**🌟 doclens - 🔍 search engine for docx files 🌟**

**Environment**

* Python 3.x
* Windows OS

**Build Instructions**

* Install pyinstaller using the following command in Command Prompt
* pip install pyinstaller
* Clone the repository from <https://github.com/raajkumar-s/doclens.git>
* cd to doclens directory
* Execute the build.bat file to create the executable
* biuld.bat
* If build is successful, the executable will be available in doclens/dist/bin

**Usage Instructions**

* Open the Command Prompt, navigate to doclens/bin directory and execute the executable using the following command
* doclens.exe
* By default, the assignment document collection with 6 docs are used for indesing and searching. To provide new document corpur pass the path of the directory which contains documents
* doclens.exe /newdoc-collection
* You can also double-click the doclens.exe file directly to run the program
* Enter your search query and hit Enter key. Wrap your query in "" (double-quotes) for searching a phrase.
* If you want to exit, hit Enter key without providing any search query.
* Supported query types
  1. One Word Queries
  2. awesome
  3. Free Text Queries
  4. you are awesome
  5. Phrase Queries
  6. "you are awesome"

**Assumptions**

* Document contains only plain text (No image, table, etc.,)
* Text is only in English Language

**Limitations**

* Only .docx files are supported
* If document contains multimedia, they are ignored
* Supports only Windows platform. For other platforms, directly execute the source code using python (Python 3.x should be installed)
* python main.py

**Libraries/References Used**

* PorterStemmer - For stemming the index tokens/terms
* List of Stop Words: <http://xpo6.com/download-stop-word-list/>