A system for measuring and ranking an overall performance of college by using online data

Aniket Nikam[1], Suraj Upare[2], Ujjawal Anand[3], Vikki Kumar[4]

[1,2,3,4] BE, Computer Engineering, MIT Academy of Engineering. Pune, India

Abstract—As a nation, we have to make college more accessible and affordable and ensure that all student graduate with an education of real value. Our students deserve to know, before they enroll, that the college they have chosen will deliver this value. we can develop a useful rating system that will help more student's to realize the dream of a degree that unleashes their potential and opens doors to a better life . It's an important step in improving transparency ,accountability and equity in higher education. College rating system will provide the analysis rating done on the basis of reviews, feedback and comments given by students and parents in various websites and social media. This system is implemented using web crawler (to extract information from various websites and social media) and sentiment analysis is done on the extracted on generate report.

A WebCrawler is an Internet bot which systematically browses the world wide web, typically for the purpose of Web Indexing, web crawler also clusters and gathers information.

It is necessary to hit the website and retrieve the data from that website. We are not going to take random data, the only data which belongs to the key factor, that we have to take and store it into the database. This phase is known as "Information Gathering". After this phase, the main phase of our project begins which is nothing but the "Data Analysis". In this phase dynamic data is taken as a input and analysis is done. At the end we are going to generate the graph for both positive and negative points.

Keyword: Web crawler, Ranking algorithm, semantic analysis, analysis methods.

I. INTRODUCTION

A rating is the evaluation or assessment of something, in terms of quality (as with a critic rating a novel), quantity (as with an athlete being rated by his or her statistics), or some combination of both.

A system is a set of interacting or interdependent components forming a complex/intricate whole. Every system is delineated by its spatial and temporal boundaries, surrounded and inuenced by its environment, described by its structure and purpose A System for measuring and ranking an overall performance of college by using online data evaluates and rank a college based on its performance and principles, and reviews or feedbacks given by the students.

1 Web Crawler

A Web crawler is an Internet bot which systematically browses the World Wide Web, typically for the purpose of Web indexing. A Web crawler may also be called a Web spider, an ant, an automatic indexer, or (in the FOAF software context) a Web scutter.

Web search engines and some other sites use Web crawling or spidering software to update their web content or indexes of others sites' web content. Web crawlers can copy all the pages they visit for later processing by a search engine which indexes the downloaded pages so the users can search much more efficiently.

2 Ranking algorithm

Algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

3 Semantic analysis

In linguistics, semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings. It also involves removing features specific to particular linguistic and cultural contexts, to the extent that such a project is possible. The elements of idiom and figurative speech, being cultural, are often also converted into relatively invariant meanings in semantic analysis.

4 Synthetic analysis

The analytic synthetic distinction (also called the analytic synthetic dichotomy) is a conceptual distinction, used primarily in philosophy to distinguish propositions (in particular, statements that are affirmative subject predicate judgments) into two types: analytic propositions and synthetic propositions.

Analytic propositions are true by virtue of their meaning, while synthetic propositions are true by how their meaning relates to the world.[1] However, philosophers have used the terms in very different ways. Furthermore, philosophers have debated whether there is a legitimate distinction.

II.   LITERATURE SURVEY

According to the college board(2013), over the last 30 years, average tuition and fees at public and private 4 years institutions rose 231 percentage and 153 percentage respectively. Public two years tuition and fees rose 164 percentage during the same period. Students have borne the brunt of rising costs, contributing to trends in student borrowing that show a full 60 percentage of 4 years college and universities graduate owing dollar 26,500 on average.

The aim of project is to generate the awareness among the students about the college, to provide proper guidance in selection of better college.

In 2009, N. Narayanan et al., proposed a method to rank to the professional colleges. They used "Fuzzy Repertory Table(FRT)" technique to acquire the necessary domain knowledge of colleges from which the ranking is made. A sample of four colleges is taken into consideration and accessing similarity is made over it. Distinction matrix is created based on similarities of colleges from which the ranking has been made.

Inspired from the above idea we are designing a system which gives a platform in which students or parents can check rank of the college, which is purely given by the true reviews.

There are various approaches to infer the patterns from big data using aggregation, filtering and tagging. Earlier research shows that data aggregation concerns about gathered data and how efficiently it can be utilized. It is understandable that at the time of data gathering one does not care much about whether the gathered data will be useful or not. Hence, filtering and tagging of the data are the crucial steps in collecting the relevant data to fulfill the need.

It is a big challenge to guarantee the quality of discovered relevance features in text documents for describing user preferences because of large scale terms and data patterns. Most existing popular text mining and classification methods have adopted term-based approaches.

III. PROPOSED SYSTEM

The higher education conversation has shifted from simply ensuring access to one that focuses on success ? supporting students through completion and readiness for careers, citizenship and life. This system will help in improving transparency, accountability and equity in higher education. We are making a system that will generate report of a college which includes college performance, principles, reviews or feedbacks given by the students and rank that college on the basis of these parameters.

