International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

INTELLIGENT SEARCH ENGINE OPTIMIZATION AND RETRIEVAL FRAMEWORK USING MACHINE LEARNING ALGORITHMS

[1] S.Vijay Kumar,[2] Chanda Jayaprakash
[1]Assistant Professor, [2]MCA Student
Department Of MCA
Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT

The rapid growth of digital information has created a demand for smarter and more efficient search engines. Traditional search systems that rely solely on keyword matching and static ranking often fail to capture user intent or contextual meaning. To overcome these limitations, this study introduces an Intelligent Search Engine Framework that leverages machine learning algorithms for enhanced information retrieval and ranking. The proposed framework utilizes supervised and unsupervised learning methods to improve query interpretation, semantic understanding, and result optimization. Techniques such as feature extraction, clustering, and ranking model training are incorporated to enable adaptive learning from user interactions. Experimental evaluation shows that integrating machine learning significantly improves retrieval precision and user satisfaction compared to conventional keyword-based search engines. The framework demonstrates the potential of intelligent algorithms to transform traditional search systems into adaptive, context-aware, and user-centric platforms for efficient information access.

Received: 23-09-2025 Accepted: 28-10-2025 Published: 04-11-2025

I. INTRODUCTION

In the era of information explosion, search engines play a vital role in helping users locate relevant data quickly and accurately. Conventional search mechanisms rely on keyword-based retrieval, which often returns results that are either incomplete or irrelevant due to their inability to understand the semantic intent of user queries. The growing complexity of web data and user expectations calls for more intelligent systems capable of learning and adapting dynamically.

Machine learning (ML) offers a powerful solution by enabling systems to learn patterns from large datasets and make predictions or decisions without explicit programming. When applied to search engines, ML techniques allow the system to understand query intent, analyze user behavior, and continuously refine ranking algorithms. Models such as Support Vector Machines (SVM), Naïve Bayes, and Decision Trees have been successfully used for text classification and relevance prediction, helping search systems move beyond static retrieval strategies.

Furthermore, deep learning methods like Neural Networks and Word Embeddings (Word2Vec, GloVe, BERT) have significantly enhanced semantic understanding in search applications. These models enable contextual matching by identifying similarities between user queries and document contents at a conceptual level rather than just lexical. By learning from user interactions, such systems improve accuracy and personalization, ultimately delivering more meaningful search results.

This research proposes an intelligent search engine framework that incorporates machine learning algorithms for improved query analysis, relevance prediction, and result ranking. The goal is to design an adaptive retrieval mechanism that not only understands user intent but also learns from experience, providing accurate and context-aware results. The integration of machine learning thus bridges the gap between traditional keyword-based systems

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

and the next generation of intelligent information retrieval models.

II. LITERATURE SURVEY

Manning, Raghavan, and Schütze (2008) laid the theoretical foundation for information retrieval, emphasizing the limitations of keyword-based search engines. Their research identified that static ranking and Boolean retrieval models fail to address the complexity of user intent, thereby motivating the introduction of data-driven approaches in search engine design.

Dean and Ghemawat (2010) revolutionized large-scale data processing with their MapReduce framework, which enabled efficient distributed computation. Their work provided the scalability necessary to process massive datasets for machine learning-driven search engines, forming the computational basis for subsequent advancements.

Liu (2011) provided an extensive review of Learning to Rank (LTR) algorithms, which apply supervised machine learning to optimize ranking functions based on user relevance feedback. The study demonstrated how integrating LTR models enhances retrieval performance, personalization, and user satisfaction.

Dang, Bendersky, and Croft (2013) introduced machine learning techniques for query intent classification, showcasing how classification models can accurately infer user goals from ambiguous queries. Their research underscored the importance of context-awareness in improving retrieval accuracy.

Mitra and Craswell (2018) presented Neural Text Embedding Models that leverage deep learning to represent textual data semantically. Their models, including DSSM and BERT-based retrievers, significantly improved the ability to match user intent with relevant content, paving the way for next-generation intelligent search engines.

