**Develop A SEO Tool to Analyse Live Web Pages**

# **Introduction**

Search Engine Optimization (SEO) is 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 mechanism that calculate the score of a keyword on the page and thus work out its ranking in the search order.

# **Synopsis**

The course project is to develop a generic SEO toolset which will help a Web developer analyse live Web pages for keywords and other components of the page which contribute to SEO. The user should be able to provide the keywords of interest and understand their density patterns across the various components of the HTML page. The Web page analysis should be developed to be done in a batch mode where you can analyse hundreds of Web pages and the results should be saved as reports in spread sheets with graph presentations wherever necessary.

# **Different Steps Used for This SEO Project**

* Read Any Website
* Extract only Relevant Words
* Make a list of words
* Remove the words from the list of words
* Create Dictionary with word as key and frequency as Value
* Create a table which should store: Website, Words, Count
* Create Xlsx file to create a chart

# **Different Part of The SEO system:**

**XLSXWriter:** 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. It supports features such as formatting and many more, including:

* 100% compatible Excel XLSX files.
* Full formatting.
* Merged cells.
* Defined names.
* Charts.
* Auto-filters.
* Data validation and dropdown lists.
* Conditional formatting.
* Worksheet PNG/JPEG/BMP/WMF/EMF images.
* Rich multi-format strings.
* Cell comments.
* Textboxes.
* Integration with Pandas.
* Memory optimization mode for writing large files.

**Dictionary:** Dictionary in Python is an unordered collection of data values, used to store data values like a map, which unlike other Data Types that hold only single value as an element, Dictionary holds key: value pair. Key value is provided in the dictionary to make it more optimized. Each key-value pair in a Dictionary is separated by a colon(:), whereas each key is separated by a ‘comma’.

A Dictionary in Python works similar to the Dictionary in a real world. Keys of a Dictionary must be unique and of immutable data type such as Strings, Integers and tuples, but the key-values can be repeated and be of any type.

Example:

Empty Dictionary: {}

Dictionary with the use of Integer Keys: {1: 'Geeks', 2: 'For', 3: 'Geeks'}

**Database:** A database is an abstraction over an operating system's file system that makes it easier for developers to build applications that create, read, update and delete persistent data. Databases storage implementations vary in complexity. SQLite, a database included with Python, creates a single file for all data per database. Other databases such as PostgreSQL, MySQL, Oracle and Microsoft SQL Server have more complicated persistence schemes while offering additional advanced features that are useful for web application data storage. These advanced features include but are not limited to:

* data replication between a master database and one or more read-only slave instances
* advanced column types that can efficiently store semi-structured data such as JavaScript Object Notation (JSON)
* sharding, which allows horizontal scaling of multiple databases that each serve as read-write instances at the cost of latency in data consistency
* monitoring, statistics and other useful runtime information for database schemas and tables

**JSON**: JSON (JavaScript Object Notation) is a popular data format used for representing structured data. It's common to transmit and receive data between a server and web application in JSON format.

In Python, JSON exists as a string. For example:

Example: p = '{"name": "Bob", "languages": ["Python", "Java"]}'

It's also common to store a JSON object in a file.

**Regular Expression:**  A regular expression in a programming language is a special text string used for describing a search pattern. It is extremely useful for extracting information from text such as code, files, log, spreadsheets or even documents.

While using the regular expression the first thing is to recognize is that everything is essentially a character, and we are writing patterns to match a specific sequence of characters also referred as string. Ascii or Latin letters are those that are on your keyboards and Unicode is used to match the foreign text. It includes digits and punctuation and all special characters like $#@! %, etc.