



Project
Prototype IR System

By

6488018 Ramita Deeprom

6488025 Thitiwut Harnphatcharapanukorn

6488079 Burit Sihabut

6488215 Pongsakorn Kongkaewrasamee

A Report Submitted in Partial Fulfillment of
ITCS414 Information Retrieval and Storage

Faculty of Information and Communication Technology

Mahidol University, 2023

Introduction

This report on a research project in the Information Retrieval and Storage (ITCS414) course for the academic year 2023 consists of

- Problem(s) that we are trying to solve
- Existing relevant systems (i.e., Are there any other similar search engines as yours?)
- Implementation
- Discussion

To achieve the objectives of the study report, which are:

- To study the implementation aspect of information retrieval and web search in the context of designing and implementing a prototype vertical search engine for a specific dataset.

Table of Contents

Problem(s) that you are trying to solve	1
Existing relevant systems (i.e., Are there any other similar search engines as yours?)	2
Implementation	4
Data collection	4
Example documents	6
Data statistics	7
Tools and Software	8
Snapshots of the system	11
Example step-by-step search sessions that highlight the following functionality:	13
1. One-word query	13
2. Multiple word query	16
3. Partial match	17
4. Ranking	17
Discussion	19
Limitations of your system	19
Technical difficulties, challenges, and lessons learned	19
Opportunities for Future Improvements	21
Conclusion	22
Reference	23

Problem(s) that you are trying to solve

1. FrontEnd

- a. **Bootstrap Integration and CSS Editing:** Extensive CSS editing has led to bugs and inconsistencies in the design, posing a challenge in maintaining a consistent and visually appealing user experience.
- b. **Logical Design and Code Behavior:** We are working on aligning the logic and design of front-end elements to ensure a seamless user experience despite different behaviors in the code.
- c. **Static Image Update:** We are actively working to address this inconsistency, ensuring that image updates are seamlessly displayed to users without requiring manual intervention.

2. BackEnd

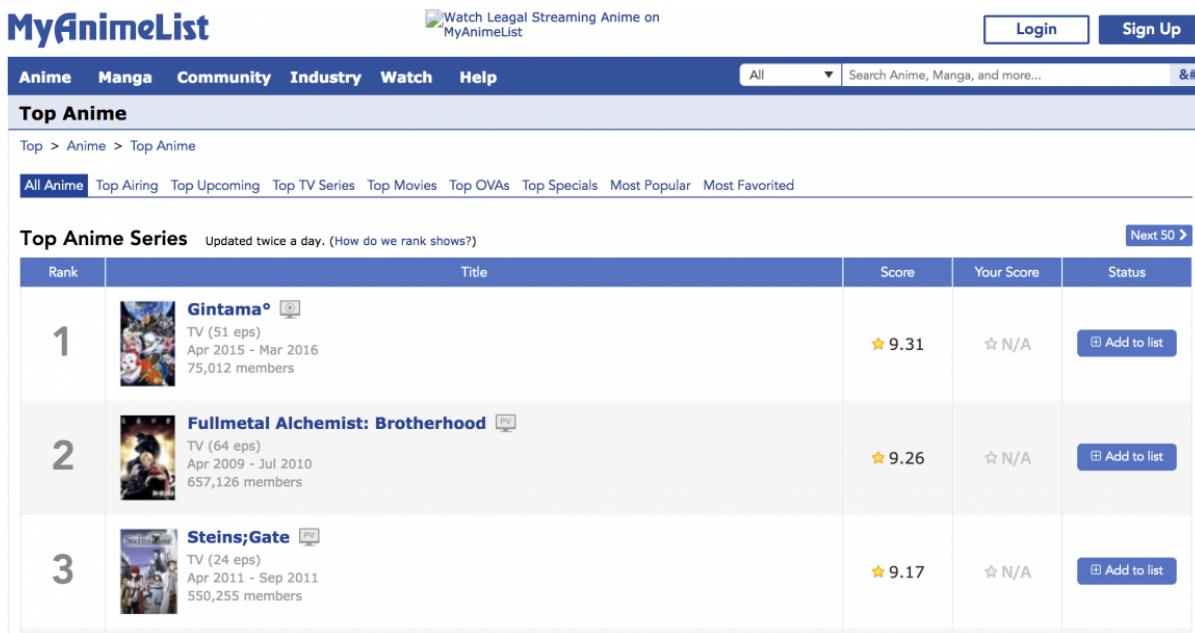
- a. **Elasticsearch and Anaconda Integration:** The necessity to exit and re-enter Elasticsearch when changing images suggests potential issues in the back-end infrastructure or data indexing processes. This problem needs to be resolved to streamline the image update process.
- b. **Localhost Reshaping:** We are looking into localhost reshaping, considering that Anaconda allows localhost to be reshaped. This suggests potential local server environment issues or configurations needing attention. We are dedicated to ensuring a stable and consistent localhost environment for effective development and testing.

In addressing these challenges, we are combining front-end and back-end development efforts, emphasizing debugging and design consistency to enhance our project's overall functionality and user experience.

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

There are other similar anime search engines that focus on indexing and retrieving information from specific domains, such as MyAnimeList (MAL), anime-planet.com, and AniList [1].

MyAnimeList is a popular online database and social networking site dedicated to anime and manga. It provides information on various anime and manga titles, including details about episodes, characters, staff, and user ratings. In essence, MyAnimeList is a comprehensive platform for searching and discovering information related to anime and manga.



The screenshot shows the MyAnimeList website's interface. At the top, there is a navigation bar with links for Anime, Manga, Community, Industry, Watch, and Help. To the right of the navigation bar are buttons for Login and Sign Up. Below the navigation bar, a search bar contains the placeholder text "Search Anime, Manga, and more...". The main content area is titled "Top Anime" and displays a list of the top anime series. The list includes three entries: "Gintama" (Rank 1), "Fullmetal Alchemist: Brotherhood" (Rank 2), and "Steins;Gate" (Rank 3). Each entry includes the title, a thumbnail image, the number of episodes, the broadcast period, and the number of members. The "Gintama" entry has a score of 9.31, while "Fullmetal Alchemist" and "Steins;Gate" both have a score of 9.26. There are also "Add to list" buttons for each entry.

Rank	Title	Score	Your Score	Status
1	Gintama <small>[TV]</small> TV (51 eps) Apr 2015 - Mar 2016 75,012 members	★ 9.31	☆ N/A	<button>Add to list</button>
2	Fullmetal Alchemist: Brotherhood <small>[PV]</small> TV (64 eps) Apr 2009 - Jul 2010 657,126 members	★ 9.26	☆ N/A	<button>Add to list</button>
3	Steins;Gate <small>[PV]</small> TV (24 eps) Apr 2011 - Sep 2011 550,255 members	★ 9.17	☆ N/A	<button>Add to list</button>

Anime-planet.com is a community-driven website that serves as an anime and manga recommendation database. Users can create profiles, track their anime and manga progress, and receive personalized recommendations based on their preferences.

anime planet

Anime Manga Characters Community

sign up log in

Discover new anime
Find your new favorites, track your progress, watch videos.

