Project 2: Prototype IR System
Information Retrieval and Storage Report

Presented by

6488002 Sirasan Srijirapattanakul

6488193 Thanapat Riluechai

6488217 Pattaradnai Kaeodumkoeng

Presented to

Aj. Suppawong Tuarob

ITCS 414 – Information Retrieval and Storage
Semester 1 / 2023
Faculty of Information and Communication Technology
**Mahidol University**

Introduction

In luxury travel, a private jet search engine is offering users a gateway to the exclusive world of private aviation. It optimizes for individuals and businesses owner who want the ultimate in convenience, comfort, and efficiency. This search engine will provide detailed insights into various private jets and encompassing their specifications, facility, and performance capabilities. It would be more useful to people who are observing a private jet or want to be an owner. This search engine simplifies the process and makes it simple in concept friendly user.

Problem(s) that you are trying to solve

To assist a potential buyer in understanding the capabilities of an aircraft and aiding their decision to purchase, providing detailed information about the plane's features, performance specifications, and benefits is crucial. There are many problems for users with too much information that can be confusing, and it is not easy to find all the details in one place. This search engine is like a tool that brings together all the important information about private jets. It makes things easier and quicker for buyers to learn about different jet models and make the best decision for themselves. It is a friendly guide for people who might be new to the world of private jets making the whole process of buying one more straightforward and enjoyable.

Existing relevant systems (i.e., Are there any other similar search engines as yours?)
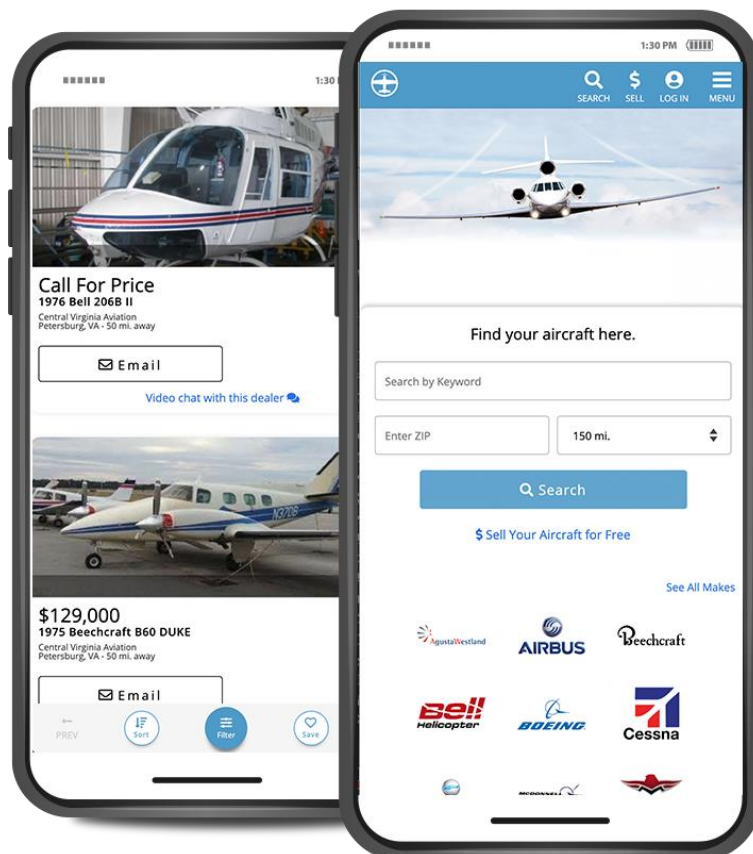
Aero Trader

AeroTrader is a leading aviation site that caters to both buyers and sellers in the aviation market. This platform's inventory is comprehensive, featuring a wide range of new and pre-owned airplanes, helicopters, and other aviation-related products and services. For sellers, Aero Trader offers an easy-to-use online inventory management tool, complete with access to metrics and email reporting, enhancing the selling experience. The site's user-friendly nature

extends to buyers as well, providing a vast selection of aircraft for sale from various locations. These include aircraft like helicopters, jets, ultralight aircraft, military planes, single-engine and multi-engine propeller aircraft, and more.

AeroTrader.com emphasizes customer-centric service, ensuring users have a seamless experience whether they are buying or selling aircraft. They offer assistance with placing ads for sellers, aiming to reach a broad audience of potential buyers. The platform also has a dedicated Help section to address any queries users might have, reflecting its commitment to user satisfaction and ease of use.

