CMPE 272 PROJECT REPORT – FUND ANALYSIS OF UNIVERSITIES IN CALIFORNIA

TEAM 9
Anshul Sharma (anshul.sharma@sjsu.edu)
Deepna Bains (deepna.bains@sjsu.edu)
Harleen Kaur (harleen.kaur@sjsu.edu)
Navneet Dhillon (navneet.dhillon@sjsu.edu)
Swetha Rathna kempahanumaiah (swetha.rathnakempahanumaiah@sjsu.edu)

Software Engineering Department, College of Engineering
San Jose State University, San Jose, CA 95192

ABSTRACT

Looking into the trending technologies and rising competition among various resources in the market, it has now become essential to do a comparison in order to make a best choice for oneself. The university research analysis helps in deciding among various universities in California so that an individual can pursue the best career in education as desired. The main goal of the project is to provide information to individuals regarding the analysis of various universities and chooses the best amongst them. The university research analysis implements the recent technology and provides information for universities across California.

INTRODUCTION

Scenario: There are many top universities in California, which attract students from all over the world like UCLA, UC Santa Barbara. Based various achievements of the universities, individuals wishing to pursue good higher education in CA can compare the universities so that they are able to make a best decision out of it.

Problem: There are many universities in CA, which offer a versatile education and program. For an individual who wants to pursue higher education expects that he/she choose the best university, which also meet his/her requirements. There are various websites, which give a lot of knowledge of comparison of one university with the other. But the comparison lacks analysis i.e. representation of information in some form of charts which gives a better insight on selecting the correct university. The University research analysis will help in accomplishing this.

Solution: The aim of our project is to present the information of various universities of CA in the form of analysis so that it can help an individual make a correct choice of which university to choose. The analysis will be done based on the awards, programs, funding etc. which determines the performance of a university. This will be presented in the form of charts and graphs. The university research analysis focuses on the funding received by various universities to present the charts.

METHODOLOGY

In order to do the analysis for the universities of California, the most important point was what methodology we must follow so that we arrive at the results we expect for. For this a methodology was planned that was needed to be followed .A brief description of methodology is described below.

For proceeding the biggest challenge was to obtain data in large number so that we could perform adequate analysis. In order to do analysis of data from various universities, a data set showing the information of universities was needed. The required data set was obtained from website **data.gov**. The data obtained was in unstructured format. In order to structure the data, Mongo DB and map reduce are used.

Mongo DB is an open source, documented oriented database i.e. it a no SQL database, providing high performance, availability and scalability. Instead of using rows and tables as in relational databases, the collection and databases form the architecture of the Mongo DB. The basic unit is key value pairs, which has a similar presentation to JSON like documents. The support of dynamic schema design by Mongo DB allows documents in a collection to have different fields and structure. In order to work with Mongo DB, it was downloaded http://www.mongodb.org/downloads. After creating DB, we can use it from the command line or the web console.

In order to fetch the meaningful data needed for our analysis, we implemented the map reduce function using the Mongo DB. In order to obtain the meaningful and the aggregated data, the **map Reduce** database command is used. The map phase is applied to each input document and the reduce phase is applied to keys having multiple values. The output is stored in a collection.

In order to present the analysis among various universities, graphs are implemented. The Google charts and the Am charts have been used. The Google charts is a tool from Google that allows creation of charts and embed them in web pages, they

provide a good degree of dynamism. These charts are supported by variety of programming languages like java,C#.net, python, PHP etc. Am charts are another type of charts that were used to display the analysis of two universities, for single university, for a particular city etc. in the form of graphs. Am charts are the standalone and independent library that any data visualization need and does not require any third party software to be installed. The am charts provides various types like bar graphs, line charts, area charts .pie charts and many more. The ones implemented in our project are the following:

- Bar graphs
- Pie charts
- Trend line

MOTIVATION

The main idea behind doing the university research analysis is to provide a website to the end users that provide useful information to the users regarding the comparison of the universities of California based upon the funding received by the universities from agencies like NASA and NSF. It should help the users, agency and students to make a correct decision on selecting the correct university for them. For this, Graphs that provide better insight in understanding the program of the university have been implemented. The main idea behind these how different users will use this.

As a student, Especially in US, student would be interested in looking for universities who support their interest on certain program. He can looks at this analysis report and find the universities which has professors who gets enough funding in this field. This might interest him since he'll be convinced that he gets enough support to pursue his area of interest. The comparison analysis done in this university research analysis in the form of bar graphs will help in the decision process. An example of the bar graphs has been presented on the analysis section of the report.

As an agency (NASA/NSF):
Agency can check how much overall funding they have made on certain research areas so that they can make a decision on how much, more funds they can provide and to which areas they should more focus on. From the perspective of a university, Universities can look into this analysis report of other universities and decide if they need to start new departments, which may cater to student's research area.

IMPLEMENTATION

In order to understand the university research analysis and the various comparisons done, it is necessary to understand the workflow and how various technologies have been implemented. The following is the list of technologies used in implementation:

- MongoDB
- Mongohq
- Bootstrap
- Github
- Codeship
- Heroku
- Coffeecupsoftware
- Nodejs
- Mapreduce

The above-mentioned technologies were implemented in the following manner:

1. Data was obtained from the website http://www.data.gov/
2. As the data was in unstructured format we refined it using CSV and hosted it on the Mongohq. Mongo Hq is a cloud-based platform for deploying and scaling Mongo databases
3. Using the mapReduce database command the data was mapped and phased in Mongo DB.
4. The UI was developed using node.js and the coffee cup software. NodeJS is a platform that is built on chrome's JavaScript and using the event based models easy and scalable applications can be developed.
5. Finally the charts were created using am charts and embedded into the web pages.
6. In order to deploy, we needed cloud services. After exploring many options, we discovered the heroku and the code ship technologies.

