International Journal of Core Engineering & Management
Volume-7, Issue-06, 2023 ISSN No: 2348-9510

ISEARCH: ENHANCED INTELLIGENT SEARCH BASED ON A QUERY USING LLMS

Praneeth Vadlapati
Independent researcher, India
praneeth.vad@gmail.com

Abstract

The growing digital landscape is occupied by vast amounts of content, making it increasingly challenging to find relevant information efficiently. Search engines are powerful tools to retrieve relevant content. However, the results from search engines are often manipulated by Search Engine Optimization (SEO) spammers who prioritize their non-relevant content for financial gains. This paper presents iSearch, an intelligent search system that utilizes Natural Language Processing (NLP) through Large Language Models (LLMs) to generate relevant queries that could be utilized to find relevant content. LLMs are utilized to filter the search results to include the most relevant results. The system is designed to address the challenges associated with discovering relevant content. By generating queries dynamically and ranking the search results based on relevance, iSearch creates a user-centered content discovery experience. The proposed solution demonstrates significant advantages over current search paradigms by offering a scalable, intention-aware, and context-aware search mechanism resistant to SEO spam. The system successfully demonstrated its ability to generate relevant search queries to uncover the relevant unpopular content for each search query from a user. The source code is available at github.com/Pro-GenAI/iSearch.

Keywords: Large Language Models (LLMs), Natural Language Processing (NLP), content discovery, intelligent search, semantic search, context-aware decision making, Search Engine Optimization (SEO)

I. INTRODUCTION

A vast amount of content is generated across the web every second [1], [2]. The high volume of content poses opportunities and challenges for the consumers of the content. Users rely on search engines to navigate vast amounts of information and access relevant content [3], [4]. Search engines are equipped with ranking algorithms [5], [6], which retrieve and rank content based on relevance. Since their inception, search engines have become indispensable tools for information retrieval [7]. However, they rely on utilizing keywords. Search engines often produce irrelevant results influenced by SEO strategies [8], [9].

A. Disadvantages with current approaches

Search engines utilized by the users prioritize the content optimized with Search Engine Optimization (SEO) keywords [10], [11] instead of only the relevance. Valuable content often remains hidden from the audience due to competition from content created by SEO spammers who utilize numerous keywords in content that is of lesser relevance to the query. The current approaches to using search engines are vulnerable to SEO manipulation and poisoning [12], [13]. The current approaches use a limited amount of semantic understanding and focus on utilizing keywords to

perform search operations. The competition from irrelevant results can frustrate users and contribute to decision fatigue and information overload in the current digital landscape.

B. Proposed system and its benefits

This paper introduces iSearch, a system designed to bridge the gap between user intent and effective content discovery. The system understands the user queries and intentions at the semantic level to leverage paraphrasing techniques and filter results to finally sort them according to relevance to the user query. The proposed system offers the potential for a transformative search experience. The system addresses the limitations of existing methodologies through an innovative strategy that improves conventional search techniques through sophisticated query processing and filtering. The system leverages user input and paraphrased variations of the query to broaden the search scope and consolidates results into a refined output. The results are filtered and sorted by an LLM according to relevance. This approach mitigates the impact of SEO spam by prioritizing user intent over matching the keywords. The system enables personalized content discovery based on the needs of users. The system allows access to relevant content that is underrepresented, such as independent creators or less-publicized academic papers. Considering the multilingual capabilities of LLMs and modern search engines currently and in the future, the system has the potential to operate across multiple languages.

C. Related work

Query Expansion Techniques [14], [15] focused on augmenting user queries by generating related items to improve recall. However, the technique lacks a focus on user intent and selecting relevant results from the search results. The technique lacks challenges that are related to highly specialized fields. Efforts to filter search results have utilized clustering algorithms [16], [17] to eliminate duplicates or prioritize high-quality results. However, the technique is not designed to understand user intent and might retain irrelevant results. Paraphrasing models [18], [19] are being used to rephrase a query multiple times to obtain numerous results. However, the technique lacks filtering and sorting relevant results based on user intent. Existing efforts in intelligent search enhancement often involve advancements in Natural Language Processing (NLP). Recent advancements in LLMs based on transformer architecture [20] created the foundation of the iSearch system. To address the gaps in the existing work, iSearch integrates advancements from these domains into a robust user-centric system that is capable of understanding the user intent based on the query, generates relevant queries, fetches and consolidates the results to filter them, and returns relevant results.