ADN - Aircraft Search Engine

ADN - Aircraft Search Engine, operating through FindAircraft.com, is a comprehensive online platform tailored for the aviation market. It primarily facilitates the buying and selling of aircraft by providing a constantly updated environment. With a massive listing of over 16,000 aircraft for sale, it connects buyers and sellers effectively. This includes a wide range of aircraft from different manufacturers such as Airbus, Beechcraft, Falcon, Learjet, Boeing, Cessna, Embraer, Gulfstream, and more. ADN's extensive database includes detailed information on each aircraft model, covering aspects like dimensions, weight, power, range, speed, and fuel consumption. This is meticulously gathered by a team of experts, ensuring that the data is comprehensive and reliable.
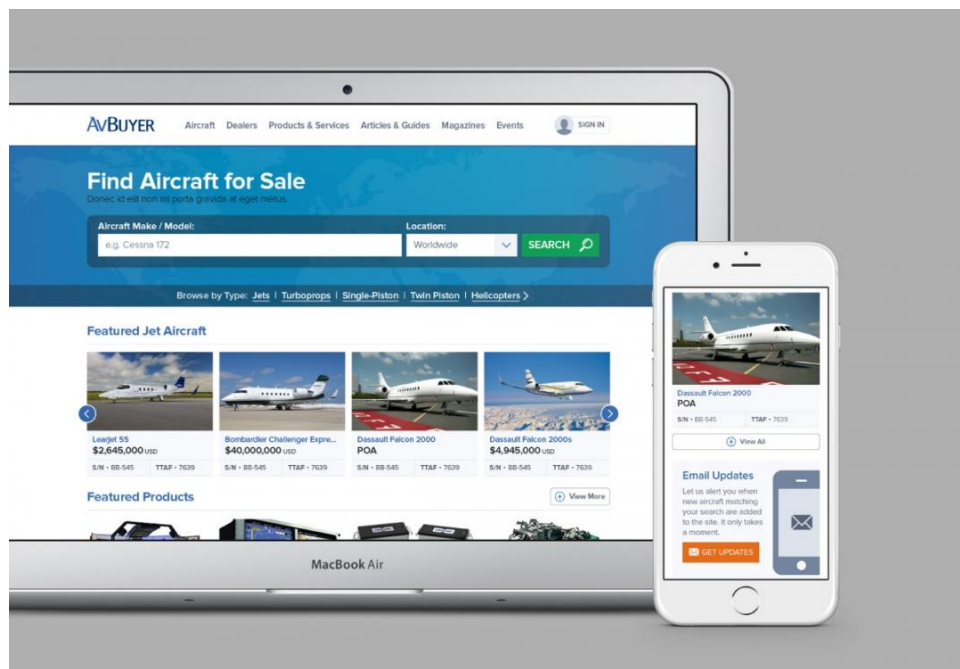
AvBuyer

AvBuyer, known for its role in the aviation industry, operates as a key online marketplace and information platform. It primarily focuses on aircraft sales, showcasing a diverse array of aircraft types such as private jets, turboprops, helicopters, and vintage models. Its reach and utility extend well beyond mere sales; AvBuyer offers comprehensive market intelligence, assisting clients in understanding current market trends and values. In addition to sales and market data, AvBuyer also provides essential services that are vital for aircraft ownership and operation. This includes detailed information and services related to aircraft inspection, maintenance, and upgrades, ensuring that aircraft owners and operators have access to the necessary resources for keeping their aircraft in top condition.

Implementation

In the implementation part, we are installing Elastic search for search and analytics engine. Next step, we install Kibana for visualization platform that works in conjunction with the Elastic Stack, a collection of open-source tools for data ingestion, storage, analysis, and visualization. We also use Konbert which is an online data conversion tool to convert our dataset CSV to NDJSON file. After that we use VScode to write the front end by using HTML and CSS to create a search. In the back-end part use Node.js to run a server. Last part we install Flask to web framework that is popular for building web applications. It will be used to create the search engine's web interface.

I. Data collection, example documents, and data statistics

We have compiled and organized aircraft information from Aero Trader and ADN - Aircraft Search Engine which include aircraft names, production years, descriptions, and image links. This dataset is collected using Excel spreadsheets and after that converted to csv file.