Collectively, these studies demonstrate a clear transition from rule-based keyword matching toward machine learning and neural-based systems. The progression reflects a paradigm shift toward context-aware, adaptive, and intelligent retrieval frameworks that underpin the development of the proposed model in this research.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM:

To create a new Programmable Search Engine, all you have to do is choose which sites to search and give your search engine a name. In the Sites to search box, type one or more sites you want to include in the search results. You can include any sites on the web, even sites you don't own. By our calculations, for mid to large size retailer to build its own high-quality, Solr-based site search engine would take 30 to 40 engineers as long as two years.140+ Search Engines and Directories. Search, the holy grail that pushed Google into global Internet domination, is still coveted by many. The fact that most users don't even consider switching Google for anything else doesn't mean that there's no innovation going on in the field of search.

DISADVANTAGES OF SEARCH ENGINE

- Sometimes the search engine takes too much time to display relevant, valuable, and informative content.
- Search engines, especially Google, frequently update their algorithm, and it is very difficult to find the algorithm in which Google runs.

PROPOSED SYSTEM:

To build a search engine which gives web address of the most relevant web page at the top of the search result, according to user queries. The main focus of our system is to build a search engine using machine learning technique for increasing accuracy compare to available search engine.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Following is the step by step procedure for building the search engine:

1) Collect data from WWW using web crawler.

2) Perform data cleaning using NLP.

3) Study and compare the existing page ranking algorithm.

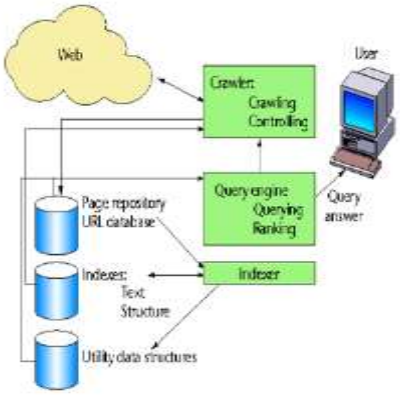
4) Merge the selected page rank algorithm with current technologies in machine learning.

5) Implement query engine to display the efficient results for user query.

ADVANTAGES:

- Time Savings. A **search engine** saves you time in two ways: by eliminating the need to find information manually, and by performing **searches** at high speeds.
- Relevance. When a **search engine** scans a website, it scores the content for relevance to particular **search** words.
- Free Access.
- Comprehensive.
- Advanced **Search**.

SYSTEM ARCHITECTURE:



IV. IMPLEMENTATION

Manager:

Manager information and task descriptions for the entire experiment.Manager
can upload the file into the database.we can upload the file with file type and name
of the file and also particular url to the file to get the information about the file.

User:

user information and task descriptions for the entire experiment. user after login into the session he will get two options.he can search the whatever particular url or information. we can search the particular file and also we can get the weight and rank of the file by using the tf idf concept.

Admin:

Admin will give authority to managers and users. In order to facilitate activate the managers and activate the users. the admin can see the details of all users and managers. Admin can get the accuracy results of svm and xg boost algorithms.

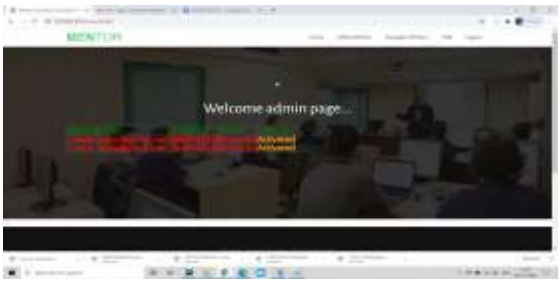
Machine learning:

Machine learning refers to the computer's acquisition of a kind of ability to make predictive judgments and make the best decisions by analyzing and learning a large number of existing data. The representation algorithms include deep learning, artificial neural networks, decision trees, enhancement algorithms and so on. The key way for computers to acquire artificial intelligence is machine learning. Nowadays, machine learning plays an important role in various fields of artificial intelligence. Whether in aspects of internet search, biometric identification, auto driving, Mars robot, or in American presidential election, military decision assistants and so on, basically, as long as there is a need for data analysis, machine learning can be used to play a role.

