INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 9, Issue 5 - V9I5-1189)
Available online at: https://www.ijariit.com

Keyword-Based Search Engine Using Cosine Similarity

Ancha Venkata Lakshmi Narasimha
narasimha.ancha@gmail.com
Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh

Athalury Rohith
rohithskbp@gmail.com
Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh

Bakkamantala Teja Kumar
tejakumarbakkamantala@gmail.com
Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh

Chakka Venkateswara Rao
venkatchinnu216@gmail.com
Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh

Rekha Sudha Kishore
sudhakishore@vvit.net
Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh

ABSTRACT

The increasing volume of digital information available on the internet necessitates the development of effective search engines that can retrieve relevant content based on user-specified keywords. This abstract presents a novel approach to building a keyword-based search engine that utilizes cosine similarity to enhance search result relevance and precision.
Our keyword-based search engine using cosine similarity offers several advantages over traditional search engines. It improves the accuracy of search results by considering the semantic relationships between words, leading to a more contextually relevant ranking. Users can expect better search experiences with increased precision and efficiency in finding the information they seek.
Keywords: *Keyword-Based Search Engine, Cosine Similarity, Search Results, Ranking, Threshold Value, Information Retrieval.*

1. INRODUCTION

The main aim of our project is to find the entertainment, motivational videos on the web, our project helps the user to find the videos directly by using the keyword, based on their mood, and based on their thoughts.
In the digital age, the relentless expansion of online information has elevated the role of search engines to paramount importance. Internet users increasingly rely on these tools to navigate the vast expanse of data and extract meaningful insights. While conventional search engines have revolutionized information retrieval, they often face challenges when it comes to delivering results that are not just keyword-matched, but truly relevant in terms of context and semantics. This is where our project, the "Keyword-Based Search Engine Using Cosine Similarity,"
This project holds the potential to redefine how we approach information retrieval on the web. By enhancing the precision and relevance of search results, it promises to streamline the search process, empowering users to discover information with greater ease and efficiency. In the following sections, we will delve deeper into the architecture and functioning of our keyword-based search engine using cosine similarity, shedding light on its advantages and practical implications. As we explore these aspects, it becomes evident that our project is not just a technological innovation but a catalyst for a more refined and efficient online search experience.

2. PROPOSED MODEL

Our proposed model for a keyword-based search engine using cosine similarity is designed to provide an innovative and efficient approach to information retrieval. This model leverages advanced techniques to improve search result relevance and precision, addressing the limitations of traditional keyword-based search engines.

2.1 QUERY PROCESSOR

The query processor is responsible for handling user queries. When a user submits a query, this component tokenizes and preprocesses the query, extracting important keywords. It may also perform tasks like removing stop words and stemming to enhance query understanding.

2.2 Cosine Similarity Technique

The cornerstone of our model is the cosine similarity algorithm, which calculates the similarity between the keywords in a user's query and the keywords in the indexed documents.
The cosine similarity algorithm transforms keywords into vector representations and measures the cosine of the angle between these vectors, providing a numerical measure of similarity.