name	create	description	link
Cessna Citation 500	1976	The Cessna Citation 500, also known as the Citation I, is a small business jet aircraft that was produced by Cessna Aircraft Company from 1971 to 1986. It is the first aircraft in the Citation family of business jets, which is one of the world's most successful and enduring lines of business aircraft. Citation 500 Eagle with upgraded Garmin 625Xi system with WAAS/LPV. Freon Air, Fresh Phase 1-5 and a fresh HSI.	https://resources.global air.com/specs/images/Jets/Cessna/Citation/50 0/Exterior/ext2.jpg?w= 650&h=430&mode=max
Cessna Citation Encore	2002	The Cessna Citation Encore, also known as the Cessna Citation C-560 Encore, is an improved version of the Cessna Citation II, a light corporate jet aircraft produced by Cessna Aircraft Company from 1991 to 2006. The Encore was designed to offer increased range, payload, and cabin comfort over its predecessor. One Owner / Pilot Since New,Only 1970 Total Flight Hours, Primus 1000 Flight Deck w/ WAAS / LPV ,Immaculate Paint & Interior	data:image/jpeg;base64 /9j/4AAQSkZJRgABA QAAAQABAAD/2wCE AAoHCBUVFRgVFRY YGBgaGBkYGBocGho sGBwcGBoZGhocGBgc lS4IHB4rHxoYJjgmKy8x NTU1HC07QDs0Py40N TEBDAwMEA8QGBISH jQhlSE0NDQxNDE0ND Q0NDQxMTQ0NDE0N DQxNDE0ND00MTQ0 NDQ0MTExNDQ0PsQ0 NDQxPsQ0P//AABElAL sBEwMBlgACEQEDEQ H/xAAbAAACAgMBA AAAAAAAAAAAAAA BAgADBAUGB//EAEE QAAEDAgMFBQYEQ IFB9AAAAEAAhEDIQ QSMQVBUWFxBoGBo fATIjJCscFSYnLRFCQS 4fGC0kODk8LTBxUw RFT/xAAZAQEBAQEB AQAAAAAAAAAAAAA AAQIDBAX/xAAgEQ
Cessna Citation X	2005 2023	The Cessna Citation X, also known as the Model 750, is an American business jet aircraft produced by Cessna and part of the Citation family. It is the fastest business jet in the world, with a maximum cruise speed of Mach 0.924 (610 mph or 982 km/h). The Citation X was introduced in 1990 and is known for its combination of speed, range, and luxury. Engines Enrolled on Rolls-Royce Corporate Care Enhanced, Honeywell Primus Elite Avionics Package, Doc 3, 6, 10, 11, 13, 15 & 18 Inspections by Textron-ICT	data:image/jpeg;base64 /9j/4AAQSkZJRgABA QAAAQABAAD/2wCE AAoHCBISEREREhERE REBEQ8REQ8REBERE Q8PGBQZGRgUGBgcl S4IHB4rHxgZJjgmKz0x QDU1Gi07QDssPy40N TEBDAwMEA8QGBISH sQklSQ0NDQ0NDQN DQxND Q0NDQNDQx NjY0ND Q0ND Q0MTE0 NDQ0NDE0ND Q0NDQ 0ND Q0ND Q0NP/AABE lALcBEwMBlgACEQED EQH/xAAbAAACAwEB AQAAAAAAAAAAAAAA AAgEDBAUGB//EAE MQAAICAQIDBQUEB g/clAwAAAAECAAMBB BIFTEGE0FRYSlpcYG RFFKhclJicpKiwcVlc OC0fBTY3ODk7LC8R YkRP/EABgBAQEBAQ EAAAAAAAAAAAAA AAAABAgME/8QAIG REB