IV. DETAILED DESIGN

There are 4 important phases in our project :-

1. Hit the website
2. Grab the information and store it into the local database
3. Apply semantic analysis and ranking algorithms
4. Generate Report

1. Hit the website :-

Enter the college name (keyword for our project).
The algorithm will display the multiple links of websites. Hit the various website and retrieve data from it.

2. Grab the information and store it into the local database:-

Copy data from the website and store it into the local
database. The data should not contain redundancy, means each site must be visited only once.

3. Apply semantic analysis and ranking algorithms :-

After storing the data into the local database we need to filter it and analyse it.The semantic and ranking algorithms are used for the analysis of data.

The analysis of data is based on various factors. The factor is "key-value" pair. Lets take example, if there is a comment like, **"Canteen is good",** then 'canteen' is considered as a key and 'good' is a value. so it will increase the counter of positive points.

If there is a comment like, **"Canteen is not good",** then 'canteen' is considered as a key and 'not good' is a value. so it will increase the counter of negative points.

4. Generate Report :-

The final phase will produce the report of a particular college. The report contain the graph of positive point and negative point.
Based on the report the grade and measurement of college is obtained.

These are phases of project which plays important role in our project. Various phases are link with each other, and need to follow the flow for their proper working.
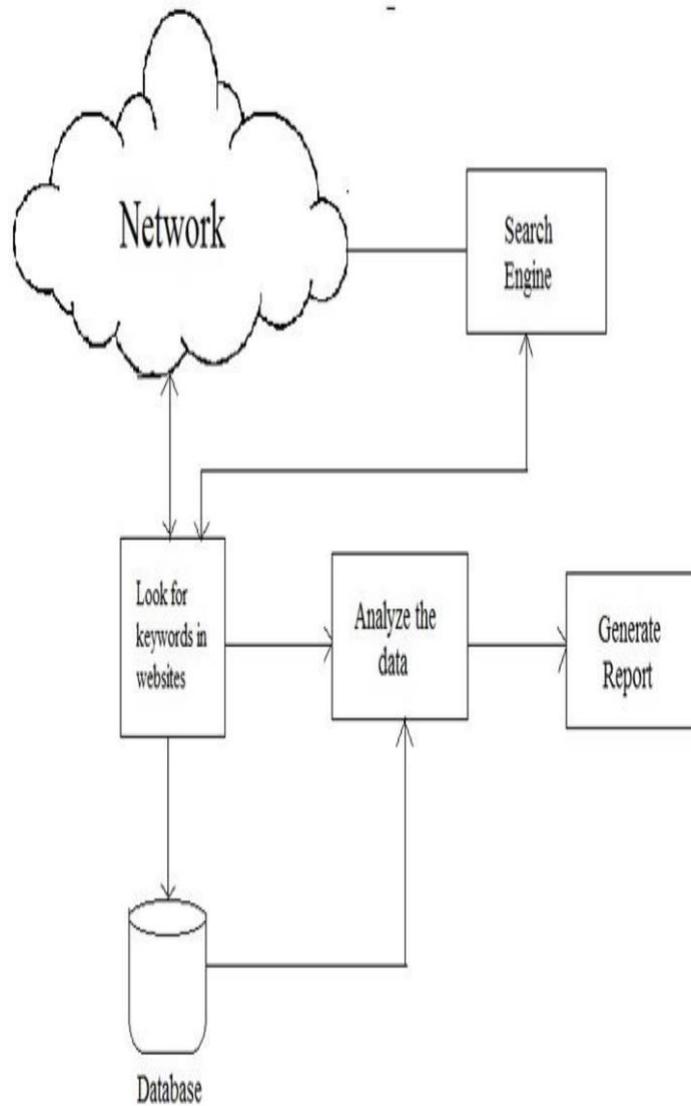
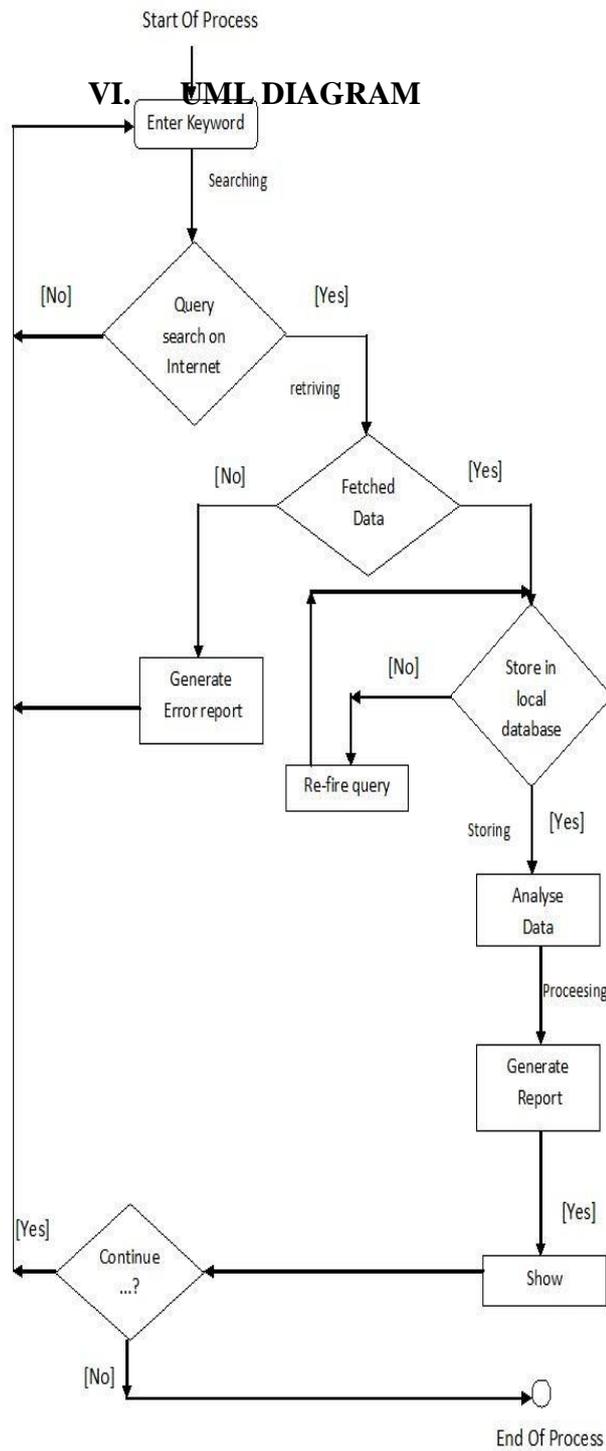## V. ARCHITECTURE DIAGRAM

