

Proposal: Intelligent Information Retrieval System for Real Estate Investment Analysis

Track	Development Track
Team Members	Jason Hu [jasonh11@illinois.edu], Nianze Guo [nianzeg2@illinois.edu], Chenhan Luo [chenhan8@illinois.edu], Haoran Tang [ht18@illinois.edu]
Project Coordinator	Jason Hu

Functions and Users

We propose to develop a real estate-focused information retrieval (IR) system with a web-based interface that helps investors search for and rank properties based on investment potential. The system will function as a web-based search engine, allowing users to query properties based on factors such as rental yield, appreciation potential, local economic trends, and real estate sentiment analysis. The tool will provide real estate investors, financial analysts, and homebuyers looking for high-return properties with ranked search results, filtering and sorting properties based on financial criteria rather than just location and price.

Significance

Most existing real estate platforms (e.g., Zillow, Redfin) focus only on basic property listings without advanced investment-based retrieval and ranking. Investors often struggle with information overload and lack efficient ways to filter and rank properties based on long-term returns. Our system addresses this gap by integrating advanced ranking models to help investors identify the best properties based on data-driven insights. By offering a specialized search engine, our system will save investors time by retrieving financially relevant properties first, reduce the need for manual research by automatically analyzing market trends, and provide personalized rankings and recommendations based on user preferences.

Approach

Our system will function as a real estate-focused search engine, aggregating and indexing property listings from sources like Zillow, Realtor.com, and public real estate databases. We will collect rental income data, historical price trends, tax rates, neighborhood statistics, and economic indicators to enhance retrieval effectiveness. Additionally, NLP techniques will be used to analyze property descriptions, detect investment-related keywords, and extract insights from real estate news articles and discussions.

For search and ranking, we will implement BM25 and Learning-to-Rank (LTR) models to prioritize properties based on rental yield, appreciation potential, and market demand. A custom scoring function will rank properties based on these financial metrics, helping investors identify high-return opportunities. To improve user experience, we will integrate query expansion techniques that suggest alternative searches based on user behavior and market trends. The system will be built using FastAPI (backend), Elasticsearch (search indexing), PostgreSQL (database), React (frontend), and NLP models from Hugging Face and spaCy. Potential risks include data accessibility limitations, which we plan to mitigate through web scraping and open datasets, and ranking model accuracy, which will be refined through continuous evaluation and feedback.

Evaluation

To evaluate our system, we will conduct relevance testing by comparing its property rankings with recommendations from real estate investment experts to ensure alignment with professional decision-making criteria. Additionally, we will perform a user study to assess user satisfaction and measure the efficiency of investment property searches, specifically tracking how quickly users identify promising opportunities. Lastly, we will use IR metrics such as Precision@K with historical performance data, Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (NDCG) to quantitatively assess the effectiveness of our ranking algorithm in retrieving high-quality investment properties.

Timeline

Initial data collection and preprocessing	March 30, 2025
Search indexing and retrieval implementation	April 10, 2025
Ranking model development and refinement	April 20, 2025
System integration and frontend development	April 30, 2025
Testing, evaluation, and final improvements	May 5, 2025
Project submission and presentation preparation	May 9, 2025

Task Division

Jason Hu	Overall system architecture, backend API development
Nianze Guo	Data collection and processing, web scraping, database management
Chenhan Luo	Search ranking model implementation, relevance evaluation
Haoran Tang	Frontend development, UI/UX design, system integration