data:image/jpeg;base64

PlaneData.ndjson

```
{} PlaneData.json > ...
  1  {"name":"Cessna Citation 500","created":1976,"description":"The Cessna Citation 500, also known as the Citation I, is a small business jet a
  2  {"name":"Cessna Citation Encore","created":2002,"description":"The Cessna Citation Encore, also known as the Cessna Citation C-560 Encore,
  3  {"name":"Cessna Citation X","created":2005,"description":"The Cessna Citation X, also known as the Model 750, is an American business jet a
  4  {"name":"Cessna Citation CJ4 ","created":2020,"description":"The Cessna Citation CJ4 is a mid-size business jet aircraft produced by Cessna
  5  {"name":"Cessna Citation Sovereign","created":2006,"description":"The Cessna Citation Sovereign, also known as the Model 680, is an Americar
  6  {"name":"Cessna Citation CJ3+","created":2018,"description":"The Cessna Citation CJ3+ is a mid-size business jet aircraft produced by Textro
  7  {"name":"Cessna Citation XLS","created":2007,"description":"Best Pedigree, Meticulous Owner 8 Passenger Configuration Engines on ESP Gold |
  8  {"name":"Cessna Citation II","created":1993,"description":"The Cessna Citation II, also known as the Model 550 or Citation S/II, is a light
  9  {"name":"Cessna Citation XLS+","created":2014,"description":"The Cessna Citation XLS+, also known as the Model 560XLS+, is an American mids
 10  {"name":"2012 Cessna Citation CJ4","created":2012,"description":"The 2012 Cessna Citation CJ4 is an upgraded version of the Citation CJ3, o
 11  {"name":"2019 Cessna Citation CJ3+","created":2019,"description":"The 2019 Cessna Citation CJ3+ is an upgraded version of the Citation CJ3,
 12  {"name":"Cessna Citation Mustang","created":2008,"description":"The Citation Mustang features a spacious and comfortable cabin that can acc
 13  {"name":"Cessna Citation Sovereign","created":2004,"description":"The Cessna Citation Sovereign, also known as the Model 680, is an Americar
 14  {"name":"Cessna Citation M2","created":2022,"description":"The Cessna Citation M2, also known as the Model 525, is an American light busine
 15  {"name":"Cessna Citation 500","created":1975,"description":"The Cessna Citation 500, also known as the Model 500 or Citation I, is an Ameri
 16  {"name":"2008 Cessna Citation Sovereign","created":2008,"description":"The 2008 Cessna Citation Sovereign is an updated version of the Citat
 17  {"name":"Cessna Citation X","created":2004,"description":"The Cessna Citation X is a family of business jets produced by Cessna Aircraft Cor
 18  {"name":"Cessna Citation Excel","created":2001,"description":"Only 2 US Owners Since New Meticulous Maintenance and Care Since New\n560 / 0
 19  {"name":"Cessna Citation III","created":1984,"description":"\nThe Cessna Citation III is a mid-size business jet aircraft that was produced
 20  {"name":"2014 Cessna Citation XLS+","created":2014,"description":"Fully Enrolled on Programs 8+1 Passenger Configuration Medevac Configurat
 21  {"name":"Cessna Citation V Ultra","created":1996,"description":"The Cessna Citation V Ultra, also known as the Model 560 Ultra, is an upgrad
 22  {"name":"Cessna Citation Bravo","created":2003,"description":"The Cessna Citation Bravo, also known as the Model 550 Bravo, is a mid-size bu
 23  {"name":"Cessna Citation CJ4","created":2010,"description":"The Cessna Citation CJ4 is a light-midsize business jet aircraft produced by Ces
 24  {"name":"2005 Cessna Citation Sovereign","created":2005,"description":"Two U.S. Owners Since New, Corporate Pedigree Fresh Doc 1, 2, 3, 5,
 25  {"name":"Cessna Citation M2","created":2019,"description":"The Cessna Citation M2, also known as the Model 525, is an American light busines
 26  {"name":"Cessna Citation X","created":1999,"description":"The Cessna Citation X is an American business jet aircraft produced by Cessna Airc
 27  {"name":"Cessna Citation Encore ","created":2002,"description":"Three U.S. Owners Since New Well Organized Records Since Delivery WAAS | EGF
 28  {"name":"2006 Cessna Citation Sovereign","created":2006,"description":"AFTT: 44636.6 | Two U.S. owners since new (13 years with current owne
 29  {"name":"2002 Cessna Citation Encore","created":2002,"description":"Dual Garmin GTN-750 NAV/COM/GPS Garmin GTX-3000 Transponders w/ADS-B Out
 30  {"name":"2007 Cessna Citation Mustang","created":2007,"description":"Garmin NXi Phase II Upgraded Avionics Excellent Pedigree and Ownership
 31  {"name":"Airbus ACJ319","created":2011,"description":"2008 EIS 2011 Airbus ACJ-319 | MSN 3542 Cabin Certified for 18 pax commercial operatic
 32  {"name":"Airbus ACJ TwoTwenty","created":2023,"description":"We are offering a turnkey package, including ferry flight from the Airbus faci
 33  {"name":"Airbus ACJ330","created":2015,"description":"The Airbus ACJ330 is a swept-wing, four-engine jet aircraft with a high-set T-tail. I'
 34  {"name":"Airbus ACJ318","created":2010,"description":"Newest ACJ318 on the market Three zones with VIP Dining and Private Office area Extenc
 35  {"name":"Airbus ACJ320","created":2002,"description":"The Airbus ACJ320 is a wide-body business jet aircraft produced by Airbus Corporate Jc
 36  {"name":"Airbus ACJ318 Elite","created":2010,"description":"The Airbus ACJ318 Elite is a derivative of the Airbus A318 business jet, equippc
 37  {"name":"Dassault Falcon 2000S","created":2018,"description":"The Dassault Falcon 2000S is a twin-engine, mid-size business jet aircraft prc
```

II. Tools and software

Elastic Search: Elastic search stands out as a prominent search engine in the digital landscape, renowned for its distributed, open-source design. This robust search and analytics engine is highly acclaimed for its ability to handle an ever-expanding array of use cases, making it a staple tool in the tech arsenals of companies worldwide, irrespective of their size. Its core functionality revolves around the efficient storage, searching, and analysis of voluminous data sets. The foundation of Elastic search rests on the Apache Lucene library, a testament to its powerful and flexible search capabilities.

Kibana: It is a visualization platform that works in conjunction with the Elastic Stack, a collection of open-source tools for data ingestion, storage, analysis, and visualization. Kibana allows users to explore, analyze, and visualize data stored in Elastic search, Elastic search's document-oriented search and analytics database.

Konbert: It is an online data conversion tool that allows you to convert tabular files, structured data, and databases between different formats, including CSV, JSON, Avro, Excel, SQL, XML, and HTML.

Vscode: It is a free and open-source code editor developed by Microsoft for Windows, Linux, and macOS. It is a cross-platform and lightweight editor that is popular among developers for its wide range of features, customizability, and integration with popular extensions.

Flask: It is a lightweight Python web framework that is popular for building web applications. It is a microframework, meaning that it does not require a lot of external libraries or dependencies. This makes it a good choice for beginners and experienced developers alike.

