
Search engine for COVID-19

IN104

Jessica López Espejel

ENSTA, Institut Polytechnique Paris

How an index works?



User query

Image taken from internet



Thousands of documents

Image taken from internet

Important dates

11/ May/ 2021 - Finish the code:

- Argparse
- Get keywords from the user query
- Docstrings

18/ May/ 2021

- Project report
- README.md

25/ May/ 2021 - Project presentation

Argparse

<https://docs.python.org/3/howto/argparse.html>

Execution:

```
python main.py -user_query "types of cancer in the human body"
```

```
/home/jessica/Documents/data/IN_104-Projet-Informatique/clean_docs/PMC8021921.txt
```

```
/home/jessica/Documents/data/IN_104-Projet-Informatique/clean_docs/PMC7222958.txt
```

```
/home/jessica/Documents/data/IN_104-Projet-Informatique/clean_docs/PMC7509820.txt
```

```
/home/jessica/Documents/data/IN_104-Projet-Informatique/clean_docs/PMC7750180.txt
```

```
/home/jessica/Documents/data/IN_104-Projet-Informatique/clean_docs/PMC7533576.txt
```

Keywords

https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

Docstrings

<https://www.programiz.com/python-programming/docstrings>