Design and development of the web based solution for analyzing and suggesting the educational opportunities

[1]Rutuja Pisal, [2]Prachi Tekawade, [3]Neha Punjabi, [4]Anuja Wararkar

Department of Computer Engineering,
Pimpri Chinchwad College of Engineering
Pune , India.

Abstract—**In order to realise the goal of demographic divide, every citizen of the country needs to be educated, be it a scientist, a doctor, engineer or whatever to meet not only the needs of our society but also the other societies elsewhere in the world. Every student aspires to become a great person and self reliant, but their passion and likings are diverse. In order to cater to the aspirations of the students and facilitate them to aim it high, they are required to be informed about the scope of various educational fields this country has in the future educational path, as also the availability of opportunities in all areas of employment. A database maintaining information about various educational fields, key interests necessary for a particular field, courses offered by various institutes in that field by both government as well as non-government institutes, future scope based on market analysis, job availability , salary and perks, research areas, etc can be maintained.The main objective of this application is to be interactive, informative and dynamic. The application provides information regarding various courses, institutes offering the course, current market trends, emerging technologies, future scope, rate of employability, salary, perks, etc**

Keywords— **Education , Path , web crawling , parsing , Data analyzing.**

I. INTRODUCTION

India is a young nation because it has the largest youth population.This youth has a lot of potential.But very few means of channelizing this potential is known.This is because most of the fields remains unexplored due to scarcity of the information sources available.A society needs not only doctors and engineers but professionals skilled in each and every field.For this the aspiring individuals must have access to information regarding various fields.A common platform thus can be designed that would help the individuals decide their educational path based on the abundant information available.It would also help them avoid making wrong decisions regarding choice of educational path which can end up in waste of time and money.Taken the right decisions at right time the career path would be on the right track.

II. LITERATURE SURVEY

Applications like Career Guide, Career Counselor, Career Guidance,Career Choice provide user with information about various fields only.It does not provide user with a particular path when a goal state is given.Some applications do provide user with minimum education needed to opt for the course however a path is not given that would give the user a more clear view.

Applications like 'My Pursuit' however does display a path,but the path displayed is a fixed path.Its does not take any input as a goal state and give the path.Thus the existing applications are just informational and not interactive.

www.stupidsid.com

Stupidsid is an online platform that helps students learn more effectively and make informed decisions about education. Stupidsid.com started off in February 2010 as a college review website that offers students' opinions on everything you have to know about colleges, courses and universities. Over time, iy has generated an extensive pool of content that helps students throughout their undergraduate journey.

www.shiksha.com

Shiksha.com is a one-stop-solution making course and college selection easy for students looking to pursue undergraduate (UG) and postgraduate (PG) courses in India and abroad; also accessible to users on the move through the website's mobile site.

The website is a repository of reliable and authentic information for over 14,000 institutions, 40,000 plus courses and has a registered database of more than 3.5 million students. We offer specific information for students interested in UG/PG courses in India

(shiksha.com) and Abroad (studyabroad.shiksha.com) across the most popular educational streams – Management; Science & Engineering; Banking & Finance; Information Technology; Animation, VFX, Gaming & Comics; Hospitality, Aviation & Tourism; Media, Films & Mass Communication; Design; Medicine, Beauty & Health Care; Retail; Arts, Law, Languages & Teaching; and Test Preparation.Thus,education seekers get a personalised experience on this site, based on educational background and career interest, enabling them to make well informed course and college decisions.

www.dtemaharsashtra.gov.in

The role of the Directorate is to maintain, enhance the standard, quality of technical education by laying the policies, establishing developing Govt. Institutions, guiding supervising the aided, private institutions, interacting with industry and national level institutions, coordinating with other departments of State Government, Government of India Statutory Organisations and to contribute to the development of industry society at large.

III.   TECHNOLOGY

DJango framework in Python-

Django is a free and open-source web framework, written in Python, which follows the model-view-template (MVT) architectural pattern. Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development. Python is used throughout, even for settings files and data models. Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models.

Some well-known sites that use Django include the Public Broadcasting Service,Instagram,Mozilla,The Washington Times, Disqus,Bitbucket, and Nextdoor.

Features of DJango -

1. Ridiculously fast

Django was designed to help developers take applications from concept to completion as quickly as possible.

2. Reassuringly secure

Django takes security seriously and helps developers avoid many common security mistakes, such as SQL injection, cross-site scripting, cross-site request forgery and clickjacking. Its user authentication system provides a secure way to manage user accounts and passwords.