III.   System diagram
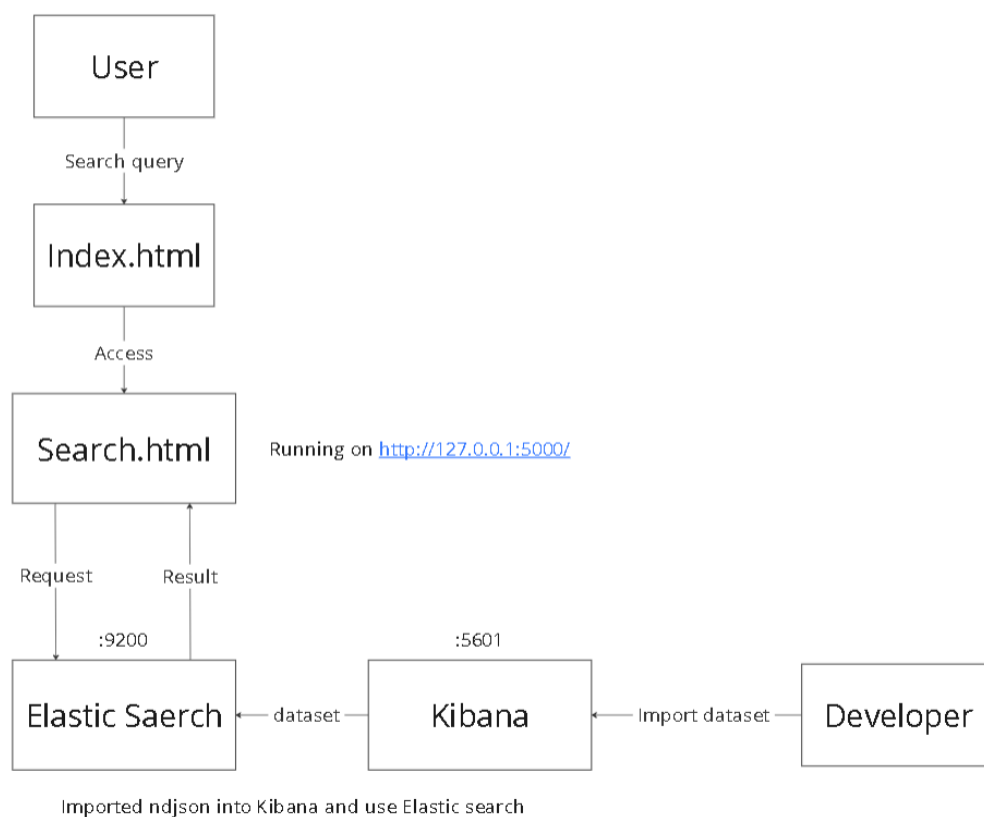
The user journey commences as individuals access the index page, providing them with the opportunity to utilize the search bar for entering specific keywords. Subsequently, the system directs them to the search result page, presenting a comprehensive list of items aligned with the search query. To optimize the search process, Elastic Search is employed for scoring and indexing. The dataset is stored in Kibana, establishing a seamless connection with Elastic search for efficient data retrieval and management.

IV.   Snapshots of the system

Index.html
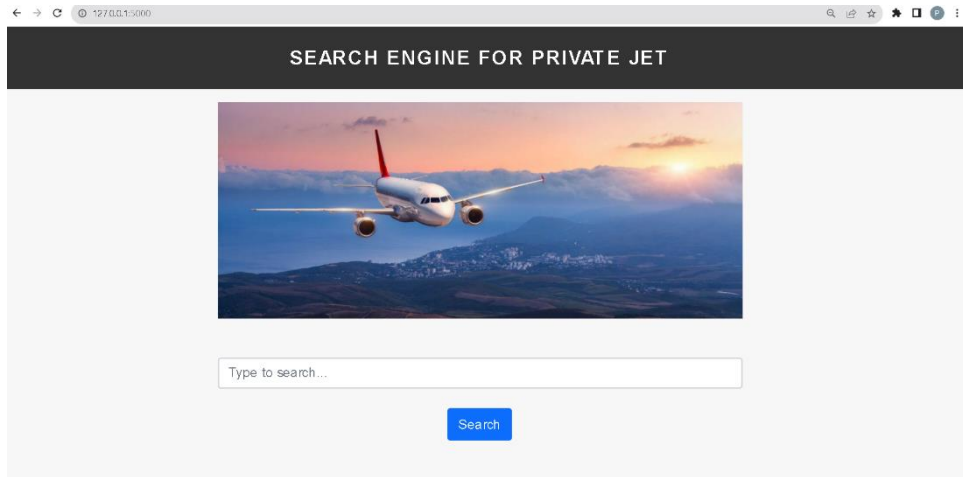
Search.html
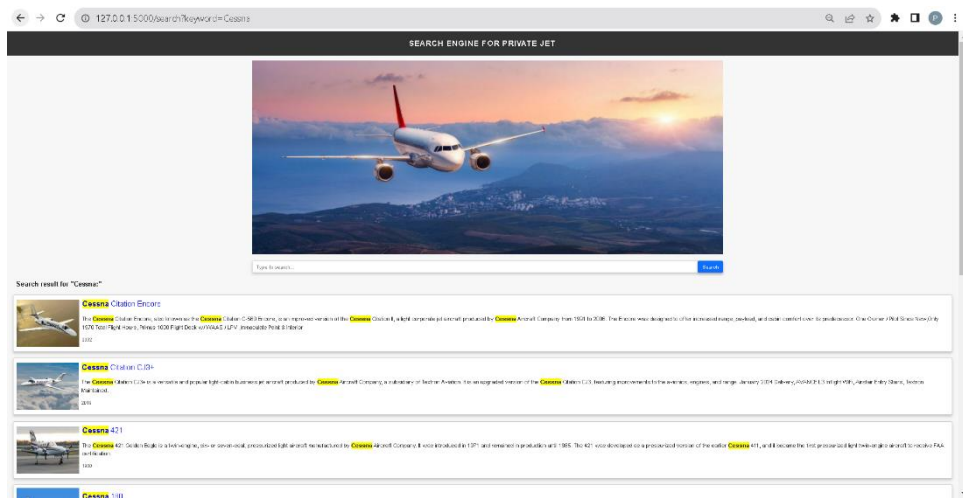
V. Example step-by-step search sessions that highlight the following functionality:

1. One word query
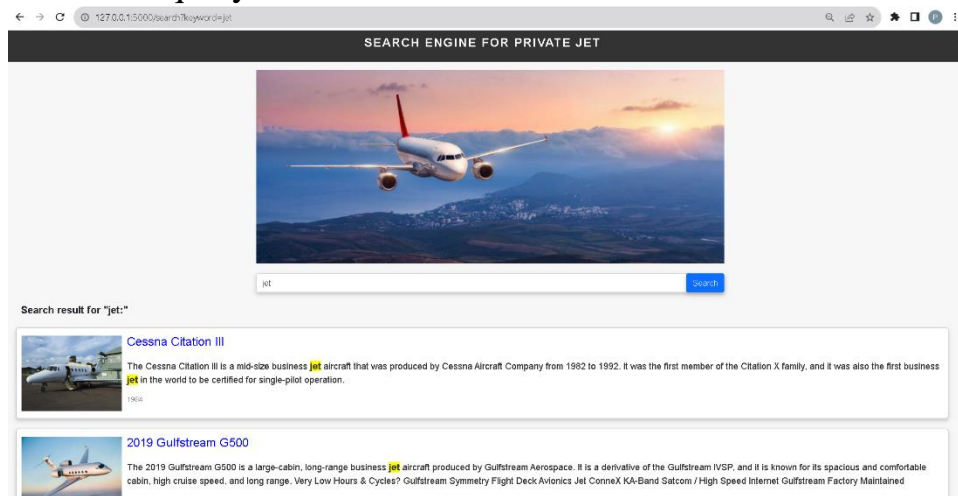
In the single-word query testing phase, we employ the term "jet" to search for private jets in the search results. The outcomes will display items featuring the word "jet" in either their description or name.

2. Multiple word query
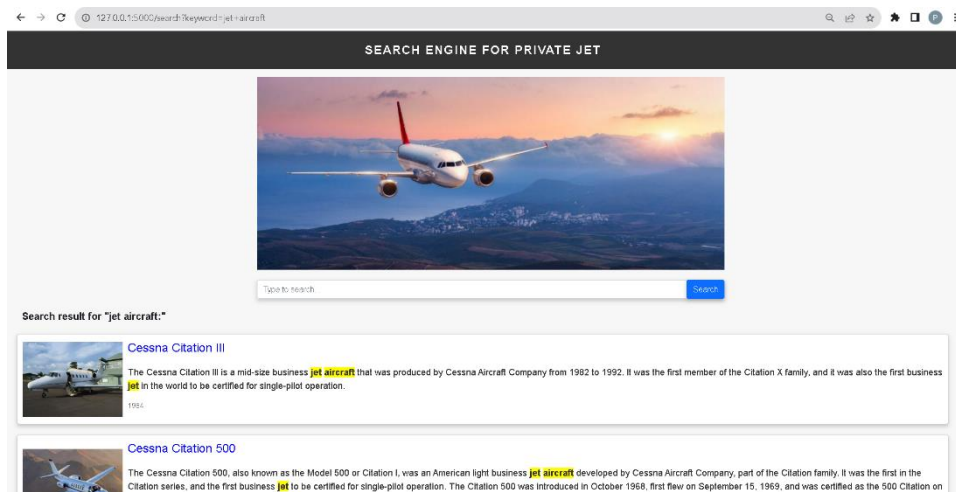
During the multiple-word query testing phase, we utilize the term "jet aircraft" to conduct a search, specifically targeting results related to jet aircraft. The displayed outcomes will showcase items that include the exact phrase "jet aircraft" in either their description or name.