3. PROPOSED MODEL ARCHITECTURE

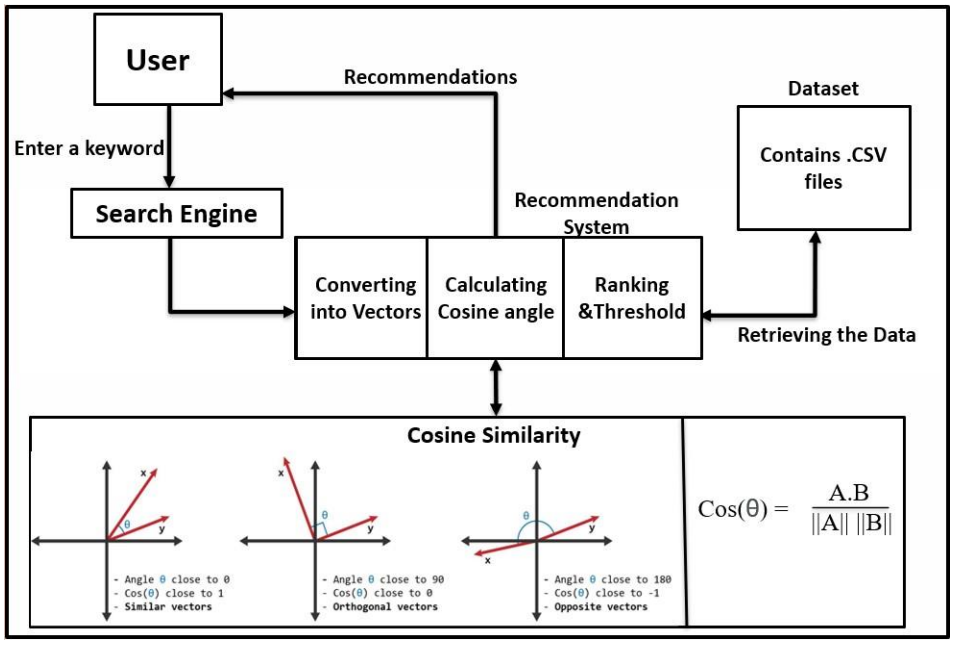

Fig: Architecture for Keyword Based Search Engine using Cosine Similarity

The project starts by entering a keyword in the Search Engine by the User. Then the keyword goes to the Recommendation system and the Recommendation system retrieves the data from the Dataset and in the Recommendation system we use a technique called Cosine Similarity. So the video title in the data set are converts into Vectors, then the cosine similarity technique calculates the cosine angle, then the titles will get ranked if they matches the threshold value then the titles are similar to the enteres keyword, then the Recommendation system sends the recommendations to the user. The working of cosine similarity technique is explained below.

4. COSINE SIMILARITY

Cosine similarity is a mathematical metric used to measure the similarity between two non-zero vectors in a multi-dimensional space. It is commonly employed in various fields, including information retrieval, natural language processing, machine learning, and data mining, to determine how closely related two vectors are based on the cosine of the angle between them.
Cosine similarity is particularly useful for text and document analysis, as well as for recommendations and clustering tasks. Specifically, it calculates the cosine of the angle θ between the two vectors. The formula for cosine similarity is as follows:
Cosine Similarity (A, B) = (A · B) / (||A|| * ||B||)
Where:
A · B represents the dot product of vectors A and B.
||A|| and ||B|| represent the Euclidean norms (magnitudes) of vectors A and B, respectively.
Cosine similarity yields a value between -1 and 1. The range is interpreted as follows:
A cosine similarity of 1 indicates that the vectors are identical (i.e., they point in the same direction).
A cosine similarity of 0 suggests that the vectors are orthogonal (i.e., they are unrelated).
A cosine similarity of -1 means that the vectors are opposed (i.e., they point in opposite directions).

5. KEY FEATURES

Keyword Tokenization: The system tokenizes user queries and indexed documents to extract and identify important keywords. Tokenization ensures that each word is processed as a separate entity, making it easier to calculate cosine similarity.

Cosine Similarity Calculation: The system uses the cosine similarity algorithm to measure the similarity between the keywords in the user's query and the keywords in the indexed documents. This feature provides a quantitative measure of relevance.

Semantic Search: The system incorporates semantic search techniques to understand the context and semantics of user queries. This feature enables the recognition of synonyms, related concepts, and context, contributing to more context-aware search results.

Contextual Ranking: The system ranks search results based on their cosine similarity scores, ensuring that documents closely related to the user's intent are presented at the top. This contextual ranking enhances the relevance of the search results.

Stop Word Removal: Stop words, such as "and," "the," and "in," are removed from user queries to improve the accuracy of the search process and reduce noise.