II. METHODS

A. Selecting and loading an LLM

An LLM is required in this step to interpret user queries and generate relevant queries that potentially yield relevant results. An LLM is selected based on the capabilities of understanding natural language and following the custom instructions mentioned in a prompt. Hence, GPT-3.5 [21] is used for the experiment based on its proven capabilities. The LLM is utilized to generate semantically diverse queries to ensure a high level of variability in search results while maintaining alignment of the results with the intent of the user.

B. Creating user queries and generating relevant queries

A user query is required to generate relevant queries during the experiment. Hence, sample user

queries are created. The selected LLM is employed to generate related queries based on the original user query by understanding the user intent. Five queries are generated from each user query to ensure semantic diversity in queries.

C. Query execution and de-duplication

The generated queries are executed to fetch results. Google [22], [23] is the search engine utilized for query execution. The search engine is selected based on its performance, efficiency, and scalability. The search engine is utilized to return the ten most relevant results for each query to focus on the relevance of the results instead of utilizing numerous results that get returned for each query. All the search results from rephrased queries are consolidated. Duplicate results are eliminated from the search results to prevent redundancy across the results.

D. Relevance filtering and sorting

The LLM is employed to analyze the search results based on queries and return the filtered results in a sorted form to ensure the results are sorted based on the relevance to the user intent. The LLM is provided with the page title, URL, and snippet to support the decision process in filtering and sorting the results.

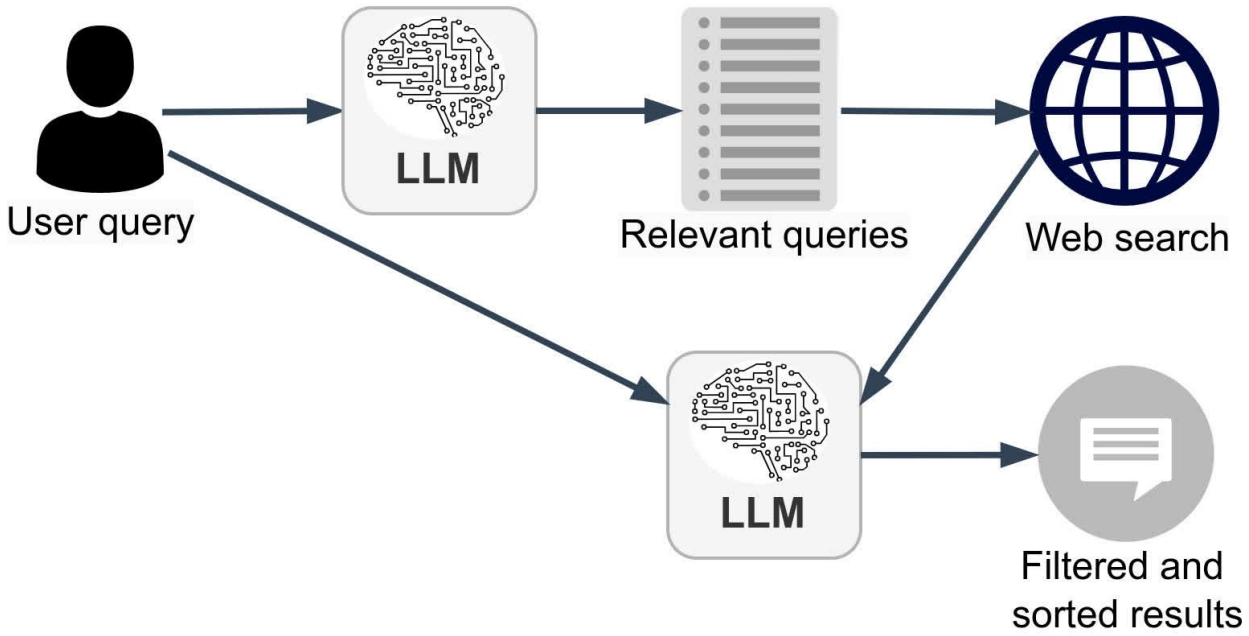

Fig. 1. Workflow of iSearch

III. RESULTS

A. Generated relevant queries

Relevant queries were successfully generated based on original queries. Each original query and newly generated queries are mentioned in the table below.

TABLE I. GENERATED QUERIES

Index	User query	Generated query
1	comedy movies by Charlie Chaplin	Charlie Chaplin comedy movie list
		What are the best comedy movies by Charlie Chaplin
		Where can I watch Charlie Chaplin's comedy movies
		Charlie Chaplin's most famous comedy movies
		Classic comedy movies starring Charlie Chaplin
2	latest ML research papers	cutting-edge machine learning research papers
		most recent ML research publications
		top academic papers on machine learning
		newest developments in machine learning research
		cutting-edge advancements in ML research papers

B. Sample results for the queries

The queries generated in the earlier step were utilized to retrieve results using the selected search engine. The sample results of the queries are mentioned in the table below.

