Anonymous Information Retrieval for News

Dwaipayan Mondal || Parnita Bokade || Manish Kumar || Piyush Kumar Saini || Pragati Agrawal

What is the problem identified in the project?

This proposal introduces a novel news aggregation platform that prioritizes user privacy while delivering personalized, up-to-date news content. Unlike traditional aggregators, our system employs dynamic web crawling, advanced natural language processing (NLP), and machine learning to offer a tailored news experience without storing identifiable user data.
This initiative addresses the growing demand for secure, reliable information sources that respect user anonymity.

Why is this problem important?

Our objective is to develop a news aggregation platform that enhances user experience through personalized content delivery, without the need for storing personal information. This approach not only safeguards privacy but also introduces a dynamic and adaptable news retrieval system. There is no platform that transforms the way users engage with news, providing instant explanations for terms and topics encountered during reading.

How different is your idea from others?

-Our proposal diverges from existing solutions by introducing a dynamic web crawling mechanism that continuously identifies and integrates recent news sources, ensuring comprehensive and up-to-date news coverage.
-Furthermore, we propose the use of advanced NLP techniques for real-time summarization of news articles, providing users with concise and informative summaries. Unlike existing systems that offer static personalization, our solution will adapt to evolving user preferences using machine learning algorithms.
-Our platform makes news more understandable with quick explanations for terms and topics users come across while reading, turning confusion into clarity.

Is there any related work?

- Information Retrieval Systems [1-5] focus on organized presentation, cross-lingual recommendations, event detection, inverted index construction, and personalized news search. However, these systems lack a comprehensive approach to privacy and dynamic personalization.

Proposed Techniques and Algorithms

Our methodology encompasses:
- Dynamic Web Crawling for real-time source updates.
- NLP for efficient summarization.
- Machine Learning for evolving personalization.
- Duplicate Detection and Clustering to enhance content delivery.

Evaluation Methodology

Evaluation will focus on:
- Coverage and Timeliness:Ensuring broad, up-to-date news representation.
- Summarization Quality:Delivering concise, informative summaries.
- Personalization Efficiency:Tailoring content to user interests effectively.
- User Satisfaction:Measured through studies and feedback on usability and content relevance.

Contribution

Frontend- Manish Kumar || Dwaipayan Mondal
Backend- Parnita Bokade || Pragati Agrawal
NLP related - Piyush Kumar Saini

References

- [1] Information Retrieval and Processing System...
- [2] A Semantic Content-Based Recommendation...
- [3] Challenges and Issues on Online News Management...
- [4] Map Reduce Programming Model...
- [5] Personalized News Search in WWW...