Smart Search Engine Project

Requirements of Smart Search Engine Project:

1. Python
2. Flask
3. Natural Language Processing
4. Database
5. OCR

1. Python:

Python is a powerful general-purpose programming language. In today's world, almost anything can be implemented in Python and it is one of the most popular programming languages of the 21st century. This project will be implemented in Python. Different libraries in Python will be used to efficiently implement this project.

2. Flask

To give GUI to our NLP model. The model will be deployed on Flask Web Framework.

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies, and several common framework-related tools. Its integration with a database can be easily done through my-sqlalchemy library of Python.

Because of its simplicity, Flask is used in top tech companies like

- Netflix
- Reddit
- Airbnb
- Lyft
- Mozilla
- Uber

3. Natural Language Processing:

Natural language processing (NLP) is a field of artificial intelligence in which computers analyse, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

In our project, NLP forms the core as most of the work will be revolved around it. It's sub-topics like TF-IDF will play a vital role in finding the most common words in a document.

4. **Database:**

A database is a systematic collection of data. They support electronic storage and manipulation of data. Databases make data management easy. Different types of databases are there like MySQL, MS SQL, etc. which makes the storage of large amount of data easy, secure and can retrieve fast.

5. **Optical Character Recognition (OCR):**

OCR (optical character recognition) is the use of technology to distinguish printed or handwritten text characters inside digital images of physical documents, such as a scanned paper document. The basic process of OCR involves examining the text of a document and translating the characters into code that can be used for data processing. OCR is sometimes also referred to as text recognition.

OCR systems are made up of a combination of hardware and software that is used to convert physical documents into machine-readable text.

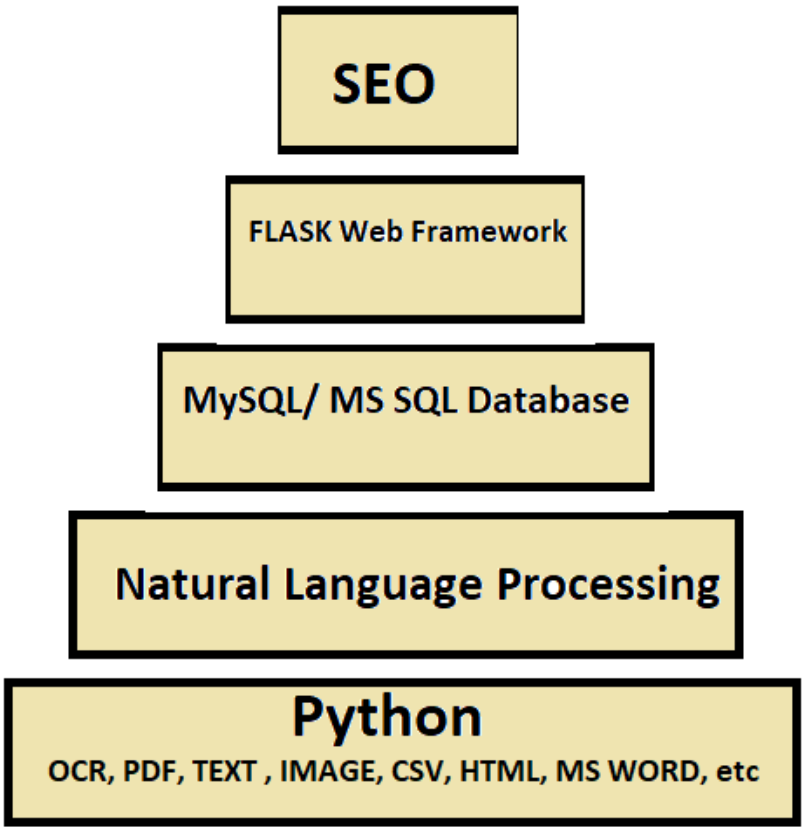

Figure 1 Framework of the Project