Intelligent Semantic Search

1. <https://medium.com/analytics-vidhya/semantic-search-engine-using-nlp-cec19e8cfa7e>
2. <https://towardsdatascience.com/natural-language-processing-document-search-using-spacy-and-python-820acdf604af>
3. <https://towardsdatascience.com/semantic-code-search-3cd6d244a39c>
4. <https://www.bloomreach.com/en/blog/2019/06/semantic-search-explained-in-5-minutes.html>

<https://github.com/cadobe/bison/tree/master/msmarco_doc_train>

Code: <https://github.com/mymycen/SW-LD_NLP_project>

<https://www.kaggle.com/ajitrajput/semantic-search-engine-using-nlp>

<https://github.com/bhavsarpratik/semantic-search>

<https://github.com/SREEHARI1994/Semantic-search-engine>

Testing: <https://www.pinecone.io/learn/train-sentence-transformers-softmax/>

<https://evergreenllc2020.medium.com/semantic-search-engine-with-s-abbfb3cd9377>

<https://www.kaggle.com/Cornell-University/arxiv?select=arxiv-metadata-oai-snapshot.json>

<https://aclanthology.org/D19-1410.pdf>

<https://www.sbert.net/examples/applications/semantic-search/README.html#>

Data Links:

1. Medium Articles : <https://www.kaggle.com/hsankesara/medium-articles>
2. BBC Articles : <https://www.kaggle.com/yufengdev/bbc-fulltext-and-category>
3. Research Articles : <https://www.kaggle.com/blessondensil294/topic-modeling-for-research-articles>
4. Financial news articles: <https://www.kaggle.com/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests>
5. News(already used) : <https://www.kaggle.com/therohk/million-headlines>
6. Economic News Articles : <https://www.kaggle.com/heeraldedhia/us-economic-news-articles>
7. Financial Articles: <https://www.kaggle.com/notlucasp/financial-news-headlines>
8. India news: <https://www.kaggle.com/therohk/india-headlines-news-dataset>
9. Covid articles : <https://www.kaggle.com/kashnitsky/covid19-articles-by-elsevier?select=covid_artilces_elsevier_train.csv>
10. Geeks for geeks articles : <https://www.kaggle.com/naidukarthi2193/geeks-for-geeks-articles-dataset>
11. BBC news articles : <https://www.kaggle.com/hgultekin/bbcnewsarchive>
12. Buzzfeed articles: <https://www.kaggle.com/promptcloud/articles-from-buzzfeed-2020>
13. NY Times Articles : <https://www.kaggle.com/parsonsandrew1/nytimes-article-lead-paragraphs-18512017>
14. Nature Research articles: <https://www.kaggle.com/wardasuhaim/published-articles-in-nature-research-2019>
15. The Tribune News Articles: <https://www.kaggle.com/waseemakramkhan/the-tribune-news-articles>

Term Paper: we are assuming that we have whole data in form of text documents, Knowledge graph for unstructured data

Train/test split

<https://oitncsu-my.sharepoint.com/:p:/g/personal/jjayaty_ncsu_edu/ERsgc81GyWpKigU6_4clQGgBOfE0Gbqo1EHQwJAoNF8FjQ?e=59I2Ua>