Query History and Personalization: The system may offer personalized search results based on a user's search history and preferences, enhancing the user experience.

6. OUTPUT

Fig: Interface of keyword Based Search Engine

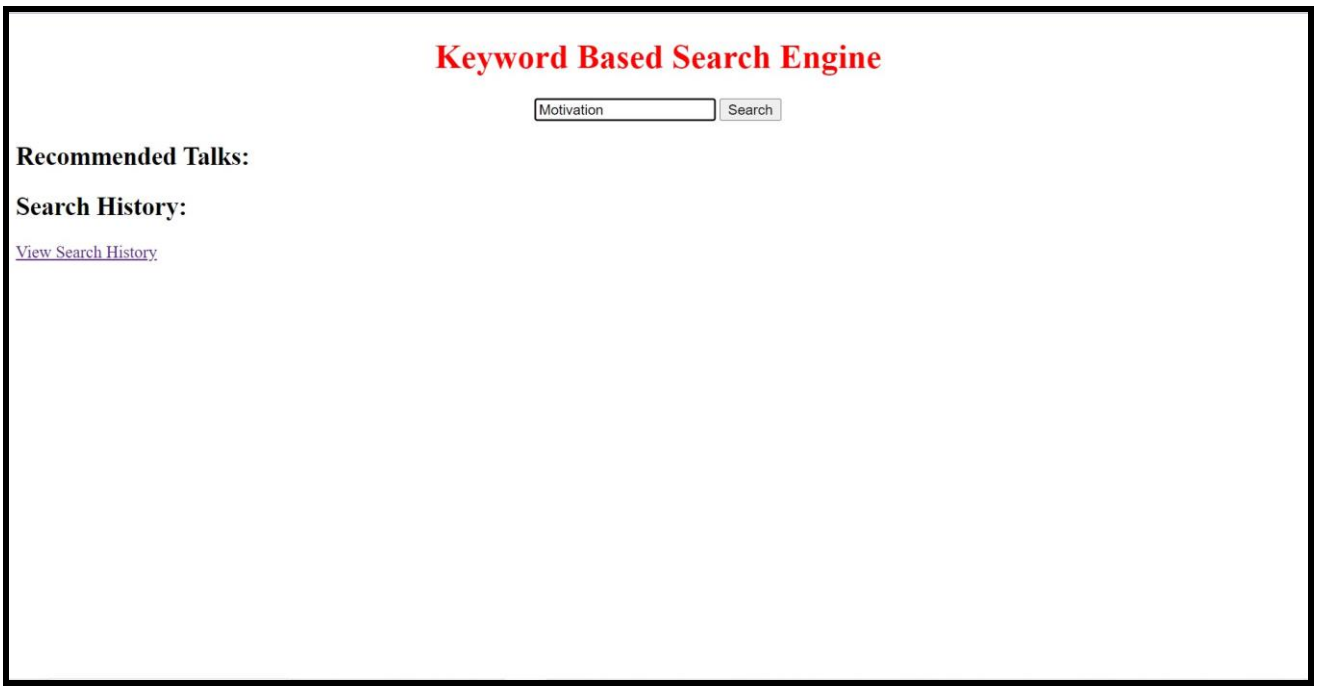

Fig: In the search engine, we have entered the keyword 'Motivation'