Popular this week

[+ Add to list](#)

Berserk of Gluttony

[+ Add to list](#)

Ragna Crimson

[+ Add to list](#)

A Playthrough of a Certain Dude's VRMMO Life

[+ Add to list](#)

The Faraway Paladin 2nd Season

[+ Add to list](#)

Dead Mount Death Play: Part II

[SEE ALL POPULAR ANIME](#)

AniList is a platform similar to MyAnimeList, providing users with a comprehensive database for anime and manga. It includes features for tracking watched and read titles, discovering new content, and connecting with other anime and manga enthusiasts. AniList also allows users to create lists, rate titles, and participate in discussions.

A

Search Social Forum Login Sign Up

Search Genres Year Season Format

TRENDING NOW

[View All](#)

Jujutsu Kaisen 2nd Season

Dr. STONE: NEW WORLD Part 2

ONE PIECE

Kage no Jitsuryokusha ni Naritakute! 2nd season

Sousou no Frieren

Implementation

Data collection

The data source is from a public anime database called animenewsnetwork.com.



Figure 1. Anime News Network logo

The website offers a complimentary Application Programming Interface (API) for accessing anime-related data. For additional details, please refer to the official documentation at <https://www.animenewsnetwork.com/encyclopedia/api.php>.

The API returns responses in XML format. An example of the raw data can be observed by accessing the following link:

<https://cdn.animenewsnetwork.com/encyclopedia/api.xml?anime=454>.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<ann>
  <anime id="454" gid="1266224097" type="TV" name="Case Closed" precision="TV" generated-on="2023-11-14T07:12:33Z">
    <related-prev rel="adapted from" id="1974"/>
    <related-next id="2679" rel="sequel"/>
    <related-next id="5140" rel="related"/>
    <related-next id="6489" rel="related"/>
    <related-next id="6706" rel="prequel"/>
    <related-next id="10605" rel="related"/>
    <related-next id="11687" rel="side story"/>
    <related-next id="11690" rel="side story"/>
    <related-next id="12912" rel="side story"/>
    <related-next id="15284" rel="related"/>
    <related-next id="22122" rel="related"/>
    <related-next id="22487" rel="alternate retelling"/>
    <related-next id="23173" rel="related"/>
    <related-next id="24820" rel="spinoff"/>
  <info gid="2750892830" type="Picture" src="https://cdn.animenewsnetwork.com/thumbnails/fit200x200/encyc/A454-1482.jpg" width="200" height="200">
    
    
  </info>
  <info gid="2232689579" type="Main title" lang="EN">Case Closed</info>
  <info gid="211993110" type="Alternative title" lang="AR">آل مهلاكه Конан</info>
  <info gid="1230032721" type="Alternative title" lang="EN">Case Closed: One truth Prevails (English)</info>
  <info gid="1300617104" type="Alternative title" lang="ES">Detectiu Conan (Catalan)</info>
  <info gid="1302888643" type="Alternative title" lang="EN">Detective Conan</info>
  <info gid="2010288027" type="Alternative title" lang="EN">Detective Conan (French)</info>
  <info gid="3507401038" type="Alternative title" lang="EN">Detectivul Conan (Romanian)</info>
  <info gid="3340497745" type="Alternative title" lang="EN">Detektif Conan (Indonesian)</info>
  <info gid="2771474056" type="Alternative title" lang="DE">Detektiv Conan</info>
  <info gid="3023967142" type="Alternative title" lang="PL">Detektyw Conan</info>
  <info gid="2936498159" type="Alternative title" lang="PT">Detetive Conan</info>
  <info gid="1517854442" type="Alternative title" lang="ES">El Detective Conan</info>
  <info gid="948541669" type="Alternative title" lang="EN">Mästerdetektiven Conan (Swedish)</info>
  <info gid="1957755074" type="Alternative title" lang="JA">Meitantei Conan</info>
  <info gid="1669176571" type="Alternative title" lang="EN">Mesterdetektivien Conan (Danish)</info>
  <info gid="3897716987" type="Alternative title" lang="EN">Salapollisi Conan (Finnish)</info>
  <info gid="1727853426" type="Alternative title" lang="EN">Thám tử lừng danh Conan (Vietnamese)</info>
  <info gid="2255894335" type="Alternative title" lang="RU">Детектив Конан</info>
  <info gid="2589267448" type="Alternative title" lang="AR">المحقق كونان</info>
  <info gid="225558012" type="Alternative title" lang="ZH-TW">名偵探柯南</info>
  <info gid="849520857" type="Alternative title" lang="JA">名探偵コナン</info>
```

Subsequently, the XML document is processed using Python to transform it into NDJSON (New Line Delimited JSON) format, facilitating subsequent indexing and ranking via Elasticsearch. The extraction process involves retrieving essential metadata, including **title**, **plot summary**, **genres**, **themes**, and **image URLs (primarily for front-end display)**. Provided below is a snippet of sample Python source code illustrating the data preprocessing steps.