TABLE II. SAMPLE RESULTS FOR USER QUERIES

Index	Page titles of sample results
1	charlie chaplin best movies | IMDB
	The Comedies of Charlie Chaplin
	Charlie Chaplin filmography
2	Papers With Code: The latest in Machine Learning
	Journal of Machine Learning Research
	Machine Learning | arXiv

C. De-duplicated results

The number of results before and after the de-duplication process is mentioned in the table below.

TABLE III. RESULT COUNTS FOR QUERIES

Index	Results of user query	Results of generated queries	Results of generated queries after de-duplication
1	7	41	25
2	10	46	38

D. Result filtering and sorting

The new results were discovered because of the new method of iSearch. Numerous relevant results

International Journal of Core Engineering & Management
Volume-7, Issue-06, 2023 ISSN No: 2348-9510

returned by this system were not returned using the original query. Results have been filtered and sorted by the LLM using custom instructions based on original queries. The samples of filtered and sorted results are mentioned in the table below.

TABLE IV. SAMPLE OF FILTERED RESULTS FOR USER QUERIES

Index	Page titles of filtered results
1	charlie chaplin best movies | IMDB
	Charlie Chaplin : Films
	Watch Six Charlie Chaplin Films in Our Virtual Silent Movie ...
2	**Cutting-Edge Research in AI and Healthcare**
	Papers With Code: The latest in Machine Learning
	Apple Machine Learning Research: Overview

IV. DISCUSSION

The system has successfully demonstrated its potential to enhance search systems by outperforming competitors driven by SEO spamming. The focus on user intent and leveraging LLMs has allowed the system to discover hidden content that rarely gets accessed using original queries used by the users. However, the challenges of the new approach include improving the instructions to optimize the prompts. Another challenge involves the increased cost and latency introduced by the utilization of LLMs. Although LLM-driven filtering introduces latency, this challenge could be mitigated by optimizing inference through processes such as edge computing and the usage of LLMs that are of smaller size. A potential improvement of the system includes fine-tuning the LLMs to specialize in multiple domains. The system faces a risk of potential biases present in the pre-trained LLMs that require rigorous testing and consideration.

V. CONCLUSION

In the current digital landscape that is dominated by SEO spammers, iSearch represents a significant step in intelligent intent-based search technology. The consolidation of results from paraphrased queries and relevance-driven ranking facilitate meaningful and efficient content discovery for end-users. The system addresses the challenges that exist in the search domain, such as SEO manipulation, redundant content, and limited semantic understanding of traditional search engines. The ability of the system to integrate query generation and result filtering demonstrates the immense potential of LLMs in enhancing content discovery. Beyond its resistance to SEO spam, iSearch provides opportunities to highlight relevant underrepresented content to the users. With continued refinement and interdisciplinary collaboration, the iSearch system has the potential to redefine the search experience, offering a robust solution to challenges posed by the ever-expanding digital ecosystem. Future research could make LLMs faster and more cost-effective, which could make this system faster and more viable for implementation in existing search applications. To fully realize the potential of iSearch, collaborative efforts between academia and industry are essential, paving the way for a new era of intention-aware search technologies.

APPENDIX

```
Generate five relevant search queries based on the user
query: "{query}"
Respond with each query in separate lines inside triple
backticks.
Generate 5 such search queries.
Don't include numbers or bullet points in the response.
Sample response:
```
is ant a mammal
is ant an insect
is ant a mammal or insect
```
```

Figure A1. Prompt template to generate relevant queries

```
Search results:
{results}
---
Filter these results for relevance to: "{query}"
Sort by relevance.
Respond with each result in separate lines inside triple
backticks.
Return only URLs of relevant pages and nothing else.
I want to find new content
Sample response:
```
https://example.com/wiki/Ant
https://www.example.com/animals/invertebrates/fac
ts/ants
```
```

Figure A2. Prompt template to filter and sort the results

REFERENCES

1. Bernard Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," Forbes. Accessed: Dec. 04, 2022. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/
2. Jeff Desjardins, "How much data is generated each day?," World Economic Forum. Accessed: Dec. 04, 2022. [Online]. Available: https://www.weforum.org/stories/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/
3. Y. Wang, L. Wu, L. Luo, Y. Zhang, and G. Dong, "Short-term Internet search using makes people rely on search engines when facing unknown issues.," PLoS One, vol. 12, no. 4, p. e0176325, Apr. 2017, doi: 10.1371/journal.pone.0176325.
4. A. Haris and S. Sarijan, "A Study On Online Search By People Using Search Engine," May