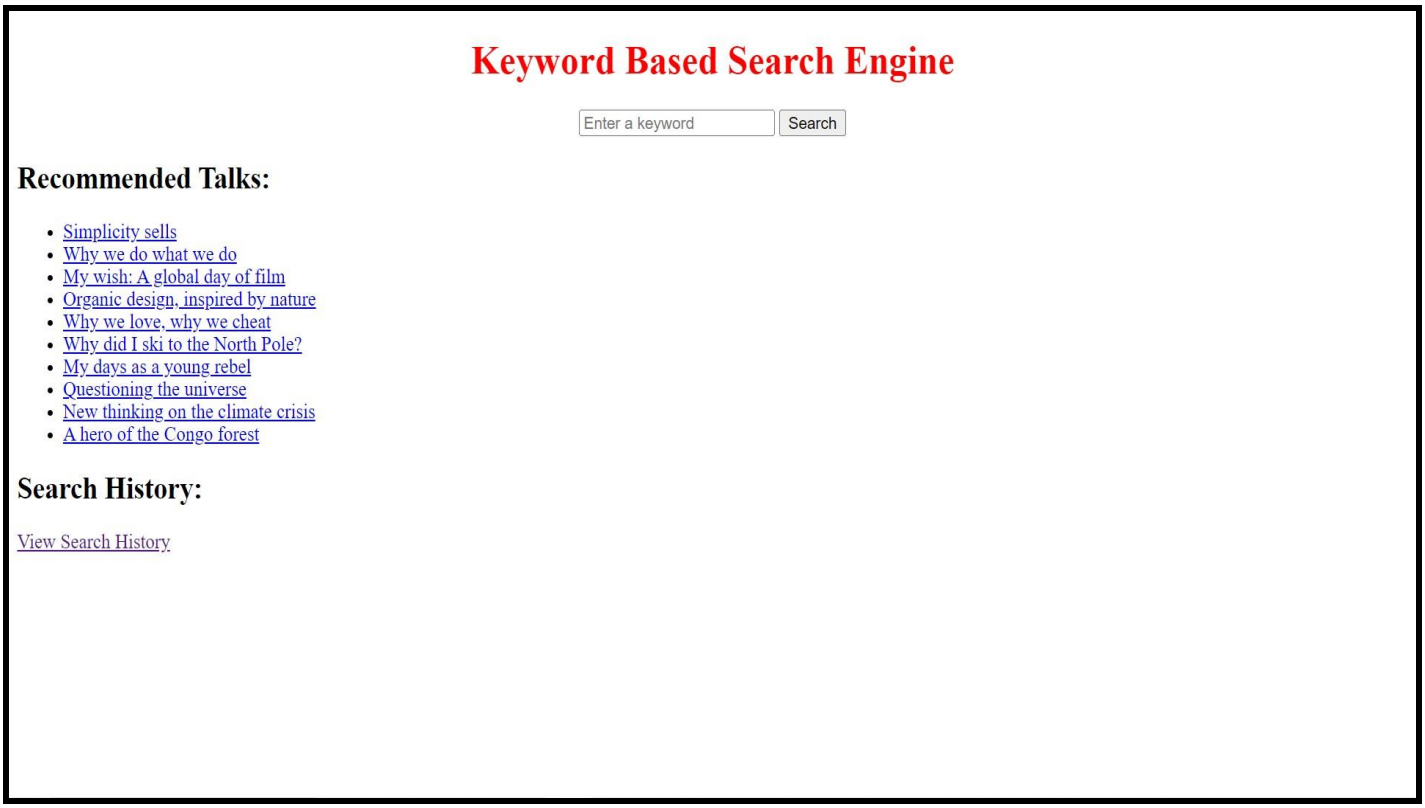

Fig: We get the links of the videos which are related to Motivation

