**Programming with Python for Data Sciences**

**Project Title: Develop an SEO Tool to Analyze Live Web Pages**

**Synopsis:**

Search Engine Optimization (SEO) is an important aspect of a Web page to gain importance for a search engine to be able to display it earlier in the search list. The optimization is based on a lot of factors such as title, description, header tags and keyword density. Different search engines will have their own mechanisms that calculate the score of a keyword on the page and thus work out its ranking in the search order.

**Course Objective:**

The course project is to develop a generic SEO toolset which will help a Web developer analyze live Web pages for keywords and other components of the page which contribute to SEO. We should be able to provide the keywords of interest and understand its density pattern across the various components of the HTML page.

**Benefits of SEO:**

1. **Cost-effectiveness** – SEO is one of the most cost-effective marketing strategies because it targets users who are actively looking for your products and services online.
2. **Increased site usability** – In an effort to make your website easier to navigate for the search engines, SEO simultaneously helps to make your website more navigable for users as well.
3. **Brand Awareness** – Since top position rankings result in significant impressions, having your website in these top positions on the result pages translates to more exposure for your website.
4. **Increased Traffic** – Top positions on the search engine result pages receive a majority of the impressions and clicks, so ranking in these top positions can result in significant traffic increases for your website.
5. **ROI** – SEO provides trackable and quantifiable results, regardless of whether you are an e-commerce or non-e-commerce site so there are no qualms when it comes to ROI.

**Modules:**

The following modules have been used in this project:

1. RE: A **regular expression** is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern.
2. SQLITE3: S**QLite** is a C library that provides a lightweight disk-based database that doesn't require a separate server process and allows accessing the database using a nonstandard variant of the SQL query language
3. URLLIB: **urllib** is a **Python** module that can be used for opening URLs. It defines functions and classes to help in URL actions.
4. XLSXWRITER: **XlsxWriter** is a **Python** module for creating Excel XLSX files. **XlsxWriter** is a **Python** module that can be used to write text, numbers, formulas and hyperlinks to multiple worksheets in an Excel 2007+ XLSX file.
5. OPERATOR: The **operator module** exports a set of efficient functions corresponding to the intrinsic **operators** of **Python**.
6. BEAUTIFUL SOUP4: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping
7. URLLIB.ERROR: The **urllib**.**error** module defines the **exception** classes for exceptions raised by **urllib**.request

**Working Pattern of The Project:**

We take the url of the first website from the user and then process the words stored in it.

We remove the frequently and most commonly used words from the website so that only the keywords will be in the system.We then also ignore all the special characters from the website content and store the first 20 keywords repeated in that website and store it the database.

We then extract the information from the database and store it in excel spreadsheet and show the graph of the keywords with their values for visual analysis. The same process is repeated for any number of websites we want to know about.

**Conclusion**:

The SEO Tool to Analyze Live Web Pages developed by us has made the best possible efforts to satisfy the needs of analysing web pages.

The details can be accessed and  the word count is done according to the rules and regulations of particular search engine in shortest time frame and avoid the commonly used words.

Efficiently manage website information with easy to use interfacing and process the most frequent words along with satisfying features like graphs and work sheet. So that it promotes clear, transparent, accountable and user friendly administration.

**THANK YOU**