**Project Plan**

**Introduction:**

The increase in globalization and wide spread of technology resulted in the deluge of data. To acquire desired information is always tough from the ocean of World Wide Web. Consider a scenario where a student needs to search for college in United States America as per his requirement say location based, gender based, course based etc. Even though if he/she can extract data it is tough to choose one from the huge list of colleges and compare on priority. To resolve such situations we have designed an application which will provide the list of colleges on the basis of the user priority. In addition to this the user can also have a view of his career i.e. the academic progress through the application.

**Goals and Objective:**

The application is basically designed for an easy search of colleges through the inputs given by user. The user can either be registered or a guest. The registered user needs to specify his profile based on his profile several recommendations can be made to him. The initial inputs given by user will be location, fees, scholarship and based on this inputs we will display a range of colleges to the user. The user can select the universities and can compare the colleges based on his priority like jobs, location etc. Based on the profile of the user we will be directly suggesting certain colleges without the specifications of user. Certain advertisements will also be displayed based on the user profile such as location, gender etc. The user change his priorities whenever he wants, but by default it will be the history of the user search that will help in displaying the college lists. The goal of the application is to give a clear view of colleges and the courses they offer. Comparisons, analysis of priorities will be done with accuracy and the search is displayed. Career growth of the user can also be maintained based on his GPA s in each semester or year. There will be also a calculator provide to convert his local percentages i.e. his academic evaluations into the GPAs, so that the search can be more precise. With less number of information more data can be provide through recommendation mechanism, search mechanisms and advertising mechanisms.

**Proposed System:**

The system typically consists of a login page, the user can be a guest or a registered user. The validation will be made initially and then the list of colleges will be displayed based on the input specifications. The history of user will be tracked and stored so that the further visits to the page will display some recommendations to the user based on his past searches and priorities. Like while comparing the colleges the user will be displayed with certain priorities so that we can analyze on this input vectors and display the comparison analysis. The user will also be mailed with list of colleges on request. The registered user can also navigate to the college websites and process his applications. This features will not be provided for the guest user such as navigation to college websites, storing of priorities, recommendations. In addition to this the registered user has the facility to track his academic growth and calculate his GPAs. He advantages of registration is based on his locality, colleges having higher percentage with the people from his location are recommended on prior i.e. before the user selects the colleges.

The application is a native app and developed using Android. So all android phones can use this application. The recommendations will be only accurate only if the user profile is accurate otherwise the matches will be completely inaccurate. The response time would be fast from the second time of the visit as there will be recommendations to the user and previous meta data will be stored. The application features can be extensible to the entire world i.e. all colleges in the world and also comparison of students profile within groups. The user will also will be provided with a proper help documentation.

The application will support android version from Froy till JellyBean 4.3.x.

The following diagram depicts the UML Activity diagram:



Figure 1UML Activity Diagram

**System Architecture:**

The brief description of the application system architecture is described below:



Figure 2System Architecture

The primary components of the system architecture includes GUI, Data Analysis, Distributed Storage data and REST services. Since we are using a huge amounts of structured it is almost impossible to use the relational databases. So in order to serve this purpose we are using the Hadoop database. Here we have collected from various data websites and will be hosting this data onto the Hadoop. Using Mahout we will be using the predefined algorithms to classify the data in such a way that we can easily form certain relations. Now these classification helps in identifying the institutions on user input basis. Now we will also be using certain Recommendation systems in order to recommend some institutions to the user based on his previous searches. However based on the output from Mahout we will be sending this formulated output to Solr. And using Solr we will be sending the output the user interface using REST services. The REST services will return the data JSON format, and the parsing of JSON needs to be done before displaying the results. We are also planning to include Lucene which helps in retrieving data from Wikipedia which can much more useful for the data classification and recommendation.

**Software Specification:**

Tools: Mahout, Solr, Android Development Kit.

Operating System: Android

Development Operating System: Windows 8

Programming Language: Java 7.0

Databases: Oracle, Hadoop

**Data Sources:**

1. <https://explore.data.gov/Education/Post-Secondary-Universe-Survey-2010-Directory-info/uc4u-xdrd>
2. <http://nces.ed.gov/ipeds/datacenter/Default.aspx>
3. <http://dumps.wikimedia.org/enwiki/20130403/>

**Class Diagrams:**

For an initial outlook, the following class diagram depicts the brief overview of classes that we will be using:



Figure 3Class Diagram

As of now we designed five major classes. The user login interface will be inherited by the registered user and guest user, which includes user validation and fetching of important data from data base. Then there will be a data analysis class which includes he data analysis part and also has the data classification and recommendations. The data output will be send through REST services using Solr from Hadoop. Since the REST services in some format here we consider JSON format so a class will be used to parse the JSON output to user readable format.

**Sequence Diagrams:**



Figure 4Sequence Diagram

The sequence of operations includes the user validation in case of a registered user other wise a guest need not login. Then we shall take the inputs from user. Based on the inputs we will be analyzing the data that has already been set through the Mahout and the result will be sent back to the user using REST services. Then based on user interest there will comparison of institution, recommendation of institutions etc. There will be inclusion of many other aspects from time to time in the application for future enhancements.

**Task Planning:**

For the entire project we will be following Agile model and the code release will be done in four iterations. In brief each iteration consists of the following tasks:

1. Designing the GUI, and hosting the data onto the Hadoop database.
2. Performing data analysis tasks which includes data classification.
3. Performing recommendation related tasks and parsing of the JSON output retrieved using REST services
4. Testing the entire application with different sets of data and performing bug fixing if found any.

The entire process and task split has been clearly mentioned is the ScrumDo tool and following is the link for the ScrumDo process for our project:

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

There will be an initial testing done on the data once it is being hosted on Hadoop so as to validate the data analysis.

**People:**

* Arapally, Ebenezer Anand ( [eakd6@mail.umkc.edu](mailto:eakd6@mail.umkc.edu))
* Bandaru, Sai Kishore ( [sbqm9@mail.umkc.edu](mailto:sbqm9@mail.umkc.edu))
* Kommineni, Siva Krishna ( [sk7x9@mail.umkc.edu](mailto:sk7x9@mail.umkc.edu))
* Ponnam, BalaKrishna ( [bp8g6@mail.umkc.edu](mailto:bp8g6@mail.umkc.edu))

**Bibliography:**

1. Hadoop Projects -- <http://hadoop.apache.org/docs/current/>
2. Developing Android Applications --<http://developer.android.com/training/basics/firstapp/index.html>
3. Microsoft research Projects -- <http://research.microsoft.com/en-us/projects/imanexpert/>
4. Developing and testing recommendation Systems: <http://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf>
5. Mahout algorithms : <https://cwiki.apache.org/confluence/display/MAHOUT/Algorithms>