**UStudy**

**A Complete Educational Portal**

**Arapally Ebenezer Anand**

**Bandaru Sai Kishore**

**Kommineni Sivakrishna**

**Ponnam Balakrishna**

**Summary:**

Our application is and educational portal which is used to search for colleges in USA for graduation by the international students. It is a Native Mobile App developed for Android Users. The colleges will be searched based on the user inputs like gre score, toefl score, and range of tuition fees, location and financial aid if required. In addition to the search results we also recommend colleges to the registered users, but not for the guest users. Recommendations are done based on the user profile. We have used machine learning techniques for clustering of data, which consists of college details and also recommendation techniques. Clustering algorithms were used to cluster the college details based on their profile and recommendation algorithms (content based recommendations) to recommend colleges to registered users.

**Framework Specification:**

The framework typically consists of three stages GUI, parser, and database. The database will deal with the required data analysis, data storage. The GUI part is being developed in Android. The parser includes the task of sending input to the database in such a way it can understand and analyze the information. And also parser is used to convert the return format of output from database to user format.

**System Architecture:**



Figure 1 System Architecture

The above diagram depicts the overview of our system architecture. Let’s follow the bottom to top approach for describing the system architecture. Initially we have a huge number of files which has different details about the colleges. So we have pruned all these files into single large file and stored it on the disk. Once we have the data we have applied the k-means algorithm technique on these data to cluster the data on the basis of “GRE”, “TOEFL”. For this we have R technology for the machine learning. The result i.e. the clustered data obtained after k-means clustering should be parsed in order to push the data into solr in JSON format. So we have designed a parser which can convert the data retrieved from the R is converted into JSON format and pushed into solr. Now we have the data in the Solr and this will be updated periodically whenever there are changes in the data.

Now coming to GUI part, each user can either login as a registered user or as a guest user. Once the user specifies his details i.e. the gre score, toefl score, range of tuition fees, location and financial aid, the respective matching data will be retrieved form the Solr. Solr pops data in the form of JSON so we have designed another parser in the middle to convert these JSON format data into a readable data and displayed to the user. For registered users we also recommend colleges based on his user profile, so whenever the user logins the recommendation algorithm will be run and a set of colleges are displayed. After the search the user will be displayed with further pruned list of colleges which are ebbing calculated form the specifications of the user.

**Domain Model:**

**Data Sources:**

Data is being collected from various government websites and is being collaged. We have got the data in the form csv files. After pruning of data we have maintained this data in the form of a single csv file which is being used in all machine learning techniques. The pruned csv file consists of college id, college name, gre, toefl, address, url, tuition fees etc. Below are the websites from which we have collected the data.

* [**http://nces.ed.gov/ipeds/datacenter/Default.aspx**](http://nces.ed.gov/ipeds/datacenter/Default.aspx)
* [**https://www.data.gov/education/**](https://www.data.gov/education/)

**Methodologies and Algorithms:**

The algorithms which we have used in our application are Clustering and Recommendation algorithm. For clustering we have used k-means clustering and for recommendation we have used the technique of content based recommendation which uses the user profiles or considers a single user characteristics for recommendations. For k-means clustering we need to give certain dimensions for clustering the data. Here we are clustering the data into 4 divisions i.e. very high, high, medium, low profile colleges based on two dimensions i.e. GRE, TOEFL. Once we have the clustered data we designed another algorithm which converts the clustered data into meaningful data and further convert that into JSON. Once we have JSON data we will be pushing the data into Solr using the curl commands. So this process of clustering goes on periodically so as to maintain the database up to date.

Once we have data clustered, the next algorithm we will be using is recommendation algorithm. Our recommendation algorithm typically works on the user profile. Based on user profile we recommend colleges to user. In the user profile we consider his marital status, gender, locality as parameters and suggest users the colleges. Consider if the user is from Asia, then we shall check which has maximum number of Asian users and recommend him/her those respective colleges.

**Analytic Tools:**

For data analysis we used R tool. We have used as an interface for clustering data base don our profiling dimensions. The steps of using R are mentioned below:

* Insert the data into R
* Specify the dimensions
* Combine the imported data and the dimensions
* Apply k-means algorithm on the combined data
* Display the data using table

**Analytical Tasks:**

The analytical task basically we designed include only the recommendation algorithms as we have done the clustering using R tool. The recommendation analysis goes with the user profile basis and the clustered data. Once we find a pattern which matches the sue profile we shall recommend such colleges to the user.

**Application Specifications:**

**Software Specifications:**

* Tools: R, Solr, Android Development Kit.
* Operating System: Android
* Development Operating System: Windows 8
* Programming Language: Java 7.0
* Databases: MySQL, HBase

**Class Diagram:**



Figure 2JSON Parser

The above class diagram depicts the parsing of the data retrieved from Solr.



