**SEARCH ENGINE:**

The current python script allows us to search and crawl the Wikipedia page of any given keyword (provided the Wikipedia page exists). The user can either crawl in the displayed links or can enter a new keyword to browse and crawl.

**Design:**

The following are the various Functions used in the code:

1. error\_handler: This code has some error handling for HTTP requests. Whenever there is an error in the spelling entered or if there is no Wikipedia page available with the keyword entered, then it throws an error.
2. download\_page: Downloads a web page using the requests library.
3. url\_parse: Parses and normalizes URLs.
4. open\_search\_results: Opens Wikipedia search results in a web browser.
5. extract\_title, extract\_introduction, extract\_links: Extract information (title, introduction, links) from the raw HTML content of a page.
6. remove\_html\_tags: Removes HTML tags from text.
7. extension\_scan: Checks if a URL has certain file extensions (e.g., image files) to skip them during crawling.
8. web\_crawl: The main web crawling function that explores Wikipedia pages, extracts information, and stores it in a dictionary.
9. main: The code enters a main loop where you can input a search query or press Enter to crawl the default Wikipedia page. It also allows you to view details of extracted pages by entering a link number.

**How does Execution occur?**

1. Start the python script.
2. You'll be prompted to either press Enter to search the default Wikipedia page or enter a search query.
3. If you enter a search query, the script will open the Wikipedia search results in the web browser and begin crawling the selected page.
4. During crawling, it extracts titles, introductions, and links from the pages it visits and stores them in a dictionary which is the database.
5. After crawling, you can enter link numbers to view details of the pages so extracted.
6. To exit the program, enter ‘q’.

**Future scope:**

The code could be made more generalised as in it could be used for any keyword entered and not specifically searching and crawling in the Wikipedia page.