```
 1  def search_anime(animeID):
 2      api_url = f"https://cdn.animenewsnetwork.com/encyclopedia/api.xml?anime={animeID}"
 3
 4      response = requests.get(api_url)
 5      if response.status_code == 200:
 6          xml_data = response.text
 7          root = ET.fromstring(xml_data)
 8
 9          # Initialize variables with default values
10          title = None
11          plot_summary = None
12          genres = []
13          themes = []
14          image_url = None
15
16          # Extract data (title, plot summary, genres, themes, image URL)
17          title_element = root.find("./info[@type='Main title']")
18          if title_element is not None:
19              title = title_element.text
20
21          plot_summary_element = root.find("./info[@type='Plot Summary']")
22          if plot_summary_element is not None:
23              plot_summary = plot_summary_element.text
24
25          genre_elements = root.findall("./info[@type='Genres']")
26          genres = [genre.text for genre in genre_elements]
27
28          theme_elements = root.findall("./info[@type='Themes']")
29          themes = [theme.text for theme in theme_elements]
30
31          image_elements = root.findall("./info[@type='Picture']/img")
32          if image_elements is not None:
33              num_images = len(image_elements)
34              if num_images > 0:
35                  big_image = image_elements[num_images-1]
36                  image_url = big_image.get("src")
37
38          # Check if at least the title and plot summary have values
39          if title is not None and plot_summary is not None:
40              # Create a JSON object
41              anime_data = {
42                  "animeID": animeID,
43                  "title": title,
44                  "plot_summary": plot_summary,
45                  "genres": genres,
46                  "themes": themes,
47                  "image_url": image_url
48              }
49
50              return json.dumps(anime_data, separators=(',', ':')) # NDJSON format
51
52      return None
```

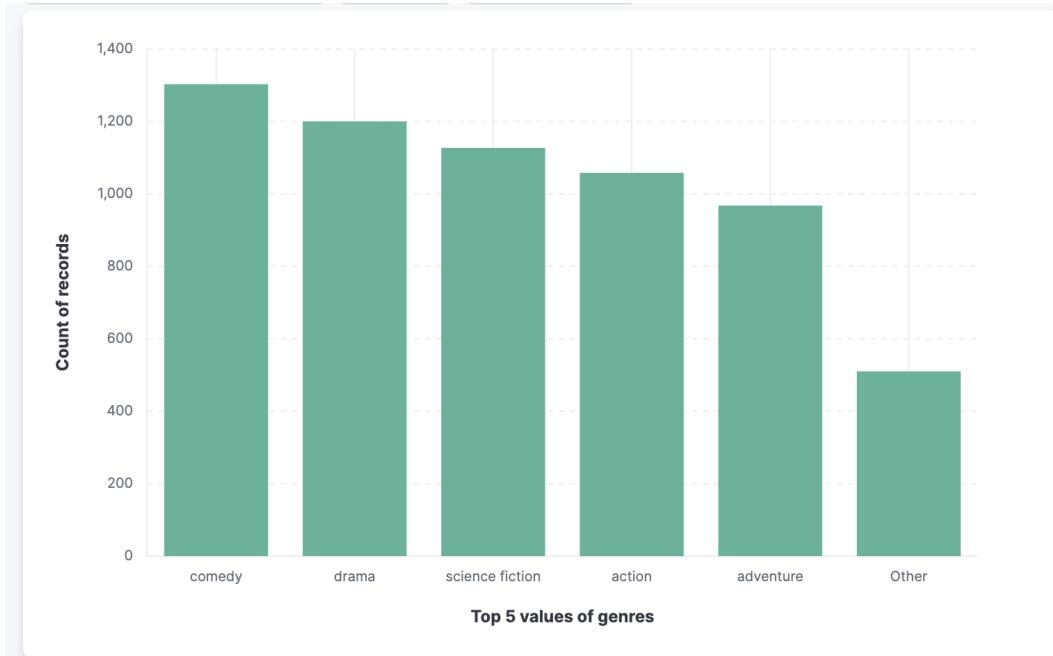
Example documents

Here are 3 examples of extracted raw data into the NDJSON document. In this project, 3244 anime data were collected.

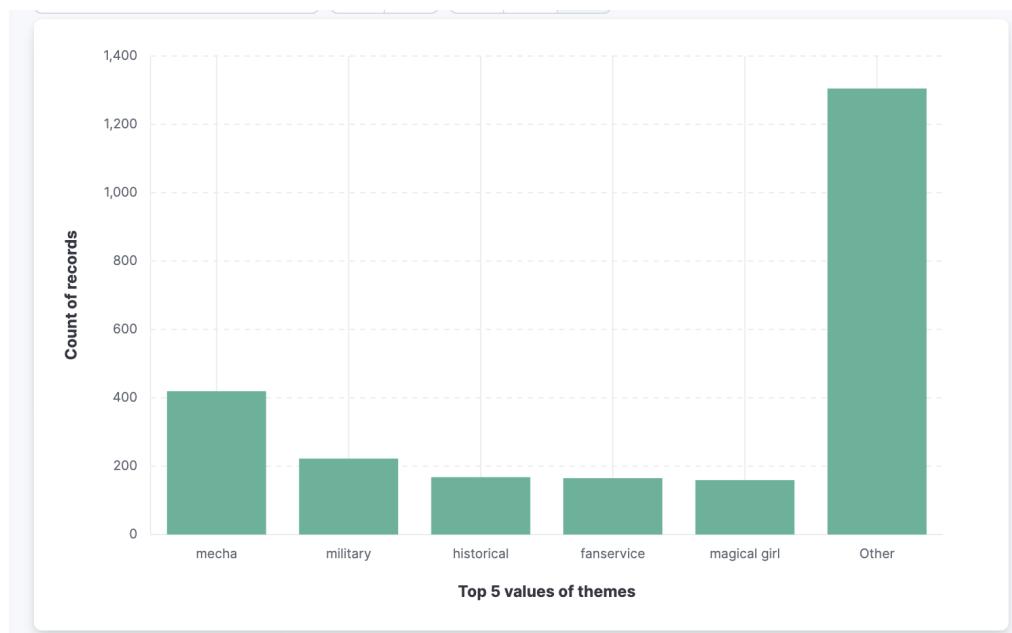
```
{"animeID":1,"title":"Angel Links","plot_summary":"Li Meifon is the head of a free, no expenses paid protection/security agency for escorting ships across outer space. Though only 16, she possesses great skill and leadership qualities while commanding her crew aboard their ship, the <i>Angel Links</i>. However, some unpleasant memories begin to trouble her and she starts questioning her own past and reason for existence. Join Li Meifon and the rest of the <i>Angel Links</i> crew on their adventures through the Oracion star system as they battle pirates, government organizations, and a Tao master, all while finally uncovering Meifon's dark and mysterious past.", "genres":["action","drama","science fiction"],"themes":["pirates","space"],"image_url":"https://cdn.animenewsnetwork.com/thumbnails/fit200x200/encyc/A1-51.jpg"}  
{"animeID":2,"title":"Apocalypse Zero","plot_summary":"Kakugo and Harara are two siblings who have been trained to fight demons, monsters, and other creatures that are now roaming about in the post-apocalyptic 21st century. Along with their superior fighting techniques and specialized skills, they were also given special suits called Zero armor, which were made from the souls of dead warriors. Unfortunately Harara succumbs to evil within the Zero armor, and Kakugo is the only human alive that has the skills and equipment to defeat his sibling.", "genres":["action","science fiction"],"themes":["demons","gore","post-apocalyptic"],"image_url":"https://cdn.animenewsnetwork.com/thumbnails/max500x600/encyc/A2-9.jpg"}  
{"animeID":3,"title":"Nasu: Summer in Andalusia","plot_summary":"Pepe is a Spanish cyclist competing in a multi-stage Iberian cycling race similar to the Tour de France. He is a support rider for one of the teams competing in the race, and his role is to assist the team's top rider in winning the overall race. The racers are set to ride through Pepe's hometown in Andalusia on the same day of the wedding of his elder brother Angel to his former girlfriend Carmen.", "genres":["tournament"],"themes":["sports"],"image_url":"https://cdn.animenewsnetwork.com/thumbnails/max500x600/encyc/A3-4.jpg"}
```

Data statistics

From our collected data, the top 5 genres that appear the most in the documents are comedy, drama, science fiction, action, adventure, and others respectively.



On the other hand, theme is the section that contains noticeable uniqueness. There are many unique terms which are useful for information retrieval. However, the top 5 themes are mecha, military, historical, fanservice, and magical girl respectively.



Tools and software

Elasticsearch and Kibana



Elasticsearch and Kibana for data management and real-time analytics including document indexing and ranking. Kibana is essential for us to access Elasticsearch and upload our NDJSON document. Elasticsearch is essential to both indexing and retrieving documents based on users' queries. Here is the fully-modified query string that we used for Elasticsearch to retrieve and rank documents efficiently

Python

