

SmartFind: Information retrieval using NLP

Team Members:

- **Nikhila Polkampally**
- **Ruchitha Vangala**
- **Eshwar Gandu**
- **Ganesh Tummalapalli**

GitHub Link:

<https://github.com/VangalaRuchitha/NLP-PROJECT-CSCE-5290-Group-12>

Motivation:

Project focused on information retrieval using NLP is to improve the efficiency and effectiveness of information retrieval for professionals, researchers, and other stakeholders who need to analyse large volumes of data. Traditional keyword-based search engines often yield many irrelevant results, making it challenging for users to find the information they need. In contrast, NLP techniques can help to identify and extract relevant information from unstructured data sources such as text, audio, or video, which can be particularly valuable in domains such as healthcare, finance, and law. By developing and implementing NLP-based techniques, the project can provide a more accurate and efficient method of information retrieval that can enhance productivity, support better decision-making, and facilitate new insights and discoveries.

Significance:

The main significance of this project is to improve the efficiency, accuracy, and user experience of searching and retrieving information from large collections of unstructured data.

Objectives:

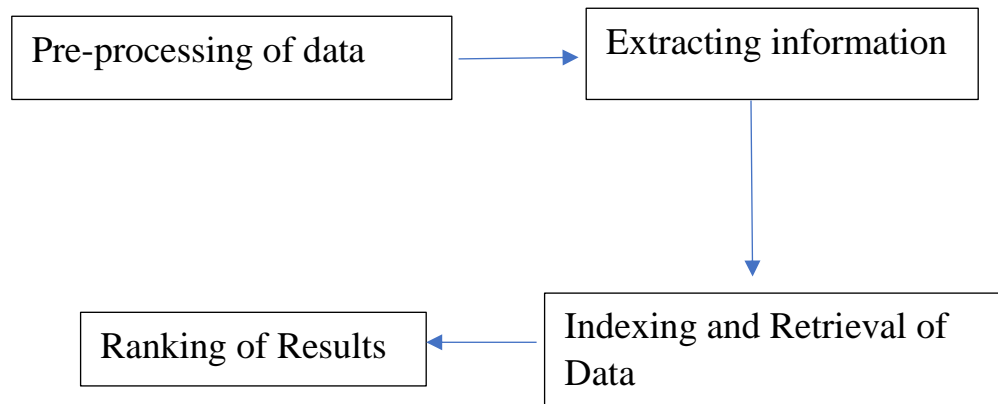
- **Improve the accuracy and relevance of information retrieval:** Develop and implement NLP techniques such as entity recognition, keyword extraction, text classification, and named entity disambiguation to improve the accuracy and relevance of information retrieval.

- **Designing and building a search engine or recommendation system:** Design and build a search engine or recommendation system that utilizes NLP techniques to retrieve and rank relevant information based on user queries or preferences.
- **Evaluating and comparing the performance:** evaluate and compare the effectiveness of various NLP techniques for information retrieval, such as comparing the accuracy of keyword-based search

Features:

Improve the accuracy score of retrieving relevant information from large volumes of text data, while also providing a user-friendly interface and incorporating feedback to continuously improve the system's performance. The algorithm should rank the retrieved documents in order of relevance to the user's query.

Visualization:



Here we pre-process the unstructured collected data and then extract the necessary information from that pre-processed data and based on the information extracted we index and retrieve the data and in final step we give ranking to the most frequently occurred results.

References:

1. Croft, W.B., Turtle, H.R., and Lewis, D.D. The use of phrases and structured queries in information retrieval. In *Proceedings of the 14th Annual International ACM SIGIR*

Conference on Research and Development in Information Retrieval (Oct. 13–16, Chicago, Ill.). ACM/SIGIR, New York, 1991, 32–45.

2. Willett, P., Ed. *Document Retrieval Systems* Taylor Graham, London, 1988
3. Salton, G. and Buckley, C. Improving retrieval performance by relevance feedback. *J. American Society for Information Science*, 41, 4 (1990), 288—297.
4. Liddy, E. D. (2001). Natural language processing.