Heroku is a cloud platform as a service, which provides a new way of deploying web applications over the cloud. Heroku, in combination with Codeship was used to upload and deploy the code of the application designed on the cloud.

Codeship is installed as an add on to Heroku which helps inn easy, integrated and continuous deployment of application. Each time the code is pushed on Github, Codeship automatically deploys it on Heroku, hence making the deployment process much easier.

7. Keeping in mind that everyone works in sink, all the team members created their accounts on Github.

In order to check the website click the following links
http://caluniversityanalysis.herokuapp.com/
and the github link is as follows.
https://github.com/iamharleen/CMPE272Project

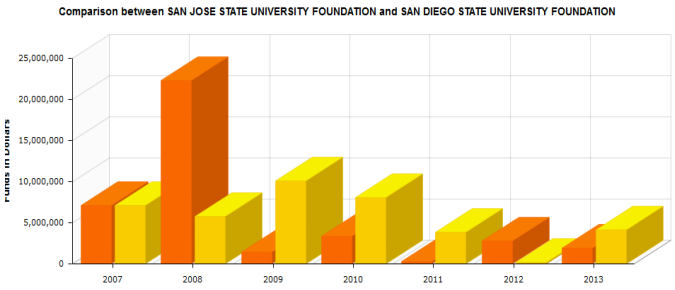
ANALYSIS
The following type of analysis is done in the application:

1. Comparison of 2 universities in terms of funding.

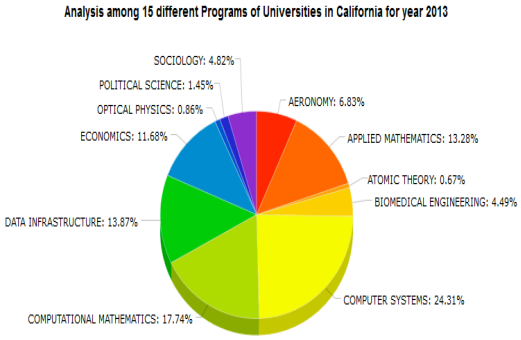
2. Funds analysis based on the cities of California
3. Share of funds provided by the agencies over the years
4. Analysis of different programs for universities of California
5. Funds received by major universities over the years
6. Comparison between funding provided by NASA and NSF

Some of the sample analysis graphs are as under:

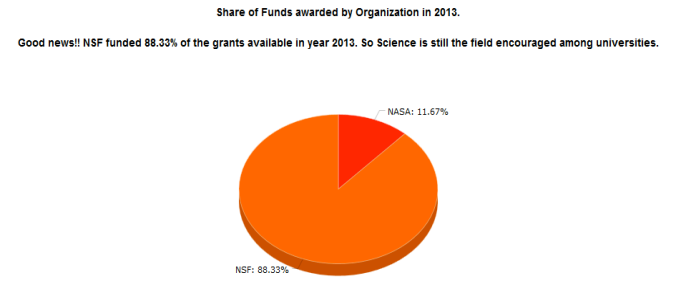
The graph for analysis based on the comparison:



The graph for analysis among 15 different programs for universities in California:



The graph for analysis of funding of two universities



CHALLENGES

Building a robust and stable application is a big challenge in itself. Moreover building an application like University analysis for California had its own challenges. Adequate and good resources were needed in order to present the analysis in a good and user-friendly way.

The biggest challenge that was encountered was obtaining a correct data set, which met the requirements. After extensive research, the data that was needed was obtained from data.gov.Data.gov is a U.S Government site that maintains large data for numerous fields. After the data set was obtained, we found that the data set was in unrefined form.in order to refine it we used the map reduce and brought it into concrete form

The second major challenge that was faced was choosing the adequate cloud platform that could easily sink with our applications. As Mongo DB is not compatible with each and every cloud platform, the process of figuring out the right cloud platform was time consuming

Another difficulty faced during the development of the application is not figure out the correct API for making analysis charts. As dynamism was our requirement for making the charts, it was hard to find tools that supported dynamic charts and to decide which one to use amongst them.

FUTURE IMPLEMENTATIONS

The current university analysis is deployed only for the universities of California. We plan it to extend it to other states of United States as well, so that the students wishing to pursue their higher studies in universities other than California also get the benefit of analysis being done by our application. We have the data sets for other states as well and plan to do the implementation using the same methodology mentioned above; we may include more analysis, if possible, depending on the requirements and time

CONCLUSIONS

During the development a of university research analysis, we have learnt and come across many new technologies like the Heroku and the Codeship.Nodejs and the Coffeecup template help in developing an easy and scalable UI. All the technologies used help in developing a stable and robust application, the main aim of this university analysis is to help out various users in making the correct decisions in terms of selecting a university and compare the universities in context of funds.

REFERENCES

Amcharts -http://www.amcharts.com/

Mongo HQ -http://www.mongohq.com/

Heroku- https://www.heroku.com/

Codeship -https://devcenter.heroku.com/articles/codeship