```
'multi_match': {  
    'query': keyword,  
    'fields': ['title^3', 'plot_summary^2', 'genres^1', 'themes^1'],  
    'fuzziness': 'AUTO'  
}  
}
```

In this query string, we rank our documents based on Title, Plot Summary, Genres, and Themes. We also include different weights assigned to each criterion. This can help us identify the importance of each criterion. In our case, the Title should contain the highest weight followed by Plot Summary, Genres and Themes (Title > Plot Summary > Genres = Themes). Moreover, to improve efficiency, we

also enable fuzziness to the search query so that typos are accepted which is practical for real usage.

Flask and Python



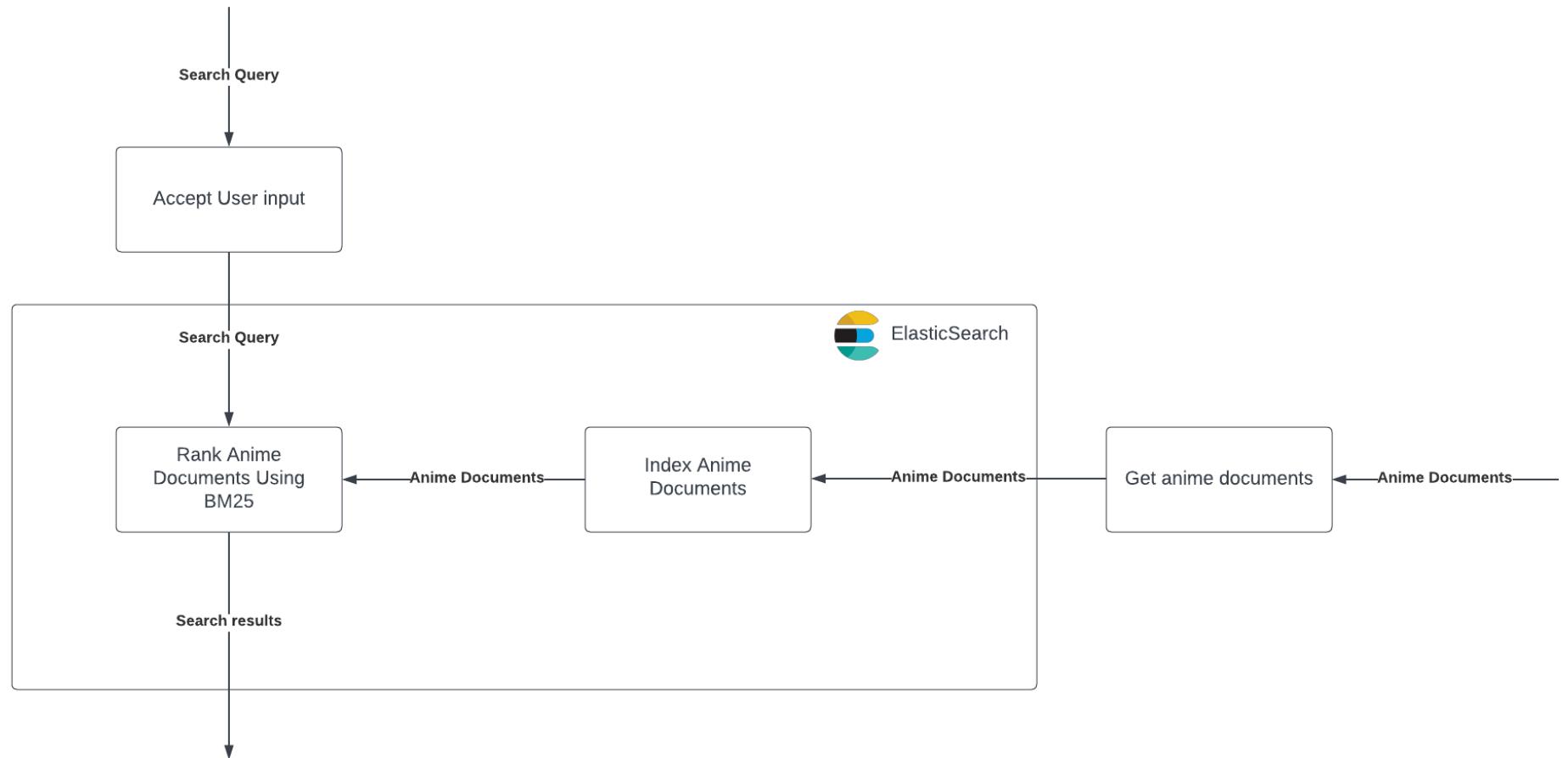
Flask is used to develop the actual implementation (mock-up prototype) of our search system. It provides flexibility to develop a web application with the integration of Elasticsearch (get search results to display on an HTML page) using Python programming language.

Git and GitHub



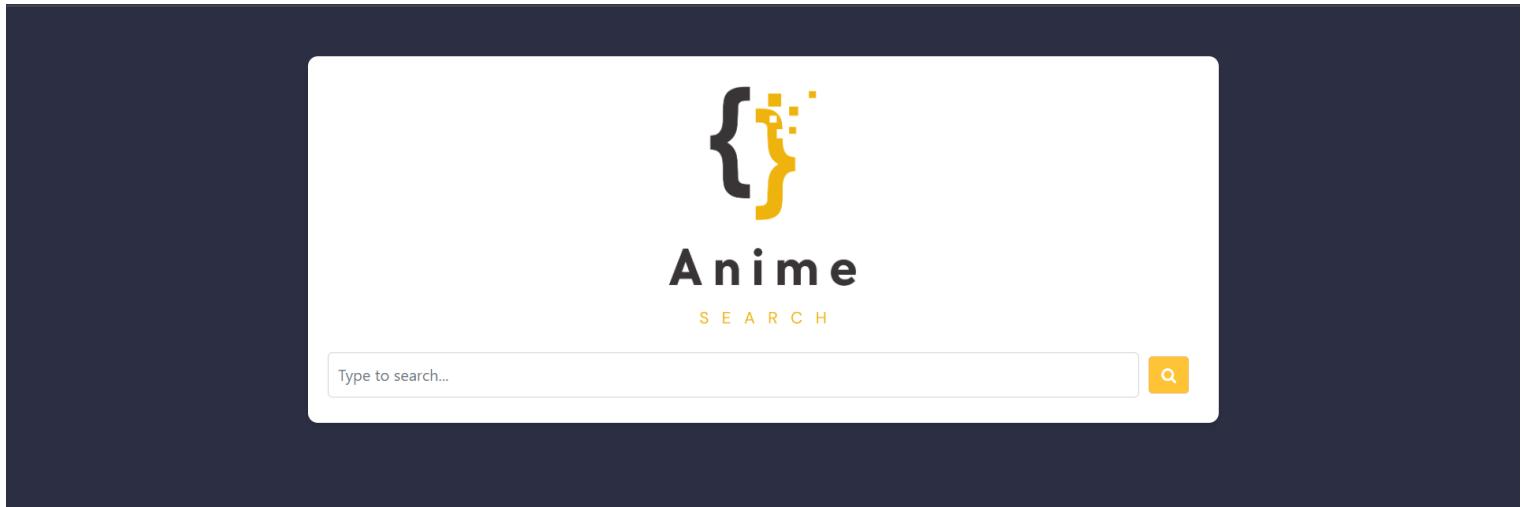
Git and GitHub are essential for version control, code collaboration, and backup. These tools provide us with the capabilities to efficiently manage data and code, collaborate with team members, and make informed decisions based on data-driven insights.

System diagram



Snapshots of the system

Home Page: Users can type in the search box to search for their desired anime.



Search Result Page: Search results will be displayed in pages (10 items each).

A screenshot of the search results page. The top section features the same logo and search bar as the homepage. Below the search bar, there are three search results, each represented by a horizontal card.

- Detective Conan: Conan vs Kid vs YAIBA**
Conan borrows a large amount of manga from Mitsuhiro and crazy things begin to happen around Beika City. Things aren't always the way they appear to be. A family asks Kogoro to help them protect a special sword in their dojo since Kaitou KID has sent word he's going to steal it. However Kaitou KID will have to steal the sword while Conan and Yaiba are after him.
Genres: adventure, comedy, drama, mystery
[More Info](#)
- Detective Conan: Conan and Heiji and the Vanished Boy**
Conan and Ran are visiting Heiji and Kazuha in Osaka but while they are touring around they come across a group of three kids that look like Ayumi, Mitsuhiro, and Genta. The group are looking for their missing friend and Conan agrees to help.
Genres: comedy, mystery
[More Info](#)
- Detective Conan: The Private Eyes' Requiem**
Some unknown man who knows Conan is in fact Kudo Shinichi lures Conan and the others (Mori, Ran, Tanteidan...) to Yokohama. This man takes Ran and the Tanteidan as hostages. He gives Conan a 12 hour time limit in which to solve a case. If it is not solved within that time, bombs attached to the hostages will explode.... To thwart this villain Conan has help from Heiji, Kaitou Kid, Hakuba.
Genres: adventure, drama, mystery, thriller
[More Info](#)



Detective Conan: A Challenge from Agasa

While creating a elaborate goose chase for the detective boys Professor Agasa becomes ensnared in a ransom plot and is abducted. It's up to Conan to deduce his location.

Genres: action, mystery

[More Info](#)



Detective Conan: Magician of the Silver Sky

Kaito Kid strikes again in this new annual installment of the Detective Conan movies. An actress asked for Mouri Kogoro to protect a precious jewel of hers which Kid has vowed to steal. On the day of the theft, Kaito Kid dressed up as Shinichi and matched wits with Conan, and fled in the end. To thank them, the actress invited Kogoro and family and friends to Sapporo, but a bigger scheme, and a great emergency is just about to unravel high above the clouds in the plane that they're taking...