3. Partial match
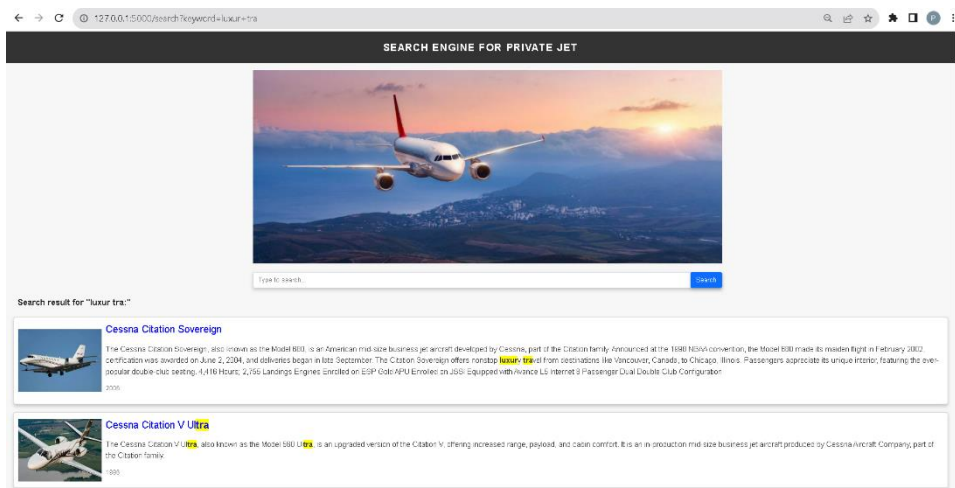
During the partial match testing, we opted to search for a segment of the phrase "luxury travel." Using the keywords "luxur tra," we conducted the search, and the results aligned with our expectations, securing the top ranking.

4. Ranking
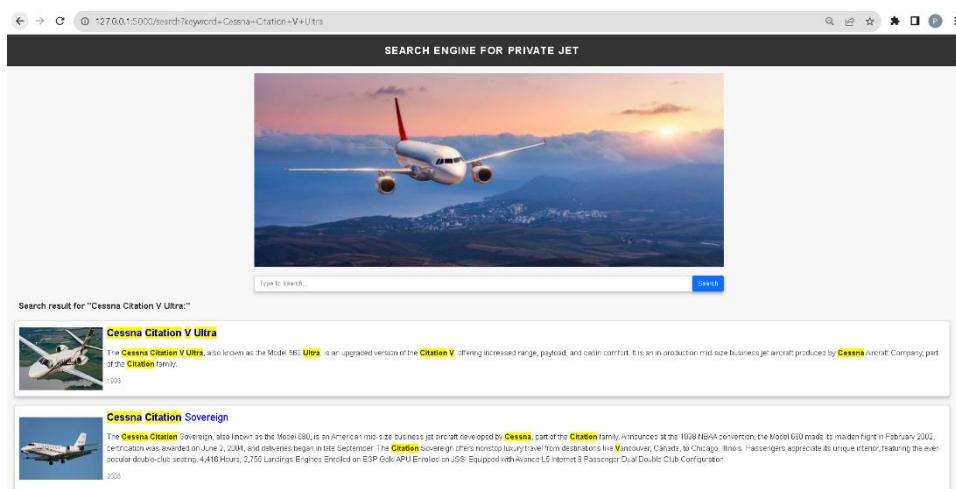
Regarding the indexing and ranking process, we will utilize the scores generated by Elastic search to determine the ranking. We made specific adjustments to the Python file to accommodate multiple-word queries and partial matching. The ranking will be presented in descending order, with the highest score positioned at the top and the lowest score at the bottom.

Discussion

I. Limitations of your system

The Limitation of dataset

The dataset for our exclusive jet search engine presented a unique challenge due to its highly specific and niche requirements. The difficulty in acquiring such specialized data stemmed from several factors. Firstly, the private jet industry is characterized by a high degree of confidentiality and exclusivity, making access to detailed and comprehensive data challenging. Information about private jet models, their specifications, availability, and pricing are not readily available or standardized, which complicates the data aggregation process.

Additionally, the dynamic nature of private jet charters, with fluctuating prices and variable availability, adds another layer of complexity. This requires not only access to current data but also the capability to frequently update and verify the accuracy of this information.

The scarcity of such specialized data necessitated a reevaluation of our project scope. It highlighted the importance of having a clear understanding of the data landscape in niche sectors and the need for innovative methods to gather and analyze data where traditional sources are limited or non-existent. This situation prompted us to explore alternative strategies, such as forming partnerships with industry insiders, leveraging crowd-sourced information, or utilizing advanced data scraping and AI technologies to synthesize information from disparate sources. These approaches underscored the need for flexibility and creativity in dealing with data acquisition challenges in specialized domains.

I. Technical difficulties, challenges, and lessons learned

Technical difficulties

- **Kibana**

  Encountering persistent timeouts in Kibana has necessitated the reinstallation of both Kibana and Elastic search. The timeouts are hindering the login process, prompting the need for a fresh token and password. This situation requires a thorough examination of the root cause of the timeouts and the subsequent reinstallation process to address the login and dataset import challenges effectively.