Figure 3 Clustered Data parsing

The above class diagram depicts the parsing of Clustered data retrieved from R. Here data read represents the data to be read from clustered data and data prune does the refining of data. JSON Conversion here converts the clustered data into JSON format.



Figure 4 User Login Interface

In the above diagram we define actvities performed at user end. In class Registered User, we have recommend(string) function which does the recommendation algorithm.

Figure 5Sequence Diagram

The above diagram depicts the sequence diagram of our application.

**Activity Diagram:**



Figure 6Activity Diagram

The services majorly consists of assisting the students in choosing right colleges. So for a guest user the services provided are less. Based on simple search criteria the colleges will be projected to him. But if a user gets registered then there will be colleges recommended to him based on his previous search, user profile. The colleges list will classified prior in the database

**Design of Mobile Interface:**

The mobile interface primarily consists of a UI page where we need to specify the search specifications. The other designs include the design of Sign Up Page, result retrieval page. The Guest Page and User registered page differs with one section i.e. the recommendations. We have completely designed the user interface as a mobile app for the Android Users.

**Implementation:**

**Implementation of data model and algorithms:**

The implementation of the algorithm is basically done in Java. We have used various number of classes in Android for the recommendation algorithm and also we have pruned the data which we have retrieved in such a way that we get only certain number of columns which are only required. Once we have data clustered, the next algorithm we will be using is recommendation algorithm. Our recommendation algorithm typically works on the user profile. Based on user profile we recommend colleges to user. In the user profile we consider his marital status, gender, locality as parameters and suggest users the colleges. Consider if the user is from Asia, then we shall check which has maximum number of Asian users and recommend him/her those respective colleges.

The below diagram shows the sequence of command that are used for R technology

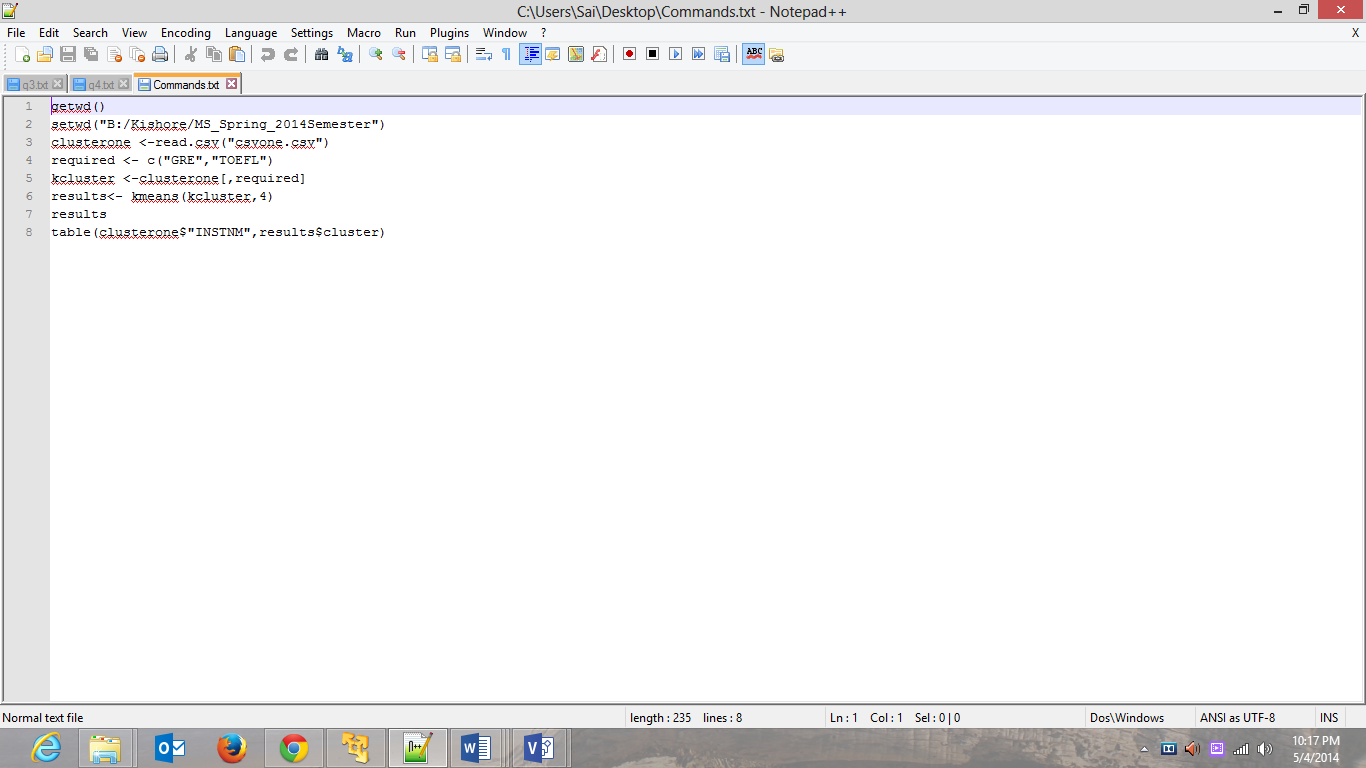


Figure 7R commands

**Implementation of services:**

For the web services we have directly used the solr connectivity for storing the data and we will be retrieving data from the solr. We haven’t created any glassfish or tomcat server as our requirement has been satisfied b usage of Solr.