Fig1. Architecture Diagram

VI.   UML DIAGRAM

Fig2. Activity Diagram

VII.   CONCLUSION AND FUTURE WORK

In this paper, we proposed an efficient as well as reliable system that will measure the college based on the performance ,reviews and comments from various websites. Parents and students will get the appropriate information about the college based on true ratings.

Clustering algorithms can be applied to categorize feedbacks based on reviews given by users. Filtering of fake resources which will make system more reliable.

ACKNOWLEDGMENT

The authors would like to thanks all the people associated with the project directly or indirectly. You all contributed to work in some way. In addition, We would like to thank the authors of all the papers, journals and books that we have referred while compiling our work. Your work means a lot to us and we have shown that in our paper.

REFERENCES

1. A Recommendation Ranking Model Based on Credit by Xu Xiaolin College of Computer and Information Shanghai Second Polytechnic University Shanghai, China.
2. Professional Tennis Player Ranking Strategy based Monte Carlo Feature Selection Ruifei Xie*, Bin Han*+, Lihua Li+, Juan Zhang, Lei Zhu College of Life Information Science and Instrument Engineering Hangzhou Dianzi University Hangzhou, P.R.China.
3. Predicting Students' grades Based on free style Comments Data by Artificial Neural Network by Shaymaa E. Sorour et al., 978-1-4799-3922-0/14/$31.00 ©2014 IEEE
4. Semantic Similarity Based Approach For Automatic Evaluation Of Free Text Answers Using Link Grammar By Udit Chakraborty., IEEE sixth International Conference On technology For Education 2014.
5. BPGM: A Big Graph Mining Tool by Yang Liu, Bin Wu, Hongxu Wang, and Pengjiang Ma.
6. Understanding Competition using Big Consumer Search Data by Daniel M. Ringel Goethe University Frankfurt and Bernd Skiera, Goethe University Frankfurt.
7. On Big data Analytics for Greener and Softer RAN by Chih-Lin , Yunlu Liu, Shuangfeng Han, Sihai Wang, Guangyi Liu, China Mobile Research Institute, Beijing 100053, P. R China
8. Fuzzy Based Framework for Professional College Rankings by N.Narayanan Prasanth, S. Mohiadeen Abdul Kader , K.Kannan, S.P.Raja, S. Arif Abdul Rahuman.
9. A Scalable Graph-based Ranking Model for Content-based Image Retrieval by Bin Xu, Student Member, IEEE, Jiajun Bu, Member, IEEE, Chun Chen, Member, IEEE, Can Wang, Member, IEEE, Deng Cai, Member, IEEE, and Xiaofei He, Senior Member, IEEE.
10. Inference Patterns from Big Data using Aggregation, Filtering and Tagging- A Survey Pathak Anand Prakashbhai M.Tech, CSE, ASET, Noida, India Hari Mohan Pandey CSE, ASET, Noida, India.
11. Relevance Feature Discovery for Text Mining by Yuefeng Li, Abdulmohsen Algarni, Mubarak Albathan, Yan Shen, and Moch Arif Bijaksana, JUNE 2015
12. On Big data Analytics for Greener and Softer RAN by Chih-Lin I, Yunlu Liu, Shuangfeng Han, Sihai Wang, Guangyi Liu China Mobile Research Institute, Beijing 100053, P. R. China, 2015