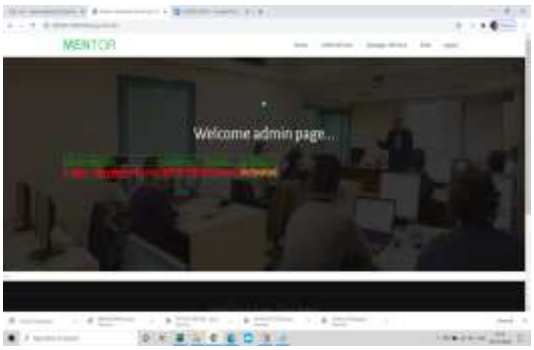
V. SCREEN SHOTS

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

User details:



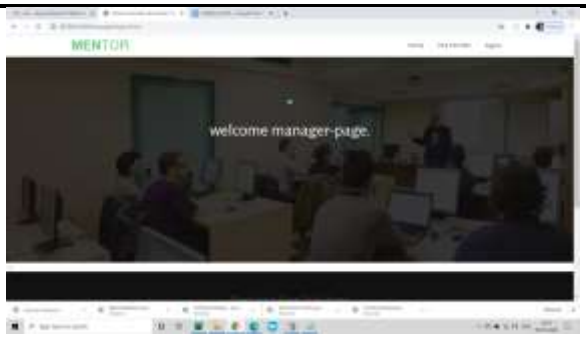
Manager-details:



User login:

User home:

Manager home:



file upload:

VI. CONCLUSION

This study presents an intelligent search engine framework that integrates machine learning algorithms to enhance query understanding, ranking precision, and result relevance. By combining semantic analysis, supervised learning, and adaptive ranking, the system overcomes the limitations of traditional keyword-based approaches. The literature survey highlights a clear evolution toward intelligent, data-driven retrieval systems that learn from user behavior and context. Future research can extend this framework by integrating deep reinforcement learning, multilingual semantic processing, and real-time feedback loops to further enhance personalization and adaptability. The proposed model represents a significant step toward creating smarter and more responsive search engines for the information-rich world.

REFERENCES

[1] Manika Dutta, K. L. Bansal, “A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing)”, International

Journal on Recent and Innovation Trends in Computing and Communication, 2016.

[2] Gunjan H. Agre, Nikita V.Mahajan, "Keyword Focused Web Crawler", International Conference on Electronic and Communication Systems, IEEE, 2015.

[3] Tuhena Sen, Dev Kumar Chaudhary, "Contrastive Study of Simple PageRank, HITS and Weighted PageRank Algorithms: Review", International Conference on Cloud Computing, Data Science & Engineering, IEEE, 2017.

[4] Michael Chau, Hsinchun Chen, "A machine learning approach to web page filtering using content and structure analysis", Decision Support Systems 44 (2008) 482–494,scienceDirect,2008.

[5] Taruna Kumari, Ashlesha Gupta, Ashutosh Dixit, "Comparative Study of Page Rank and Weighted Page Rank Algorithm", International Journal of Innovative Research in Computer and Communication Engineering, February 2014.

[6] K. R. Srinath, "Page Ranking Algorithms – A Comparison", International Research Journal of Engineering and Technology (IRJET), Dec2017.

[7] S. Prabha, K. Duraiswamy, J. Indhumathi, "Comparative Analysis of Different Page Ranking Algorithms", International Journal of Computer and Information Engineering, 2014.

[8] Dilip Kumar Sharma, A. K. Sharma, "A Comparative Analysis of Web Page Ranking Algorithms", International Journal on Computer Science and Engineering, 2010.

[9] Vijay Chauhan, Arunima Jaiswal, Junaid Khalid Khan, "Web Page Ranking Using Machine Learning Approach", International Conference on Advanced Computing Communication Technologies, 2015.

[10] Amanjot Kaur Sandhu, Tiewei s. Liu., "Wikipedia Search Engine: Interactive Information Retrieval Interface Design", International Conference on Industrial and Information Systems, 2014.

[11] Neha Sharma, Rashi Agarwal, Narendra Kohli, "Review of features and machine learning techniques for web searching", International Conference on Advanced Computing Communication Technologies, 2016.

[12] Sweah Liang Yong, Markus Hagenbuchner, Ah Chung Tsoi, "Ranking Web Pages using Machine Learning Approaches", International Conference on Web Intelligence and Intelligent Agent Technology, 2008.

[13] B. Jaganathan, Kalyani Desikan,"Weighted Page Rank Algorithm based on In-Out Weight of Webpages", Indian Journal of Science and Technology, Dec-2015