**Implementation of User Interface:**

For the user interface we have used the Android apps and designed the application. The user interface that we implemented include are retrieval page, search page, login page, recommendations page. We have used the Linear Layout for building the GUI interfaces. For stroing the suer profile we have sued MySQL database.

**Documentation:**

Let’s see the output of after executing the k means algorithm on the data.

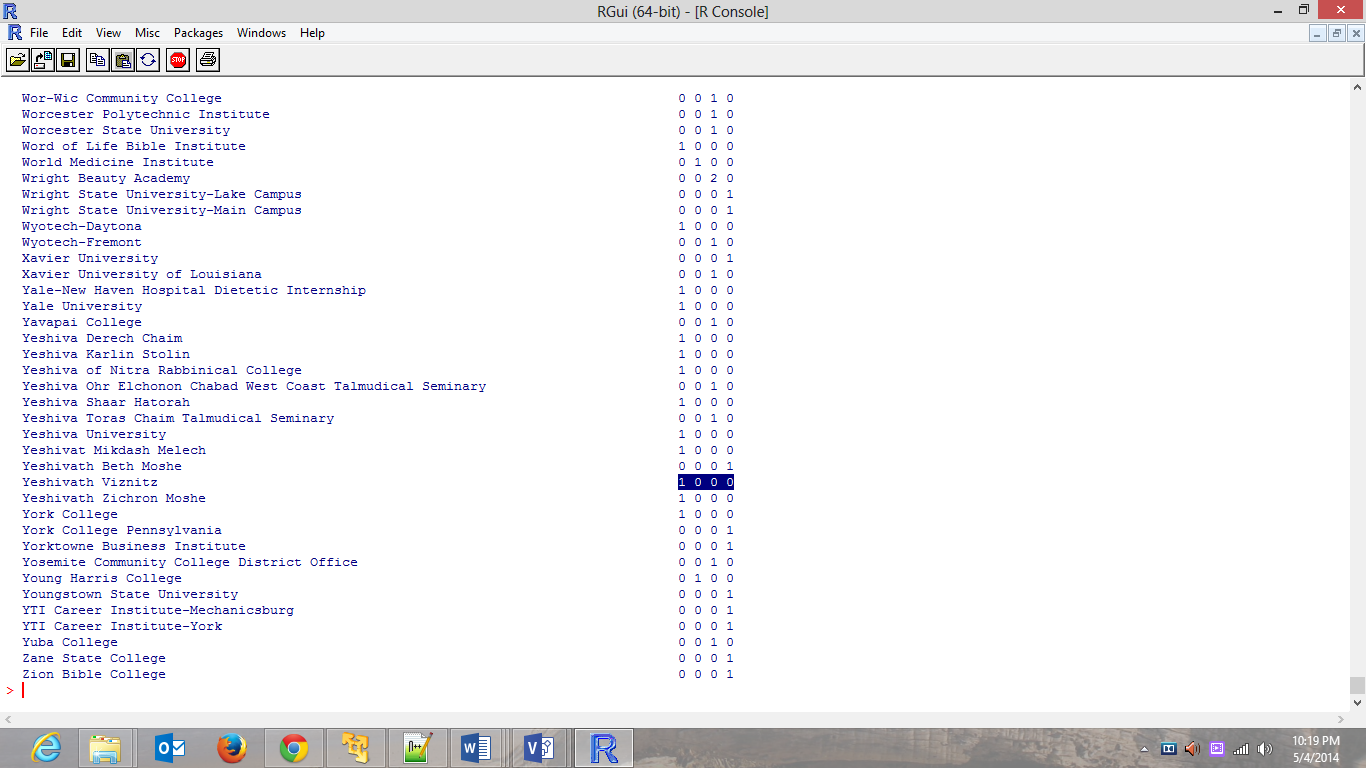


Figure 8Result of k means clustering

Here each column represents very high profile, high profile, medium, low profile colleges respectively. So each college will be profiled which forms as the end result.