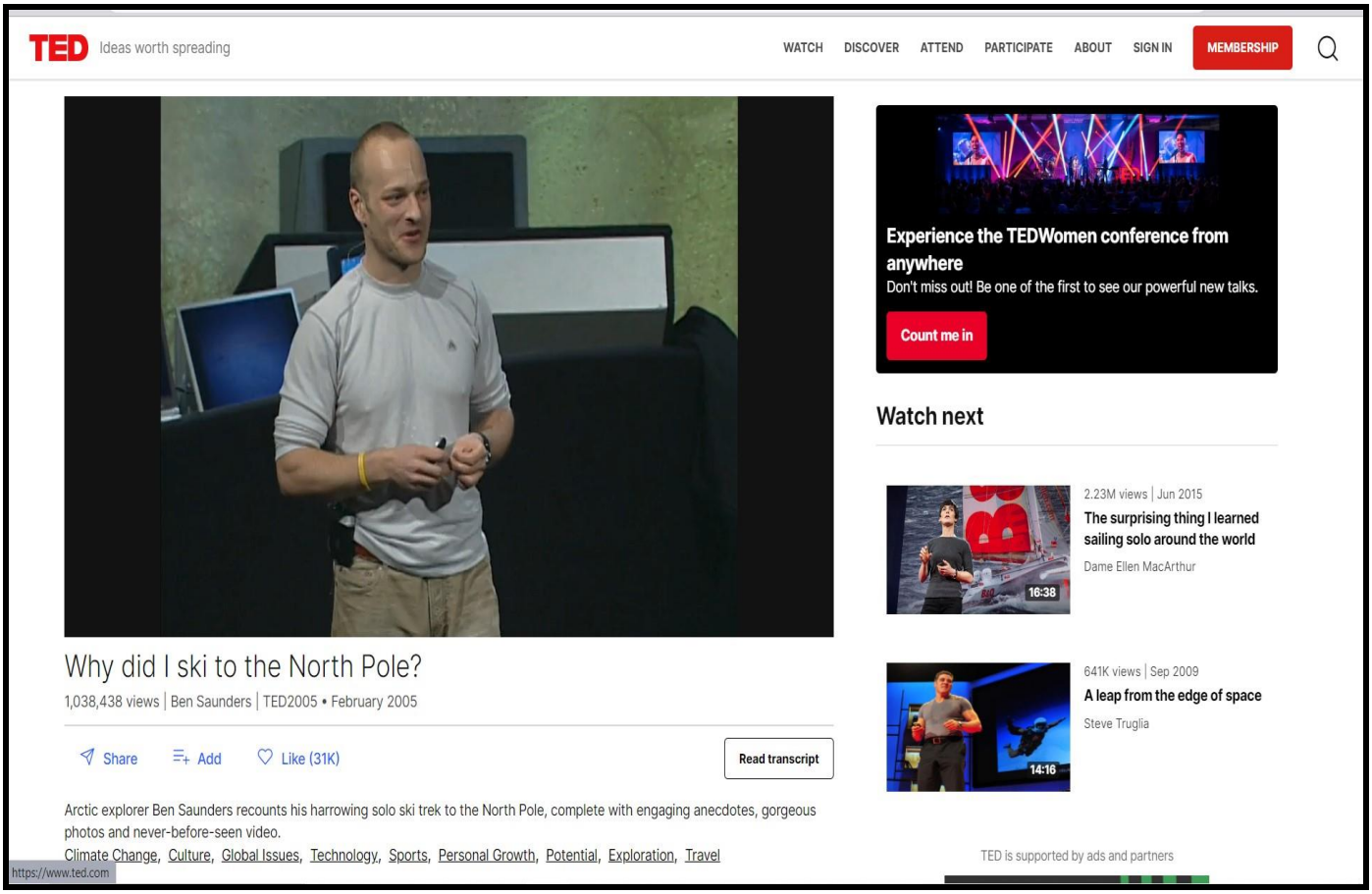

Fig: By clicking the link, we will be directed to the video.

