

# Introduction

Semantic search is a search technology that leverages natural language processing (NLP) to better understand the intent behind user queries and to deliver more relevant search results. Traditional keyword-based search engines are limited in their ability to understand the nuances and context of language, which can lead to irrelevant or incomplete search results.

In contrast, semantic search engines use NLP to analyze the meaning and context of words and phrases in a search query. This allows the search engine to consider the user's intent, the relationships between words, and the context in which they are used, enabling it to provide more accurate and relevant search results.

NLP techniques such as named entity recognition, sentiment analysis, and text classification are used to extract meaning from unstructured data sources such as web pages, social media, and documents. These techniques help to identify the key concepts and entities mentioned in a search query and to match them with relevant content.

Overall, semantic search using NLP is a powerful tool that enables more intelligent and intuitive search experiences, making it easier for users to find the information they need, even when they don't know exactly what they're looking for.

## DATA DESCRIPTION

### Overview

About 100 Politics text data refers to any text-based information related to political issues, events, or discussions. This can include news articles, political speeches, social media posts, government reports, academic studies, and more.

### Content

### Politics text data

## TOOLS

- Python
- Nltk
- Numpy
- Tensorflow
- Genism
- Tf-idf
- Cosine-similarity

# Baseline Experiments

THE GOAL → searching some of words in English articles then extract hot keywords.

Steps of Experiments :

1. Load data (making function to reading 100 txt data )
2. Text cleaning and pre-processing  
Text preprocessing is the process of transforming raw text data into a format that is more suitable for analysis by machine learning algorithms or other natural language processing (NLP) techniques. The goal of text preprocessing is to clean, normalize, and transform raw text data so that it can be effectively analyzed and used to extract meaningful insights.
  - Tokenization
  - Stopword removal
  - Stemming and Lemmatization
  - Lowercasing
  - Removing special characters and punctuation
  - word tokenize
3. making a function that takes a word as an input and search about in the article and generate an article that related to this word using
  - preprocessing function (step 2)
  - Tf-idf vectorizer to fit and transform words
  - Calculate similarity by using Cosine similarity
  - Put the words that the most similarity with the input words in a list and print it
  - Making a loop to search in all the trained articles and printing the articles that related and most similarity with the input words

## RESULTS

Printing the most words related to the input words and the articles that the most similar

## CONCLUSION

In conclusion, semantic search is a powerful technology that uses natural language processing (NLP) to better understand the context and meaning of user queries and to deliver more accurate and relevant search results. By analyzing the relationships between words, identifying key concepts and entities, and considering the context in which they are used, semantic search engines are able to provide a more intuitive and personalized search experience

## References

1. <https://www.deepset.ai/blog/how-to-build-a-semantic-search-engine-in-python>
2. <https://medium.com/captech-corner/semantic-search-with-deep-learning-and-python-3062b101a953>
3. <https://www.google.com/search?q=Semantic+search+NLP+Python&oq=semantic+sea&aqs=chrome.1.69i57j69i59l3j0i512j69i60l3.7359j0j7&sourceid=chrome&ie=UTF-8>
4. <https://blog.ml6.eu/semantic-search-a-practical-overview-bf2515e7be76>