Genres: action, drama, mystery

[More Info](#)



Detective Conan: Crossroad in the Ancient Capital

Mouri Kogoro is called to a special case in the ancient capital of Kyoto. There, Conan meets Heiji and they team up once again to solve the case, recover the stolen Healing Buddha statue, and even discover the identity of Heiji's first love.

Genres: action, mystery, romance

[More Info](#)

[1](#) [2](#) [»](#) [Last](#)

When users click on “More info”, the system will redirect users to the information page of that anime on animenewsnetwork.com, which is the original source of our anime data.

[News](#) [Views](#) [New Anime](#) [Encyclopedia](#) [Forum](#) [My ANN](#) [About](#)

Subscribe to the ANN Newsletter • Wake up every Sunday to a curated list of ANN's most interesting posts of the week. [read more](#) ×

Detective Conan: Conan vs Kid vs YAIBA (OAV)

[!\[\]\(22750f03d825e3229d78f80f6f680b4d_img.jpg\)](#)

Have you seen this? [want to](#) / [seen some](#) / [seen all](#)

[related to [Case Closed \(TV\)](#)]

Related anime:

- [Detective Conan: 16 Suspects \(OAV\)](#) (sequel)
- [Detective Conan: Conan and Heiji and the Vanished Boy \(OAV\)](#) (sequel)
- [Detective Conan: Conan and Kid and Crystal Mother \(OAV\)](#) (sequel)
- [Detective Conan: Hyōteki wa Kogorō Shōnen Tanteidan Maruchichōsa \(OAV\)](#) (sequel)
- [Detective Conan: Conan Edogawa & Heiji Hattori versus Kid the Phantom Thief \(OAV\)](#) (sequel)
- [Detective Conan: A Challenge from Agasa \(OAV\)](#) (sequel)

Alternative title:

Metsanet Conan: Conan VS KID VS Yaiba (Japanese)
 名探偵コナン: Conan VS KID VS YAIBA (Japanese)
 名探偵コナン: コナンvsキッドvsヤイバ 宝刀争奪大決戦!! (Japanese)

Genres: adventure, comedy, drama, mystery
Themes: Crossover
[find similar anime based on genres & themes](#)

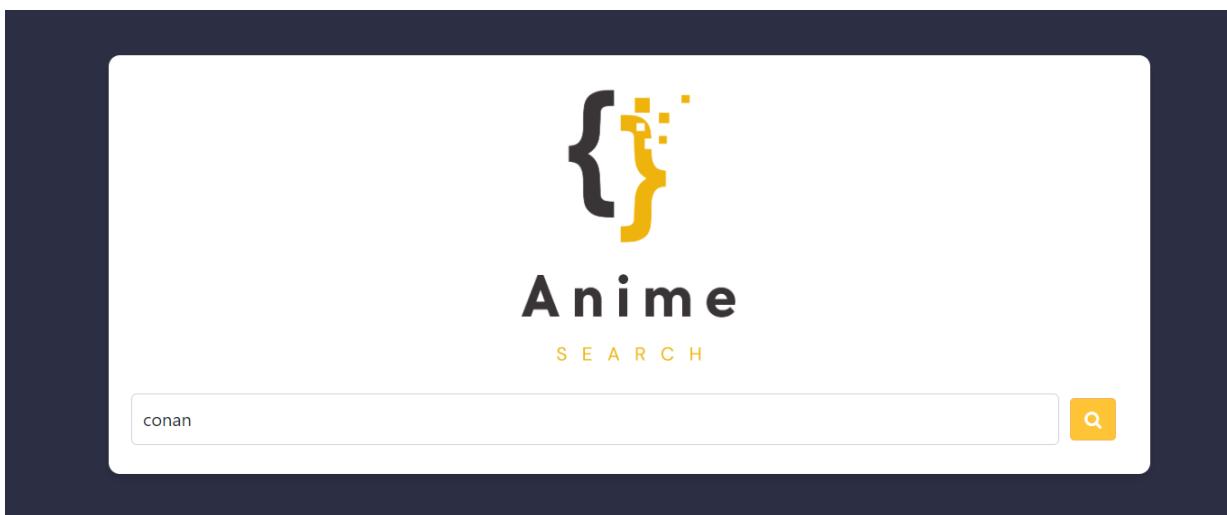
Plot Summary: Conan borrows a large amount of manga from Mitsuhiko and crazy things begin to happen around Beika City. Things aren't always the way they appear to be. A family asks Kogoro to help them protect a special sword in their dojo since Kaitou KID has sent word he's going to steal it. However Kaitou KID will have to steal the sword while Conan and Yaiba are after him.



Example step-by-step search sessions that highlight the following functionality:

1. One word query

Let's start with the **search term: “conan”**. In this case, we would like to search for Detective Conan's anime

A screenshot of the search results page for the query "conan". The page has a dark background with a white header featuring the Anime Search logo and a search bar. Below the header, there are four search results, each with a small thumbnail image, the title, a brief description, genre information, and a "More Info" button.

- Detective Conan: Conan vs Kid vs YAIBA**
Conan borrows a large amount of manga from Mitsuhiko and crazy things begin to happen around Beika City. Things aren't always the way they appear to be. A family asks Kogoro to help them protect a special sword in their dojo since Kaitou KID has sent word he's going to steal it. However Kaitou KID will have to steal the sword while Conan and Yabai are after him.
Genres: adventure, comedy, drama, mystery
[More Info](#)
- Detective Conan: Conan and Heiji and the Vanished Boy**
Conan and Ran are visiting Heiji and Kazuha in Osaka but while they are touring around they come across a group of three kids that look like Ayumi, Mitsuhiko, and Genta. The group are looking for their missing friend and Conan agrees to help.
Genres: comedy, mystery
[More Info](#)
- Detective Conan: The Private Eyes' Requiem**
