Intellimark - Intelligent Bookmark Aggregator

Anirudh Ragavender Ramarathnam Madhavan
ar80@illinois.edu

Krishna Anandan Ganesan
kag8@illinois.edu

Saairam Venkatesh
saairam2@illinois.edu

Sujithra Rajan
rajan11@illinois.edu

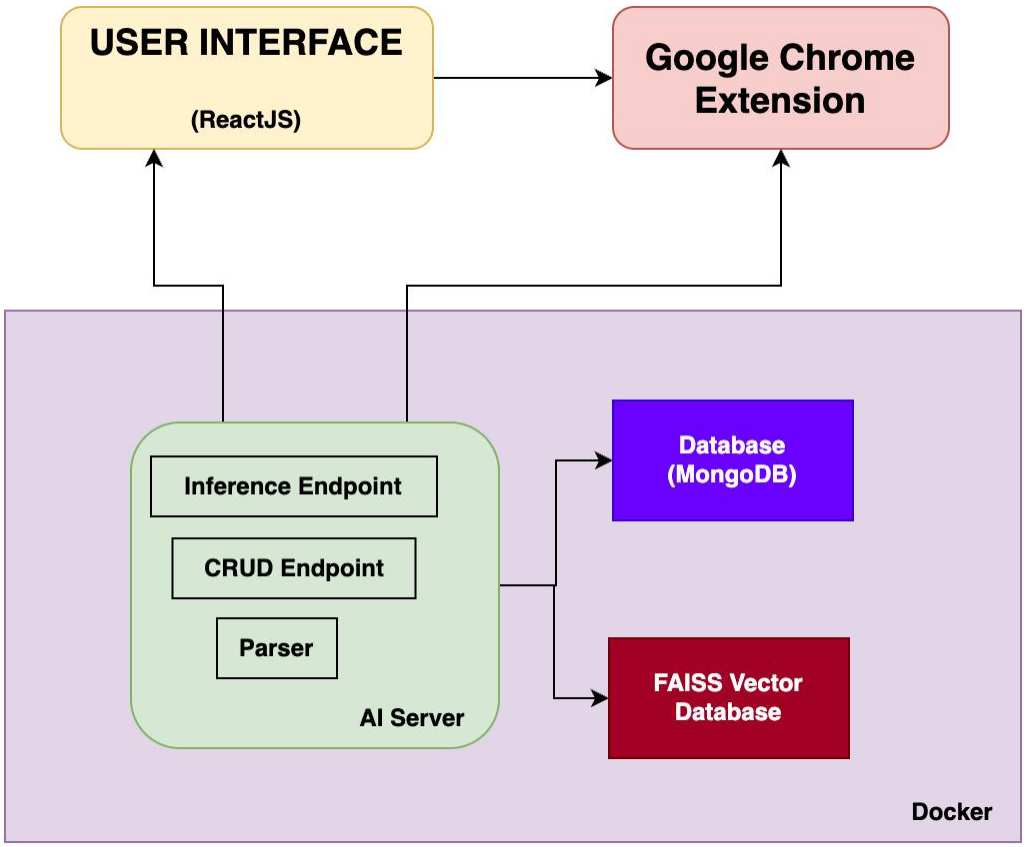

Figure 1: Architecture

ACM Reference Format:
Anirudh Ragavender Ramarathnam Madhavan, Krishna Anandan Ganesan, Saairam Venkatesh, and Sujithra Rajan. 2018. Intellimark - Intelligent Bookmark Aggregator. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

Unpublished working draft. Not for distribution.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

1 PROJECT LINK

Intellimark - Github

2 INDIVIDUAL CONTRIBUTION

- *Anirudh Ragavender* - ar80: Anirudh worked on setting up the base architecture for the entire application. He developed the components of the AI server and worked on beautifying the UI of the web-app.
- *Sujithra Rajan* - rajan11: Sujithra used the base architecture and worked on creating the extension. She also worked on integrating extension with web-app.
- *Krishna Anandan* - kag8: Krishna worked on UI of the web-app and integrating extension and MongoDB with the web-app.

- *Saairam Venkatesh* - saairam2: Saairam worked on connecting FAISS vector database with MongoDB which was the crucial connecting component of the application. He additionally worked on integrating Google OAuth with the web-app.