2017. [Online]. Available: https://www.researchgate.net/publication/328527875_A_Study_On_Online_Search_By_People_Using_Search_Engine
5. E. Roberts, "The Google PageRank Algorithm," Nov. 2016, Stanford University. [Online]. Available: https://web.stanford.edu/class/cs54n/handouts/24-GooglePageRankAlgorithm.pdf
6. M. A. Hingoro and H. Nawaz, "A Comparative Analysis of Search Engine Ranking Algorithms," International Journal of Advanced Trends in Computer Science and Engineering, vol. 10, no. 2, pp. 1247 – 1252, Apr. 2021, doi: 10.30534/ijatcse/2021/1081022021.
7. S.-S. Liaw and H.-M. Huang, "Information retrieval from the World Wide Web: a user-focused approach based on individual experience with search engines," Computers in Human Behavior, vol. 22, no. 3, pp. 501–517, May 2006, doi: 10.1016/j.chb.2004.10.007.
8. T. Jones, D. Hawking, P. Thomas, and R. Sankaranarayana, "Relative effect of spam and irrelevant documents on user interaction with search engines," in Proceedings of the 20th ACM International Conference on Information and Knowledge Management, in CIKM '11. New York, NY, USA: Association for Computing Machinery, Oct. 2011, pp. 2113–2116. doi: 10.1145/2063576.2063903.
9. N. L. Wilczynski, K. A. McKibbon, and R. B. Haynes, "Search filter precision can be improved by NOTing out irrelevant content.," AMIA Annu Symp Proc, vol. 2011, pp. 1506–1513, Oct. 2011.
10. M. Rahman, "Search Engines Going beyond Keyword Search: A Survey," International Journal of Computer Applications, vol. 75, no. 17, pp. 1–8, Aug. 2013, doi: 10.5120/13200-0357.
11. A. Malve and P. Chawan, "A Comparative Study of Keyword and Semantic based Search Engine," International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 11, pp. 11156–111614, Nov. 2015, doi: 10.15680/IJIRSET.2015.0411039.
12. "SEO Poisoning – A Gootloader Story," The DFIR Report. Accessed: Dec. 04, 2022. [Online]. Available: https://thedfirreport.com/2022/05/09/seo-poisoning-a-gootloader-story/
13. J. Bar-Ilan, "Manipulating search engine algorithms: the case of Google," Journal of Information, Communication and Ethics in Society, vol. 5, no. 2/3, pp. 155–166, Oct. 2007, doi: 10.1108/14779960710837623.
14. L. Afuan, A. Ashari, and Y. Suyanto, "A New Approach in Query Expansion Methods for Improving Information Retrieval," JUITA: Jurnal Informatika, vol. 9, p. 93, May 2021, doi: 10.30595/juita.v9i1.9657.
15. H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: A survey," Information Processing & Management, vol. 56, no. 5, pp. 1698–1735, Sep. 2019, doi: 10.1016/j.ipm.2019.05.009.
16. G. Mecca, S. Raunich, and A. Pappalardo, "A new algorithm for clustering search results," Data & Knowledge Engineering, vol. 62, no. 3, pp. 504–522, Sep. 2007, doi: 10.1016/j.datak.2006.10.006.
17. M. Jan, S. Khusro, I. Alam, I. Khan, and B. Niazi, "Interest-Based Content Clustering for Enhancing Searching and Recommendations on Smart TV," Wireless Communications and Mobile Computing, vol. 2022, no. 1, p. 3896840, Jan. 2022, doi: 10.1155/2022/3896840.
18. Y. Fu, Y. Feng, and J. P. Cunningham, "Paraphrase generation with latent bag of words," in Proceedings of the 33rd International Conference on Neural Information Processing Systems,

Red Hook, NY, USA: Curran Associates Inc., 2019.
19. E. Bandel, R. Aharonov, M. Shmueli-Scheuer, I. Shnayderman, N. Slonim, and L. Ein-Dor, "Quality Controlled Paraphrase Generation," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 596–609. doi: 10.18653/v1/2022.acl-long.45.
20. A. Vaswani et al., "Attention is all you need," in Proceedings of the 31st International Conference on Neural Information Processing Systems, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 6000–6010. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
21. "Introducing ChatGPT," OpenAI. Accessed: Dec. 04, 2022. [Online]. Available: https://openai.com/index/chatgpt/
22. "Google [Data source]." Accessed: Dec. 04, 2022. [Online]. Available: https://www.google.com/
23. "SerpAPI [Data source]." Accessed: Dec. 04, 2022. [Online]. Available: https://serpapi.com/search-api