- **Elastic search**

  The Elastic search template provided by P'Dos initially supports only single-word queries, posing limitations on the system's capability to handle multi-word queries and partial matching efficiently. In response to this constraint, adjustments were made to the Elastic search configuration files to enhance support for multiple word queries and partial matching, thereby expanding the search capabilities and improving overall user experience.

- **Dataset**

Encountered a challenge while importing the dataset into Kibana due to compatibility issues with the CSV file format. Kibana faced difficulties processing the CSV directly, prompting the need for conversion to NDJSON (Newline Delimited JSON) format. To overcome this hurdle, an external tool, such as Konbert website, was utilized to convert the CSV file to NDJSON, ensuring seamless integration with Kibana.

- **Web coding**

During the project development, it was identified that essential concepts in web programming were overlooked. To address this, the team revisited foundational web programming principles to ensure a comprehensive understanding. Subsequently, the creation of the search.html and index files, along with their corresponding CSS styles, commenced. However, this process unveiled unexpected bugs and code related challenges that required careful debugging and troubleshooting.

Challenges

- **Changing topic cause of lack of data information**

A pivotal challenge in our project was the inadequacy of the dataset for our private jet search engine. This limitation necessitated a change in the project's focus. The scarcity of comprehensive and detailed data highlighted the critical role of robust datasets in success.

The change in topic, driven by data constraints, redirected our efforts towards exploring alternative approaches and applications of IR technologies. This shift underscored the adaptability required in research and the importance of having flexible methodologies when faced with unforeseen challenges.

- **Miscommunication**

In the beginning, we did not have an appropriate agenda to share our common objective, consequently, we lacked much of the communication process. Moreover, when we were having a discussion together, we did not use communication tools such as sharing documents since the beginning, therefore, we had a lot of miscommunications as we have different understandings during the discussion.

However, after we had spent a mutual time brainstorming the ideas and helped to set the goal of the project, we had a similar direction toward the assignments. For this reason, we can set the agenda for each person in the group and divide a role for everyone.

Later, everyone reached their goal which reached the agenda that we set, and we considered this progress as the lessons learning which we totally success in each's individual role.

lessons learned

This project makes us have more improvement about teamwork and communication in the ways of working and planning in the team. It is very important to do the project because we must understand the topic and work as a team to make the project successful and reach the target that we want to get. Another lesson is endeavor because it makes us to never give up

about problem or doing data research. It always has the solution to make our work successful and complete. The last thing that we learn from this project is web programming skills. As we study in the second year, we were studying web programming but sometimes we forget about it, so we must review old lessons that we learn from second year. We also get some new knowledge about web programing as well.

Opportunities for future improvements

As we progressed with the development of our website dedicated to private jet rankings, we immersed ourselves in the dynamic realm of data analytics and user experience design. This project is rooted in the integration of cutting-edge data analysis and intuitive design principles. Additionally, cultivating collaborations with academic institutions and industry connoisseurs can deliver distinctive insights and content. This strategy is aimed at solidifying our website's position as a reliable and thorough resource for specific audiences like business travelers, affluent families, and sports teams.

A key area for further development is the diversification of our data sources. While we currently focus on core parameters, expanding to include a broader spectrum of metrics could be beneficial. These new metrics could encompass lifestyle factors pertinent to our target audiences, accomplishments of notable alumni, and metrics that reflect real-world job readiness and success. Incorporating these elements can provide a more holistic view of the private jet experience, tailored to the sophisticated needs of our users.

Additionally, implementing interactive features like customizable ranking systems based on user preferences, and real-time feedback mechanisms can enhance user engagement. Integrating AI-driven recommendations could also personalize the experience for each user, making the platform more engaging and useful for decision-making.

By broadening our scope and enhancing user interaction, we aim to make our website not just a source of information, but a comprehensive tool for informed decision-making in the luxury travel sector.

Conclusion

Our journey to create an exclusive private jet search engine has been a tapestry of challenges, adaptations, and lessons. Focused on simplifying the process of private jet selection, we faced hurdles due to the confidential nature of industry data. Adapting to data scarcity, we pivoted towards alternative applications of information retrieval technologies, emphasizing the need for adaptability. Technical challenges, including Kibana timeouts and Elasticsearch limitations, drove us to refine our search engine meticulously. Converting the dataset to NDJSON format revealed the complexities of handling diverse data structures. Early communication missteps underscored the importance of clear agendas and effective team communication, fostering lessons in teamwork and planning. Looking forward, our vision involves diversifying data sources, incorporating interactive features, and fostering collaborations for unique insights. By broadening our scope and intensifying user interaction, our goal is to transform our website into a comprehensive tool for informed decision-making in luxury travel. As we celebrate this milestone, our commitment to continuous improvement remains steadfast. The lessons learned propel us toward a future where our search engine becomes an immersive and indispensable resource for discerning travelers in the exclusive world of private aviation.