3 MOTIVATION

In the contemporary era marked by an exponential increase in digital information, individuals and organizations face the daunting task of managing a deluge of data that surpasses human capacity for consumption and organization. Traditional bookmarking tools, while useful, often fall short due to their static and context-insensitive methodologies. The proposed IntelliMark system seeks to revolutionize this domain through an intelligent, adaptive framework that utilizes semantic content analysis for data aggregation and retrieval.

By leveraging cutting-edge natural language processing technologies and employing a BERT-based AI model, IntelliMark automatically categorizes bookmarks and uses a FAISS vector database to enhance the accuracy and speed of information retrieval. This system is designed not only to address the ubiquitous challenge of information overload but also to significantly improve the relevance and usability of managed data. With its user-friendly interface, including a seamless browser extension and robust backend infrastructure, IntelliMark transforms traditional data management into a more efficient, intuitive process, making it essential for navigating today's dynamic information landscapes across various educational and professional fields.

4 INTENDED USERS

The IntelliMark system is designed with a specific focus on students and working professionals who frequently deal with large amounts of informational resources. These users often face the challenge of managing diverse types of digital content, including scholarly articles, industry reports, multimedia resources, and online databases. IntelliMark is crafted to integrate effortlessly into their daily routines, streamlining the organization and retrieval of digital content.

By automating key processes, IntelliMark substantially reduces the cognitive load and manual labor traditionally involved in bookmark management. This allows users to dedicate more time and energy to analytical and creative tasks, rather than getting bogged down by the mechanics of data handling. The result is a significant boost in productivity, particularly in educational and professional environments where the quick retrieval of accurate and relevant information is critical.

This enhancement of digital resource management with IntelliMark not only makes daily tasks more efficient but also sets new standards for productivity in various settings. Its impact is especially notable in sectors where staying updated and quickly accessing precise information can have profound implications on the success of educational and professional projects.

5 MAJOR FUNCTIONS

The *IntelliMark* system is designed to make managing digital bookmarks easy and efficient, from the moment you save a bookmark to when you need to find it again. Here's a detailed look at how each part of the system works:

5.1 Document Parsing and Processing

At the core of *IntelliMark* is a tool called a parser, which can handle different types of documents like PDFs, web links, images, and Word documents. Currently, it's especially good at working with web links. The parser starts by cleaning up the bookmarks. It removes common words that aren't necessary for understanding the content, known as stop words. This cleaning step is important because it makes sure that the main ideas of the documents are clear and ready for further analysis.

5.2 Content Analysis and Topic Modeling

After the documents are cleaned up, the system analyzes them in more depth. It uses BERT model, which has learned from a lot of Wikipedia articles to understand text better. This model turns the text into a series of *numbers* (called embeddings) that represent the deeper meaning of the words. These *numbers* are then stored in a special database called FAISS, known for being very good at storing and finding these number series quickly. FAISS is chosen because it can handle lots of data efficiently and find documents that are similar to each other quickly when needed.

5.3 Query and Retrieval System

IntelliMark makes searching for bookmarks straightforward. When you search for something, the system uses the same BERT model to turn your search terms into embeddings. Then, it looks through the database to find documents that match these embeddings. It uses a method called cosine similarity to see how close your search is to the bookmarks in the database, making sure that the results are relevant and closely related to what you're looking for.

5.4 Database Management

To keep track of all the bookmarks and their details, *IntelliMark* uses a database called MongoDB. MongoDB is flexible, which means it can handle different types of data from various documents easily. This flexibility is crucial because it helps maintain accurate and detailed information about each bookmark without slowing down the system.

6 SYSTEM ARCHITECTURE

The *IntelliMark* system is designed to enhance the management and retrieval of digital bookmarks through a well-structured, modular architecture that combines advanced front-end and back-end technologies. This system aims to improve user experience and backend efficiency, making it ideal for handling large volumes of data while remaining responsive to user queries.

6.1 System Overview

6.1.1 User Interface: Developed using ReactJS, the user interface offers a robust and intuitive platform for users to manage their bookmarks effectively. It supports a variety of user actions such as searching, querying, and bulk uploading of bookmarks. This interface is tailored to be user-friendly, facilitating seamless interaction with the system's extensive functionalities.