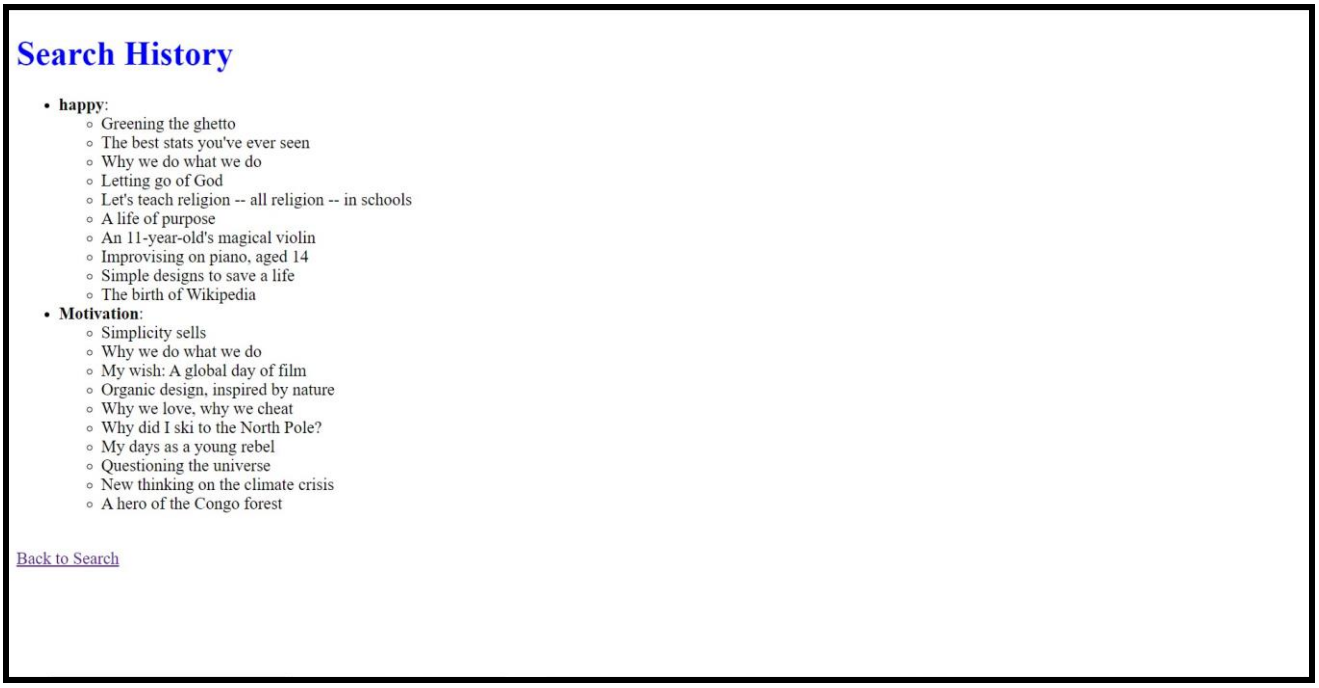

Fig: Here we can find the browsed history

7. FUTURE ENHANCEMENTS

Mobile-Friendly Design: Optimize the search engine's interface and functionality for mobile devices, ensuring a seamless and user-friendly experience on smartphones and tablets.

User Interface (UI) Customization: Allow users to customize the appearance and layout of the search interface to suit their preferences and needs.

Voice Search: Implement voice recognition and search capabilities, allowing users to perform searches using voice commands.

8. CONCLUSION

The "Keyword-Based Search Engine Using Cosine Similarity" project represents a significant advancement in the realm of information retrieval and search technology. By introducing an innovative approach to search engines, this project aims to overcome the limitations of traditional keyword-based systems and provide users with more contextually relevant and precise search results.

Through the implementation of cosine similarity, vector space models, and semantic search techniques, the project enhances the understanding of user queries, effectively reducing keyword ambiguity and improving the accuracy of search results. It offers real-time indexing, adaptability to large datasets, and scalability, ensuring that the system remains up-to-date and efficient in an ever-expanding digital landscape.

9. REFERENCES

[1 Recommendation system using Collaborative Filtering
https://www.researchgate.net/publication/373992678_Recommendation_System_Using_Collaborative_Filtering
[2] Movie Recommendation System Using Collaborative Filtering
https://www.researchgate.net/publication/354963619_Movie_Recommendation_System_by_Using_Collaborative_Filtering