Some unknown man who knows Conan is in fact Kudo Shinichi lures Conan and the others (Mori, Ran, Tanteidan...) to Yokohama. This man takes Ran and the Tanteidan as hostages. He gives Conan a 12 hour time limit in which to solve a case. If it is not solved within that time, bombs attached to the hostages will explode... To thwart this villain Conan has help from Heiji, Kaitou Kid, Hakuba.
Genres: adventure, drama, mystery, thriller
[More Info](#)
- Detective Conan: Strategy Above the Depths**
On a luxurious cruise trip paid for by Sonoko, Conan and his gang enjoy the high life on the open seas. However, a shrewd plot for revenge was brewing right under Conan's nose. And this time Mori takes the spotlight, proving his own intuitions right, and beats Conan at his own deduction.
Genres: adventure, drama, mystery
[More Info](#)

The search engine returns the expected result of multiple “Detective Conan” animes. Since this is a one-word query, there **can be some non-relevant documents** retrieved. For example,



Future Boy Conan

Based on The Incredible Tide by Alexander Key. Post nuclear future. Earth has been almost completely submerged by oceans. Indastria, an hyper-technological isle, is the new world capital. Conan is the only survived boy in the attempt of a bunch of astronauts to escape the earth when nuke bombs were falling. He will meet a girl named Lana (Italian name), and when she will be kidnapped by Indastria's soldiers, he'll go and try to rescue her.

Genres: adventure, drama, science fiction

[More Info](#)



Case Closed: The Phantom of Baker Street

Conan was invited to a party with many other guests. They were there to test a new virtual game system, when a murder mystery occurred. Now Conan must go into the game system to figure out who the murderer is with the help of the famous book character Sherlock Holmes. The lives of 49 kids testing the game are in his hands.

Genres: adventure, drama, mystery, science fiction, thriller

[More Info](#)



Case Closed: Countdown to Heaven

As a pair of towers in Tokyo are being prepared for their grand opening, there is a series of murders of people connected to the towers. Conan suspects that the mysterious Syndicate may also be involved.

Genres: comedy, drama, mystery, thriller

[More Info](#)



Case Closed: The Last Wizard of the Century

Kaitou Kid announces to the police that he intends to steal the Russian Imperial Easter Egg, which is currently being held in Osaka. As Conan pursues his rival, he discovers that there's more to this case than simply stopping a robbery.

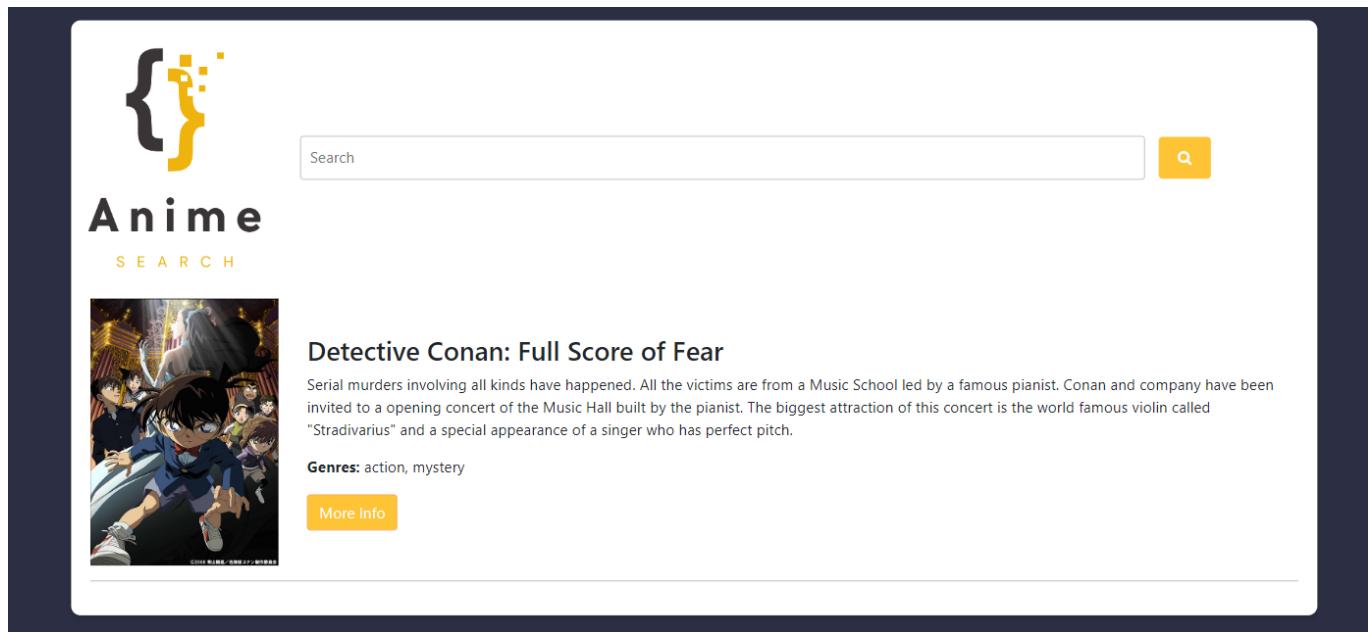
Genres: adventure, drama, mystery

[More Info](#)

The search engine also retrieved the animes of Detective Conan but written under different titles because the term “conan” appears in the plot summary. In this case, some Conan animes are named “Cased Closed”. However, these documents

will receive lower ranking scores as terms that appear in the title will have more weight than in the plot summary.

If users input a **rare term**, the search system can even scope down to improve efficiency. This is the example of when users put in the term “**stradivarius**”



2. Multiple word query

Users can also type in multiple words to search for an anime. For example, when users type in “**robotic boy with super power**”, they can expect to see the result of “Astro Boy”.

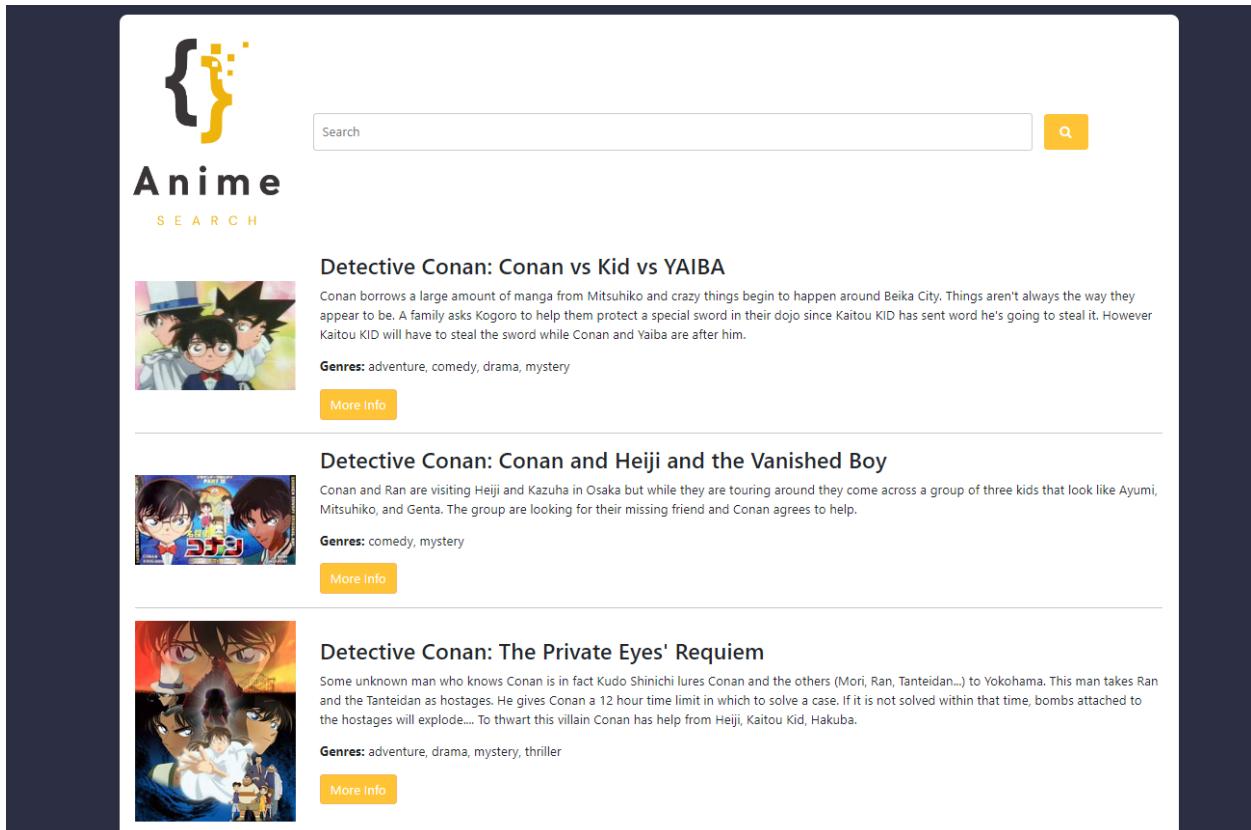