6.1.2 Google Chrome Extension: A key component of the *IntelliMark* system is the Google Chrome extension. This extension integrates directly with the user's browser, allowing them to easily save new bookmarks, mark pages as favorites, and receive recommendations for similar content based on their browsing history. This proactive feature enhances the browsing experience by bringing relevant information and suggestions to the user, thus enriching the way they interact with online content.

6.2 Backend Architecture

6.2.1 AI Server: At the core of the backend is the AI server, which houses the system's intelligence. It includes several key functionalities:

- **Inference Endpoint**: This component uses advanced machine learning models to analyze bookmarked content and generate meaningful insights, such as categorizing content based on themes or subjects.
- **CRUD (Create, Read, Update, Delete) Endpoint**: These endpoints are essential for managing the bookmarks database, allowing users to modify or update their bookmarks as needed.
- **Parser**: The parser is crucial for processing incoming data from bookmarks. It preprocesses the data by removing irrelevant content (like stop words), making the data cleaner for further processing.

6.2.2 Databases:

- **MongoDB**: This NoSQL database is used to manage unstructured metadata from user bookmarks. Its flexibility is ideal for handling diverse data types that the system encounters with different bookmarks.
- **FAISS Vector Database**: Known for its high-performance in handling high-dimensional data vectors, this database stores the processed data vectors from bookmarks. It enables quick retrieval of similar bookmarks through efficient similarity search algorithms, significantly speeding up the query response time.

6.2.3 Dockerization: To ensure that the backend is scalable and its environment isolated, the entire backend infrastructure, including the AI server and databases, is dockerized. Docker provides a consistent environment that is crucial for maintaining the system's performance and reliability across different deployment scenarios.

7 SYSTEM INTERFACE VISUALIZATIONS:

8 STEPS TO USE

8.1 Installation:

- Clone the repository:
  `git clone https://github.com/SaaiVenkat/cs510-project`
- Navigate to the extension directory: `cd extension`
- Install dependencies: `npm install`
- Navigate to the frontend directory:
  `cd frontend`

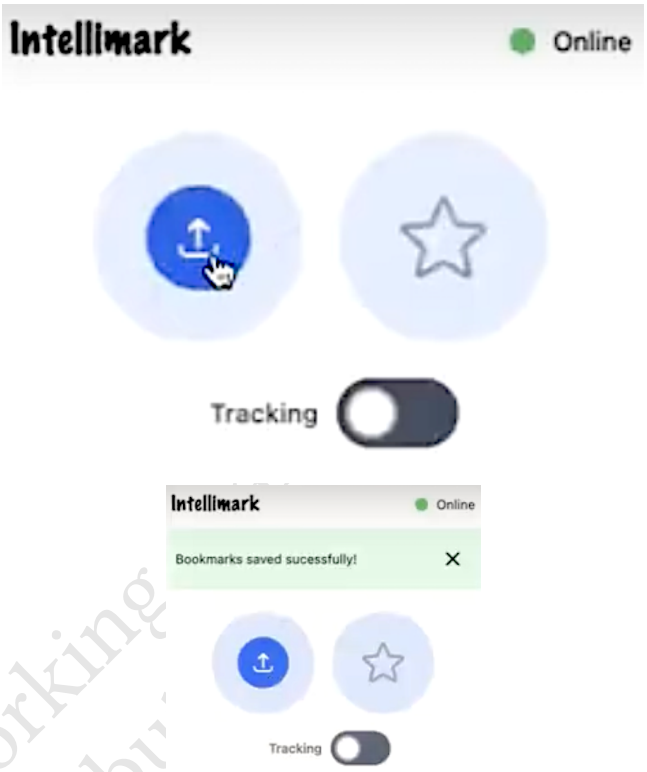

Figure 2: Chrome Extension with Two Buttons: "Bookmark Existing Bookmarks" and "Save Current Bookmark"

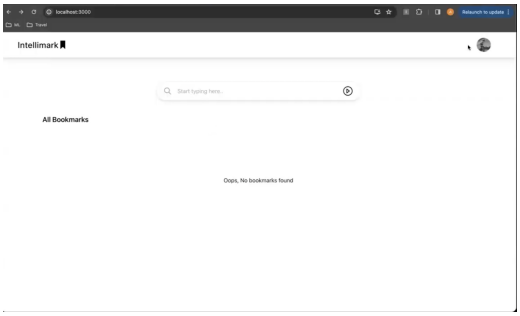

Figure 3: User Interface Prior to Bookmark Addition

- Install dependencies: `npm install`
- Navigate to the server directory:
  `cd ai-server`

