## ITCS441 Information Retrieval and Text Mining

## Faculty of Information and Communication Technology

# Search engine system for Sanrio Characters

**Submitted by**

| | |
|---|---|
| 6688072 Kemjira | Nugboon |
| 6688099 Nuttanan | Reamprasert |
| 6688129 Patsatraporn | Thongdeesakul |
| 6688167 Thamonwan | Suebma |
| 6688212 Natpapavee | Pramualsuk |

**Presented to**

**Assoc. Prof. Dr. Suppawong Tuarob**

## Table of contents

# Introduction

This project aims to create a search engine system where users can search for Sanrio characters efficiently by attributes such as name, color, species, personality, friends, and description. The system demonstrates a full information retrieval pipeline including data collection, preprocessing, indexing, ranking, and search interface. It was developed for the ITCS441 Information Retrieval and Text Mining course and includes important details for each character, such as images, names, debut dates, and other useful information.

Sanrio has been popular with people of many ages for a long time. However, there are very few websites that give clear and complete information about all characters with a simple and friendly interface. Many fans also find it difficult to search for accurate and organized resources about their favorite characters. This project aims to solve this problem by creating a search engine made specifically for Sanrio characters.

The main goal of this project is to build a search engine that offers a smooth and effective way for users to explore Sanrio characters. Users can search by name, description, debut year, creator, species, and gender, allowing them to quickly find the information they need. We also aim to provide a clean and easy-to-use interface with high-quality images and simple navigation that can support both dedicated fans and casual viewers.

In conclusion, this project not only completes the requirements of the ITCS441 course but also aims to provide a helpful and user-friendly platform where fans can easily access correct and complete information about Sanrio characters. We believe this system can improve the fan experience and help users feel more connected to the Sanrio world. Although this project was first designed for academic purposes, we believe it has the potential to become a valuable tool for the Sanrio community. It shows our interest in improving user experience and our passion for technology and creativity.

## Problem We Are Trying to Solve

### 1. Difficulty for fans and collectors to find characters based on traits or partial information.

Many users want to search for Sanrio characters using only a few details, such as color, animal type, size, or accessories. However, most websites do not support this type of trait-based search. For example, searching for “pink rabbit” or “a character with a red bow” usually does not return accurate results. In addition, if users remember only part of a character’s name, such as “Kit,” the system may not be able to match it with “Hello Kitty.” This makes the search process slow and inconvenient for fans and collectors.

### 2. Existing websites provide only simple keyword search without semantic understanding.

Most platforms related to Sanrio characters still rely on basic keyword matching. They cannot understand the meaning behind the user’s search or recognize similar terms. As a result, the search results are often incomplete or not relevant, especially when users do not know the correct spelling or use general descriptions instead of exact names.

### 3. Lack of unified search across multiple data sources

Information about Sanrio characters is spread across many websites, such as the official Sanrio site, fan wikis, and merchandise platforms. Each source presents different details, and some of them may be outdated or incomplete. Users often need to check many websites to gather full information, which is time-consuming. Without a unified search system that collects data from different sources, it is difficult for users to find reliable and complete character information in one place.