The screenshot shows the Anime Search website interface. At the top left is the logo, which consists of a stylized 'A' made of orange and yellow pixels. To its right is a search bar with the placeholder "Search" and a magnifying glass icon. Below the search bar is a list of anime entries. The first entry is "Astro Boy", featuring a thumbnail image of the robot boy in a dynamic pose. The title "Astro Boy" is in bold black font. A brief description follows: "Astro is a robotic boy that possesses super human powers and an artificial intelligence system that is unparalleled to any robot. His creator, Dr. Tenma, created him to replace his late son, Tovio. Dr. Tenma soon destroys his laboratory, after the creation of Astro, and shuts down Astro. Soon after, Dr. O'Shay, the head of the Ministry of Science revives Astro, and tries to give him a normal life as a 6th grade student that helps the police agency keep renegade robots and bigot humans from causing harm. Astro faces extreme racism for being a robot, and he must discover the truth about his creator Dr. Tenma." Below the description, the genres "action, science fiction" are listed, followed by a "More Info" button. The second entry is "Power Dolls", with a thumbnail image of three female mech pilots. The title "Power Dolls" is in bold black font. A brief description follows: "It's the 25th century and Planet Omni calls in a elite team of female Mech Pilots Called the Detachment Of Limited Line Service or The Power Dolls to Handle all foes." Below the description, the genres "adventure, science fiction" are listed, followed by a "More Info" button.

“pirate king” should return the result that contains “One Piece”.

The screenshot shows the Anime Search website interface. At the top left is the logo, which consists of a stylized 'A' made of orange and yellow pixels. To its right is a search bar with the placeholder "Search" and a magnifying glass icon. Below the search bar is a list of anime entries. The first entry is "One Piece", featuring a thumbnail image of the main cast on a boat. The title "One Piece" is in bold black font. A brief description follows: "Before he was executed, the legendary Pirate King Gold Roger revealed that he had hidden the treasure One Piece somewhere in the Grand Line. Now, many pirates are off looking for this legendary treasure to claim the title Pirate King. One pirate, Monkey D. Luffy, is a boy who had eaten the Devil's Fruit and gained rubber powers. Now he and his crew are off to find One Piece, while battling enemies and making new friends along the way." Below the description, the genres "adventure, comedy, drama, fantasy" are listed, followed by a "More Info" button. The second entry is "Jing: King of Bandits", with a thumbnail image of Jing performing a move. The title "Jing: King of Bandits" is in bold black font. A brief description follows: "Jing may appear to be a young boy, but his remarkable skills make him one of the most feared thieves on the planet. Along with his feathered partner Kir, Jing travels from town to town, stealing anything of value regardless of the amount of security. But when he's in a pinch, he has one more trick up his sleeve: Kir bonds with Jing's right arm to perform the effectively deadly "Kir Royale" attack. And because of all this, Jing is infamously known by many as the "King of Bandits." Below the description, the genres "adventure, comedy, fantasy, science fiction" are listed, followed by a "More Info" button.

3. Partial match

The search system also supports “**fuzziness**”. If the user makes a typo and types in “conam” instead of “conan”, the search results still remain relevant to the user’s information need.



4. Ranking

Elasticsearch uses BM25 to rank documents. The ranking score can be seen in either Kibana tool or Python source code.

Ranking score for search query “detective conan”. Our system then uses this score to rank the order of each retrieved document.