8.2 Usage:

- Run extension: `npm run build`
  `npm run watch`
- Run frontend: `npm run start`
- Start server: `python3 main.py`

Load the generated extension directory to the browser and load it

- Access the application using http://localhost:3000

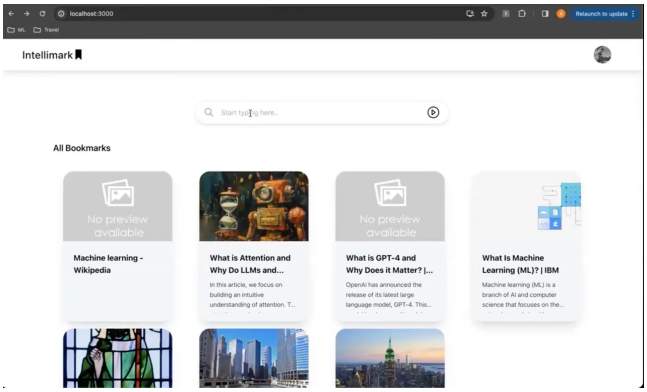

Figure 4: Results after uploading bookmarks

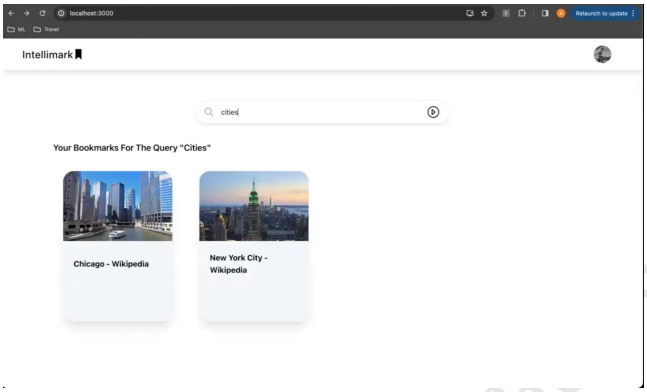

Figure 5: Search Results for 'Cities'

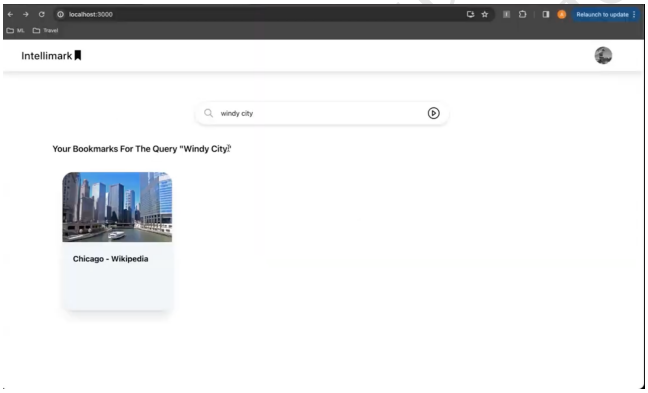

Figure 6: Search Results for 'Windy city'

- Click on the *Upload* button to upload all the existing bookmarks
- Log in using the *Sign in* button
- Click on the *Favorites* button to save current bookmark.

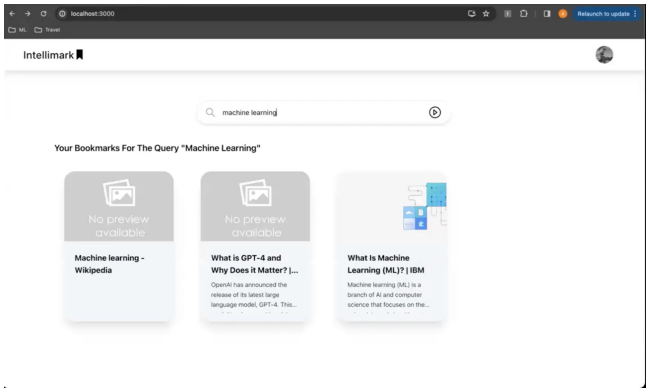

Figure 7: Search Results for 'Machine learning'