3. Fully loaded

Django includes dozens of extras you can use to handle common Web development tasks. Django takes care of user authentication, content administration, site maps, RSS feeds, and many more tasks — right out of the box.

4.**Incredibly versatile**

Companies, organizations and governments have used Django to build all sorts of things — from content management systems to social networks to scientific computing platforms.

5.Exceedingly scalable

Some of the busiest sites on the planet use Django's ability to quickly and flexibly scale to meet the heaviest traffic demands.

DJango Components-

● A lightweight and standalone web server for development and testing
● A form serialization and validation system that can translate between HTML forms and values suitable for storage in the database
● A template system that utilizes the concept of inheritance borrowed from object-oriented programming
● A caching framework that can use any of several cache methods
● Support for middleware classes that can intervene at various stages of request processing and carry out custom functions
● An internal dispatcher system that allows components of an application to communicate events to each other via pre-defined signals
● An internationalization system, including translations of Django's own components into a variety of languages

- A serialization system that can produce and read XML and/or JSON representations of Django model instances
- A system for extending the capabilities of the template engine
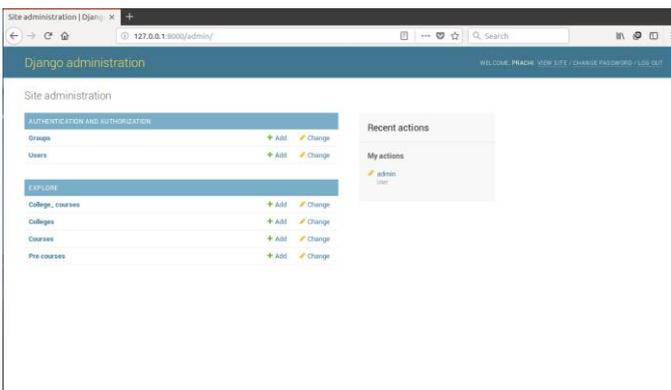- An interface to Python's built-in unit test framework

Web Crawlers-

A web crawler is an automated program that scans or crawls through the internet pages to create an index of the data. It is also known as Web spider, web robot or an automatic indexer. Search engines makes use of web crawlers to collect information about the data on public web pages, their primary purpose is to collect data so that when a user enters a search term on their site, they can quickly be provided with relevant websites.

Different types of crawlers-

1. Traditional crawlers-Visits entire Web and replaces index.
2. Periodic crawlers-Visits portions of the Web and updates subset   of index.
3. Incremental crawlers-Selectively searches the web and incrementally modifies index.
4. Focused crawlers-Visits pages related to a particular subject.

Different Web Crawlers-

1. Heritix
Language: Java

   2.   Nutch

Language: Java

   3.   Scrapy:

Scrapy is a free and open source web crawling framework, written in Python. Originally designed for web scraping, it can also be used to extract data using APIs or as a general purpose web crawler.

Scrapy project architecture is built around 'spiders', which are self-contained crawlers which are given a set of instructions. Following the spirit of other don't repeat yourself frameworks, such as Django, it makes it easier to build and scale large crawling projects by allowing developers to reuse their code.

NLTK-

The **Natural Language Toolkit**, or more commonly **NLTK**, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit,plus a cookbook.

NLTK is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning.NLTK has been used successfully as a teaching tool, as an individual study tool, and as a platform for prototyping and building research systems. There are 32 universities in the US and 25 countries using NLTK in their courses. NLTK supports classification, tokenization, stemming, tagging, parsing, and semantic reasoning functionalities.

Programming languages and platform used:

- Python
- SQL
- Angular JS
- DJango framework
- Bootstrap framework
- HTML
- CSS

Python-

Python is an interpreted high-level programming language for general-purpose programming. Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

How The Python Interpreter Works
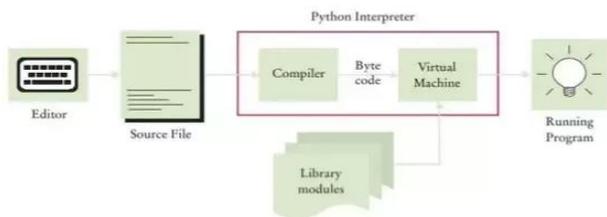
1. Simple
This pseudo-code nature of Python is one of its greatest strengths. It allows you to concentrate on the solution to the problem rather than the language itself.

2. Easy to Learn
As you will see, Python is extremely easy to get started with. Python has an extraordinarily simple syntax, as already mentioned.

3. Free and Open Source
FLOSS is based on the concept of a community which shares knowledge. This is one of the reasons why Python is so good - it has been created and is constantly improved by a community who just want to see a better Python.