```
Score: 30.177303, Source: {'animeID': 5141, 'title': 'Detective Conan: 16 Suspects', 'plot_summary': "The detective boys, the Mouri family, Professor Agasa, Heiji, Sonoko, Kazuha, and some of the polices were invited to Inspector Shiratori's party at Karuizawa. A certain case happens, and now there are 16 suspects from those invited to the party.", 'genres': [], 'themes': [], 'image_url': 'https://cdn.animenewsnetwork.com/thumbnails/fit200x200/encyc/A5141-14.jpg'}
Score: 28.700304, Source: {'animeID': 5140, 'title': 'Detective Conan: Conan vs Kid vs YAIBA', 'plot_summary': "Conan borrows a large amount of manga from Mitsuhiro and crazy things begin to happen around Beika City. Things aren't always the way they appear to be. A family asks Kogoro to help them protect a special sword in their dojo since Kaitou KID has sent word he's going to steal it. However Kaitou KID will have to steal the sword while Conan and Yaiba are after him.", 'genres': ['adventure', 'comedy', 'drama', 'mystery'], 'themes': [], 'image_url': 'https://cdn.animenewsnetwork.com/thumbnails/max500x600/encyc/A5140-6.jpg'}
Score: 26.800867, Source: {'animeID': 5143, 'title': 'Detective Conan: Conan and Kid and Crystal Mother', 'plot_summary': "Kaito and Aoko are traveling on a train along with Aoko's father, Inspector Nakamori, in attempts to help stop Kaitou KID from stealing a precious gem known as Crystal Mother. Little do Aoko and Inspector Nakamori know that Kaito is Kaitou KID and he will do whatever it takes to retrieve the gem. Unfortunately for Kaito, he is also being followed by Conan and a little boy named Philip who will do anything to stop him.", 'genres': [], 'themes': [ ], 'image_url': 'https://cdn.animenewsnetwork.com/thumbnails/max500x600/encyc/A5143-8.jpg'}
Score: 25.163357, Source: {'animeID': 8767, 'title': 'Detective Conan: A Challenge from Agasa', 'plot_summary': "While creating a elaborate goose chase for the detective boys Professor Agasa becomes ensnared in a ransom plot and is abducted. It's up to Conan to deduce his location.", 'genres': ['action', 'mystery'], 'themes': [], 'image_url': 'https://cdn.animenewsnetwork.com/thumbnails/max500x600/encyc/A8767-3.jpg'}
Score: 25.154741, Source: {'animeID': 5142, 'title': 'Detective Conan: Conan and Heiji and the Vanished Boy', 'plot_summary': "Conan and Ran are visiting Heiji and Kazuha in Osaka but while they are touring around they come across a group of three kids that look like Ayumi, Mitsuhiro, and Genta. The group are looking for their missing friend and Conan agrees to help."}
```



Detective Conan: 16 Suspects

The detective boys, the Mouri family, Professor Agasa, Heiji, Sonoko, Kazuha, and some of the polices were invited to Inspector Shiratori's party at Karuizawa. A certain case happens, and now there are 16 suspects from those invited to the party.

Genres:

[More Info](#)



Detective Conan: Conan vs Kid vs YAIBA

Conan borrows a large amount of manga from Mitsuhiro and crazy things begin to happen around Beika City. Things aren't always the way they appear to be. A family asks Kogoro to help them protect a special sword in their dojo since Kaitou KID has sent word he's going to steal it. However Kaitou KID will have to steal the sword while Conan and Yaiba are after him.

Genres: adventure, comedy, drama, mystery

[More Info](#)



Detective Conan: Conan and Kid and Crystal Mother

Kaito and Aoko are traveling on a train along with Aoko's father, Inspector Nakamori, in attempts to help stop Kaitou KID from stealing a precious gem known as Crystal Mother. Little do Aoko and Inspector Nakamori know that Kaito is Kaitou KID and he will do whatever it takes to retrieve the gem. Unfortunately for Kaito, he is also being followed by Conan and a little boy named Philip who will do anything to stop him.

Genres:

[More Info](#)



Detective Conan: A Challenge from Agasa

While creating a elaborate goose chase for the detective boys Professor Agasa becomes ensnared in a ransom plot and is abducted. It's up to Conan to deduce his location.

Genres: action, mystery

[More Info](#)



Detective Conan: Conan and Heiji and the Vanished Boy

Conan and Ran are visiting Heiji and Kazuha in Osaka but while they are touring around they come across a group of three kids that look like Ayumi, Mitsuhiro, and Genta. The group are looking for their missing friend and Conan agrees to help.

Genres: comedy, mystery

[More Info](#)

Discussion

Limitations of your system

Elasticsearch

1. **Data Consistency:** While Elasticsearch is excellent for full-text search and analytics, maintaining data consistency across multiple indices or data sources can be a challenge. Keeping data in sync and ensuring accuracy is important for reliable analysis.
2. **Scalability:** One of the limitations we have encountered is related to the scalability of the system. As data volume and user activity grow, Elasticsearch may face challenges in handling the increased load efficiently. Ensuring scalability while maintaining performance can be a complex task.

Git and GitHub

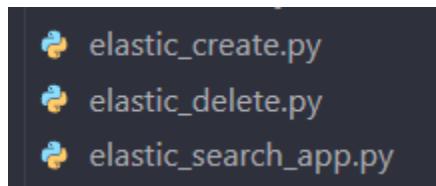
1. **Version Control:** When using Git and Github for code version control, managing large repositories and resolving merge conflicts can sometimes be time-consuming and require careful coordination among team members.

Technical difficulties, challenges, and lessons learned

1. **Query Optimization:** Writing efficient Elasticsearch queries can be challenging, particularly for complex use cases. We've learned that query optimization is an ongoing process, and constant monitoring and refinement

are necessary. In this case, we've added fuzziness to improve efficiency. Different weights are also assigned to different sections of each document. However, there can still be some improvement to the query string, such as adding synonyms, and more optimization.

2. **Difficulties accessing Kibana:** We've had difficulties using Kibana in our project since most of the time Kibana took so much time to load up and often contained errors. However, we overcame this by using Python to communicate with Elasticsearch instead of Kibana.



3. **Long Data Collection Period:** Since Animenewsnetwork's API only allows 1 request per second per IP address, it took a long time to collect sufficient anime information to apply to this project.
4. **HTML/XML Tags not filtered:** Some documents contain HTML or XML tags. For instance,

Doraemon



Nobita Nobi is so hapless that his 22nd century descendants are still impoverished as a result of his 20th century bumbling. In a bid to raise their social status, their servant, a robotic cat named Doraemon, decides to travel back in time and guide Nobita on the proper path to fortune. Unfortunately Doraemon, a dysfunctional robot that the family acquired by accident (but chose to keep nonetheless), isn't much better off than Nobita. The robot leads Nobita on many adventures, and while Nobita's life certainly is more exciting with the robot cat from the future, it is questionable if it is in fact better in the way that Doraemon planned.

Genres: action, comedy, science fiction

[More Info](#)

5. **Git Workflow:** Managing a Git workflow with multiple contributors can lead to conflicts and challenges in code integration. We've found that clear branching and merging strategies, along with effective communication, are crucial.

6. **Monitoring and Maintenance:** Regular monitoring of Elasticsearch clusters and GitHub repositories is essential. We've realized the importance of proactive maintenance and automating routine tasks to ensure system health.

Opportunities for future improvements

1. **Automated Testing:** Implementing automated testing and continuous integration for code in Github to minimize integration issues and ensure code quality.
2. **Data Quality Control:** Enhancing data consistency and quality control measures to reduce inaccuracies and maintain reliable analytics.
3. **Enhanced Security:** Improving security measures, such as role-based access control and encryption, to safeguard sensitive data stored in Elasticsearch and GitHub repositories.
4. **Scalability Planning:** Developing a comprehensive scalability plan to handle future data growth efficiently and without performance degradation.

Conclusion

In conclusion, the integration of advanced tools and software has significantly enhanced our data management, analysis, and code collaboration processes. These technologies have empowered us to efficiently process data, gain insights, collaborate effectively, and maintain code integrity. This strategic use of technology ensures that our organization remains competitive and agile in an ever-evolving digital landscape.

Reference

- [1] Anisearch.com competitors - top sites like anisearch.com | similarweb,
<https://www.similarweb.com/website/anisearch.com/competitors/> (accessed Nov.
20, 2023).