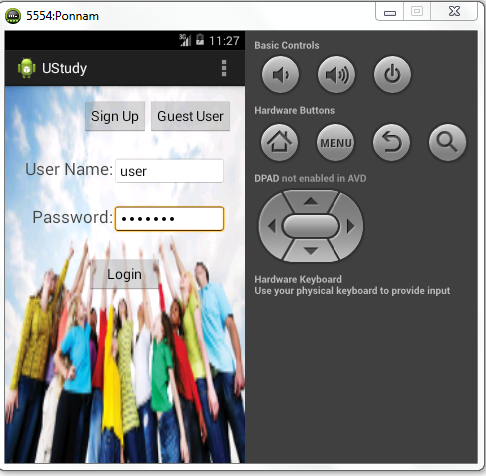


Figure 9Login

The above diagram depicts the login page view. The user can either rlogin as guest or registered user.

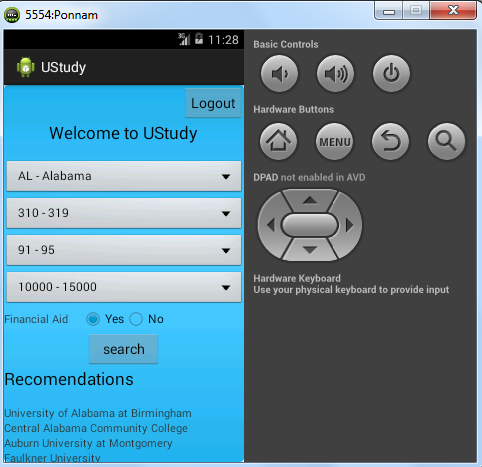


Figure 10User specification

The above diagram depicts the view page where we will specify the criteria and as he is a registered we can see recommendations at the bottom.

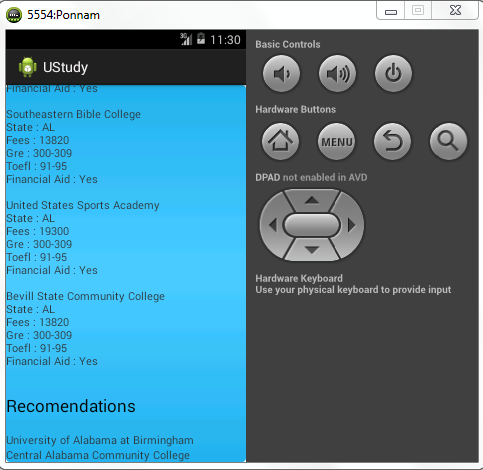


Figure 11Results

The above diagram depicts the end results and recommendations after clicking search button in prior page

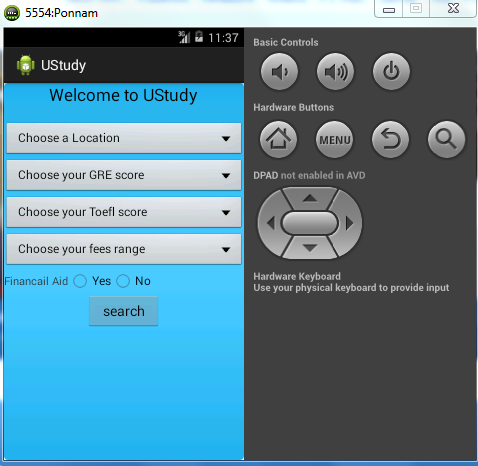


Figure 12Guest user search page

The above diagram shows the guest user page which doesn’t have recommendation’s.

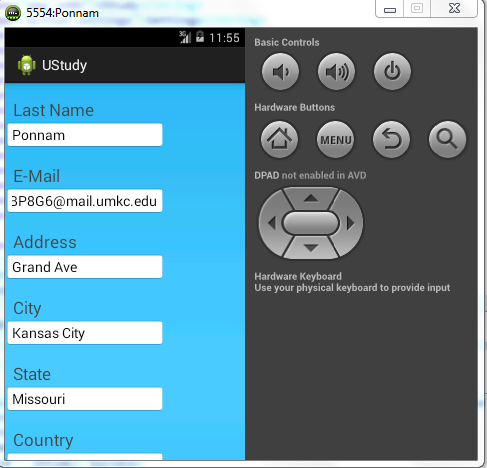


Figure 13Sign Up Page

The above diagram shows the sign up page.

**Project Management:**

All the tasks and their day today increments are being updated in the scrumdo.

* <https://www.scrumdo.com/organization/university-of-missouri-kansas-city5/dashboard>

The work has been equally distributed between the four members of the data with two members dealing with back end and two members dealing with GUI design in each successive increment.

All the stories with time allocations are updated in the scrum do for all the increments.

**Deployment:**

The code and the respective data files are updated to the GitHub in the following link:

* <https://github.com/SaiKishoreBandaru/UStudy_FinalProject>

And the YouTube link for video presentation is given below:

* <https://www.youtube.com/watch?v=cyT7NctLaBs&feature=youtu.be>