4. High-level Language
When you write programs in Python, you never need to bother about the low-level details such as managing the memory used by your program, etc.

5. Portable
You can use Python on Linux, Windows, FreeBSD, Macintosh, Solaris, OS/2, Amiga, AROS, AS/400, BeOS, OS/390, z/OS, Palm OS, QNX, VMS, Psion, Acorn RISC OS, VxWorks, PlayStation, Sharp Zaurus, Windows CE and even PocketPC .

6. Object Oriented

In object-oriented languages, the program is built around objects which combine data and functionality. Python has a very powerful but simplistic way of doing OOP, especially when compared to big languages like C++ or Java.

7. Extensible

If you need a critical piece of code to run very fast or want to have some piece of algorithm not to be open, you can code that part of your program in C or C++ and then use them from your Python program.

8. Embeddable

You can embed Python within your C/C++ programs to give 'scripting' capabilities for your program's users.**Python** is a high-level programming language. An interpreted language. Python features a dynamic type system and automatic memory management and supports multiple programming paradigms, including object-oriented, imperative, functional programming, and procedural styles. It has a large and comprehensive standard library. Python interpreters are available for many operating systems, allowing Python code to run on a wide variety of systems.

Bootstrap framework -

Bootstrap is a free and open-source front-end web framework for designing websites and web applications. It contains HTML- and CSS-based design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions. Unlike many web frameworks, it concerns itself with front-end development only.

Bootstrap is modular and consists of a series of Less stylesheets that implement the various components of the toolkit. These stylesheets are generally compiled into a bundle and included in web pages, but individual components can be included or removed. Bootstrap provides a number of configuration variables that control things such as color and padding of various components.

Html-

Hypertext Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript it forms a triad of cornerstone technologies for the WWW.

Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as <img /> and <input /> introduce content into the page directly. Others such as <p>...</p> surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language.[1] Although most often used to set the visual style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any XML document, including plain XML, SVG and XUL, and is applicable to rendering in speech, or on other media. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging web pages , user interfaces for web applications, and user interfaces for many mobile applications.

CSS is designed primarily to enable the separation of presentation and content, including aspects such as the layout, colors, and fonts.This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple HTML pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.

Separation of formatting and content makes it possible to present the same markup page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. It can also display the web page differently depending on the screen size or viewing device. Readers can also specify a different style sheet, such as a CSS file stored on their own computer, to override the one the author specified.

IV. ALGORITHMS

A- Depth First Search Algorithm

Depth-first search (DFS) is an algorithm for traversing or searching tree or graph data structures. One starts at the root

(selecting some arbitrary node as the root in the case of a graph) and explores as far as possible along each branch before backtracking.

Input: A graph G and a vertex v of G

Output: All vertices reachable from v labeled as discovered

Procedure -
```
DFS-iterative(G,v):
    let S be a stack
    S.push(v)
    while S is not empty
      v = S.pop()
    if v is not labeled as discovered:
       label v as discovered
       for all edges from v to w in  G.adjacentEdges(v) do
          S.push(w)
```

B- Breadth First Search Algorithm

Breadth-first search (BFS) is an algorithm for traversing or searching tree or graph data structures. It starts at the tree root and explores the neighbor nodes first, before moving to the next level neighbours

Pseudocode-

```
create a queue Q
mark v as visited and put v into Q
while Q is non-empty
   remove the head u of Q
   mark and enqueue all (unvisited) neighbours of u
```

V. SOFTWARE REQUIREMENT SPECIFICATION

DETAIL PROBLEM DEFINITION

As the problem statement describes , the purpose of the system is to help students to decide their career path by providing them information about various options available according to their current education and preferences.

The first step is to get academic details of student , like which class he/she is studying in or the course completed recently , whether he/she has given some competitive exams .Then based on this data and preferences students will be given options to explore.

