Cross-Language Information Retrieval Using NLP and Deep Learning

Abstract:
Information retrieval across languages is a challenging task, especially when the user query and the relevant content are in different languages. This project aims to develop a **cross-language information retrieval (CLIR)** system using **Natural Language Processing (NLP)** and **deep learning** to bridge language gaps in information searching.

Methodology:
The system will preprocess text in multiple languages using standard NLP techniques, including **tokenization**, **stopword removal**, and **semantic embedding** (e.g., **Word2Vec** or **BERT**). **Cross-lingual embedding models** will be employed to map different languages into a shared space, allowing the system to understand queries in one language and retrieve relevant documents in another language. Deep learning models such as **Transformers** and **BERT-based multilingual models** will be used for semantic matching and ranking of search results.

Outcome:
The expected outcome is a cross-language search engine that can understand user queries in one language and retrieve documents from various languages. This system will improve the accessibility and effectiveness of information retrieval for global users and support organizations in managing multilingual content efficiently.