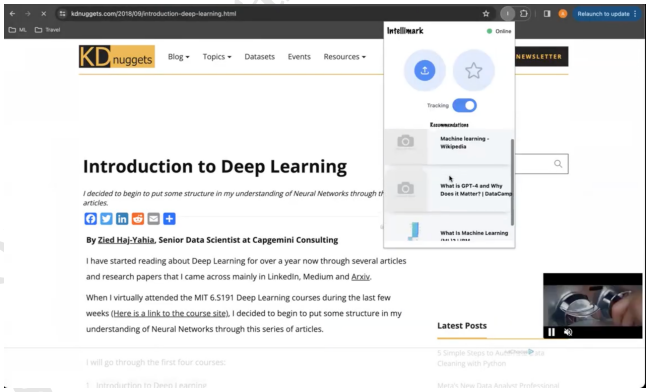

Figure 8: Recommendations

9 FUTURE WORKS

The IntelliMark system, while robust in its current form, presents several opportunities for enhancement and expansion to meet the evolving demands of digital information management. The roadmap for future development focuses on extending the system's capabilities, improving user engagement, and expanding its accessibility across different platforms and devices. Below are the envisioned enhancements along with additional ideas for further innovation:

9.0.1 Advanced Personalized Recommendations: Building on the current framework, future iterations of IntelliMark will incorporate more sophisticated machine learning algorithms to analyze user interactions and bookmark data extensively. This analysis will facilitate the delivery of highly personalized content recommendations, tailored to the individual preferences and historical behaviors of users. By enhancing user discovery and engagement, IntelliMark aims to transform how users interact with and benefit from their digital resources, making information consumption more intuitive and aligned with personal interests.

9.0.2 Cross-Platform Browser Extension: To cater to a wider audience and enhance user experience across various digital environments, the development of the browser extension will be extended to support multiple browsers, including Firefox and Safari, as well as

mobile browsers on iOS and Android platforms. This expansion will ensure that users can leverage IntelliMark's functionalities regardless of their chosen browser or device, fostering a more inclusive and flexible user interface.

9.0.3 Seamless Synchronization Across Devices: A key area of future development is the enhancement of synchronization capabilities to ensure that users can access their bookmarks seamlessly across different devices and platforms. By improving cloud synchronization technologies, IntelliMark will enable users to maintain a consistent and up-to-date collection of bookmarks, accessible from anywhere at any time. This feature is particularly crucial for users who switch between multiple devices, such as smartphones, tablets, and desktop computers, in their daily activities.

9.0.4 Social Bookmark Sharing: An exciting avenue for expansion is the introduction of social features that allow users to share bookmarks with peers. This functionality will not only enhance discovery by exposing users to curated content from trusted sources but also foster collaboration and knowledge exchange among communities. Social sharing could be particularly beneficial in educational and professional settings, where teams or study groups can benefit from shared resources.

9.0.5 Integration with Academic and Research Tools: Recognizing the significant number of users in academic and research settings, future versions of IntelliMark could integrate with research databases and academic tools. This integration would streamline the process of managing and citing scholarly articles and data, enhancing productivity for researchers and students alike.

9.0.6 Enhanced Data Security and Privacy: As the system evolves to handle more user data, especially with the integration of social sharing and personal recommendations, a heightened focus will be placed on data security and privacy. Future developments will include advanced encryption methods and robust privacy controls to ensure that user data is protected against unauthorized access and breaches.

10 CONCLUSION

The IntelliMark system marks a significant step forward in digital bookmark management, combining advanced technologies to enhance both user experience and efficiency. By automating the organization and retrieval of bookmarks with AI, IntelliMark allows users, especially students and professionals, to focus more on their main tasks rather than on managing vast digital information. Its user-friendly interface, powered by ReactJS, and a robust backend featuring AI technologies such as BERT and FAISS, ensure scalability and responsiveness. The Chrome extension further enriches the user experience by enabling real-time bookmarking and recommendations.

In conclusion, IntelliMark sets a new benchmark in digital content management, offering a powerful tool that not only meets today's demands but is also poised to adapt to future technological advancements. This makes it an essential resource for enhancing productivity and managing digital information effectively.

Received 10 May 2024; revised 10 May 2024