The second step is to show the education path. If the student has decided his/her goal in terms of education , the system then should display the path student has to follow from his current position to reach this goal. The description of path will contain information about courses , time required to complete that course , various subjects included , probable estimated cost , universities and colleges offering that particular course , etc. If more than one path exists then the users are given difference between these paths which will make choice easy for the user.
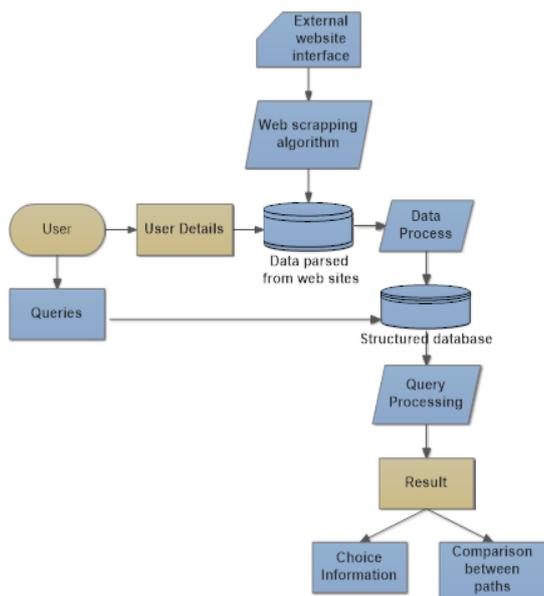
Block Diagram of System

Software requirement

Platform :

1. Operating System:
Linux
Windows
Android

2. Programming Language
    Python
    HTML , CSS , Java Script
SQL

System Features

1.Scalable
Although the focus of this system is on engineering domain , The scope of the system can be increased with little or no changes made to the system .Only additional data must be added to the database , and structure remains the same.

2.User friendly
The user will find the system easy to use and understand due to interactive user interfaces and display of result in form of graph which is easily understandable. This is very useful as it allows user to easily make comparisons .Also the information is personalized not generalized unlike many government and commercial websites.

3.Dynamic and automatic
As the system uses web crawlers and then parsers to gather data , the data is dynamically pulled from various documents and websites , thus there is no need to manually enter data every year . The admin makes sure that data is reliable.

4. Informative
The system provides information regarding various parameters like courses offered , time required , eligibility criteria , probable cost , etc . All this information makes it convenient for users to choose the right education path.

Other Non- functional requirements

1. Personal Blogs
The system facilitates the students to share their experience by writing blogs which will be helpful to other students.

2. Group Discussion
Students can discuss any topic among themselves , they can start a new discussion or participate in an ongoing discussion.

3. Recent tweets regarding an educational domain
The home page of the system can be populated with tweets tweeted by a knowledgeable person , which will help keep students updated with current scenario and trends.

Performance requirements

1.The system should respond to a query in minimal time , i.e it should be quick to respond.
2.The system should work flawlessly even when number of users logged in increases . It should not show any delay denoting a system overload.
3.The result displayed should be correct and reliable. It is responsibility of system admin to authenticate the data entered in the structured database if its source is less reliable.
4.The system should not show any speed or memory bottleneck , due to its client and server architecture.
5.The system should have backup of the database to avoid any single point of failure in system.

Safety requirements

1.The application should be secure enough to prevent a data breach and maintain the integrity of database.
2.The privacy of user details should be maintained.
3.Admin must administer the tweets , blogs and conversations to keep them threat free.

Software Quality attributes

1.Adaptability:The software is adaptable by all users.

2.Availability:The availability of the software is easy and for everyone.The server should be available 24*7.

3.Correctness:The path information should be correct, accurate and reliable..

4.Flexibility:Our system can be integrated into any educational institute.Its scope can also be widened to integrate into the corporate world.

5.Maintainability:After the deployment of the project if any error occurs then it can be easily maintained by the software developer.

6.Reliability:The system ensures that each component delivers its primary functionality and interacts with others in harmony.

7.Testability:The system will be tested considering all the aspects.

8.Usability:To perform and to understand the functioning of the systems is very easy.

VI.   CONCLUSION

"EDUSCOPE" is implemented as per the identified requirements of the education system.The system gives the educational choices for a particular individual according to their preferences and interest. System successfully Performs the Following Functionalities:

1.   Gives proper educational path as per the individual interest.

2.   Comparative study between the paths.

REFERENCES

[1]   Leif Azzopardi and David Maxwell ,Tango With Django: A beginner's Guide to Web Development With Python / Django 1.9, 2016.
[2]   Ayman Hourieh , Learning Website development with Django, 2008.
[3]   Antonio Melé, Django By Example-Create your own line of successful web applications with Django,2015.
[4]   Dimitrios Kouzis-Loukas, Learning Scrapy -Learn the art of efficient web scraping and crawling with Python.
[5]   Career360 [Online]. Available : https://www.careers360.com/courses-after-10th-12th

[6]   Shiksha [Online]. Available : https://www.shiksha.com/b-tech/ranking/top-engineering-colleges-in-india.
[7]   Education After 12th [Online]. AVAILABLE :   HTTPS://WWW.EDUCATIONAFTER12TH.COM/DIGITAL-MARKETING-SEO-COURSE-DETAIL-SYLLABUS