. **Conclusion**
2. **References**
3. **Appendices**
   1. **ER Diagram**
   2. **Use Case Diagram**
   3. **Use Case Scenarios**
   4. **Frontend Interfaces**