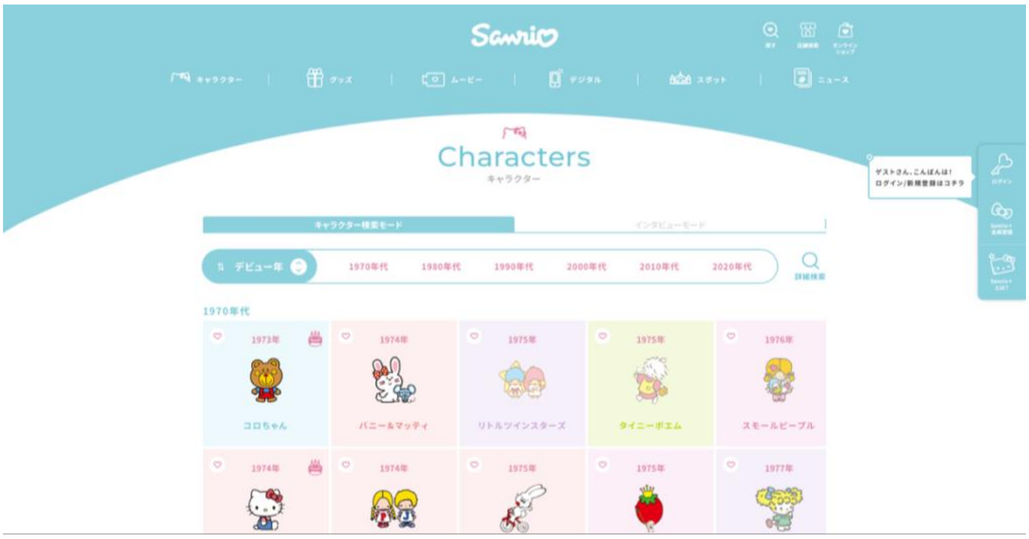
### 4. No ranking mechanism to prioritize relevant or popular characters.

Even when search results appear, most platforms do not provide ranking based on relevance or popularity. This means users may have to scroll through many unrelated characters before finding the one they want. For example, searching for “cat” might show characters with only weak connections before the ones that match the search best. Without a proper ranking system, the search experience becomes less efficient and less user-friendly.

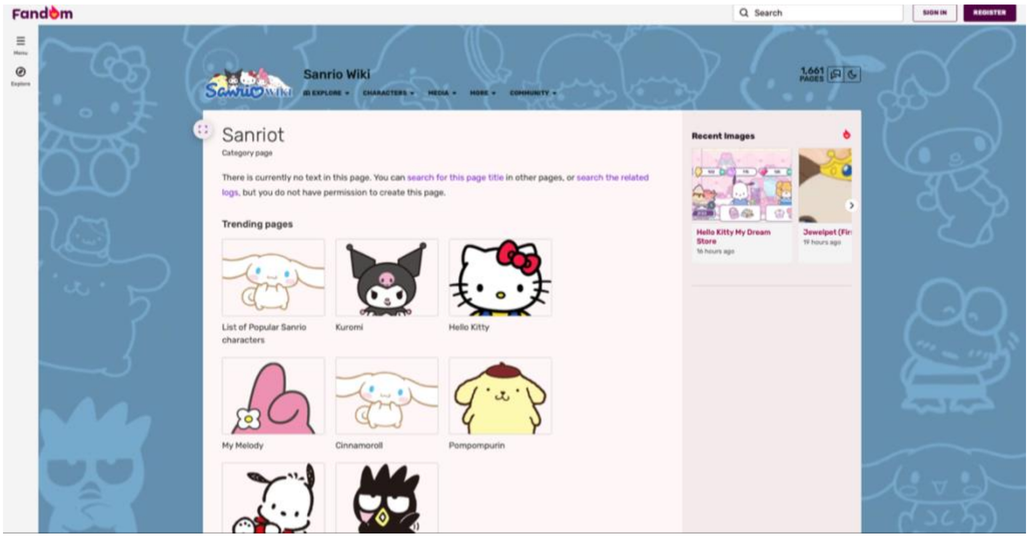
## Existing relevant systems

1. **Official Sanrio website search (supports simple keyword search)** The official Sanrio website lets users browse characters through basic filtering rather than a true search engine. Characters can be grouped by debut year, birthday, or Japanese alphabetical order, but the system does not support natural-language queries or multi-attribute searches. Users mainly explore characters through categories and individual profile pages.

   Website Link: https://www.sanrio.co.jp/characters/?search-category=debut



2. **Fandom wiki pages (searchable but not standardized or ranked)** Fandom provides a general search function that looks through all Sanrio-related pages. It works through keyword matching in page titles and text. However, because the content is fan-edited, the information is not standardized, and search results may vary in quality. The system does not rank characters by relevance or support detailed attribute-based queries.
   Link: https://hellokitty.fandom.com/wiki/Category:Sanriot



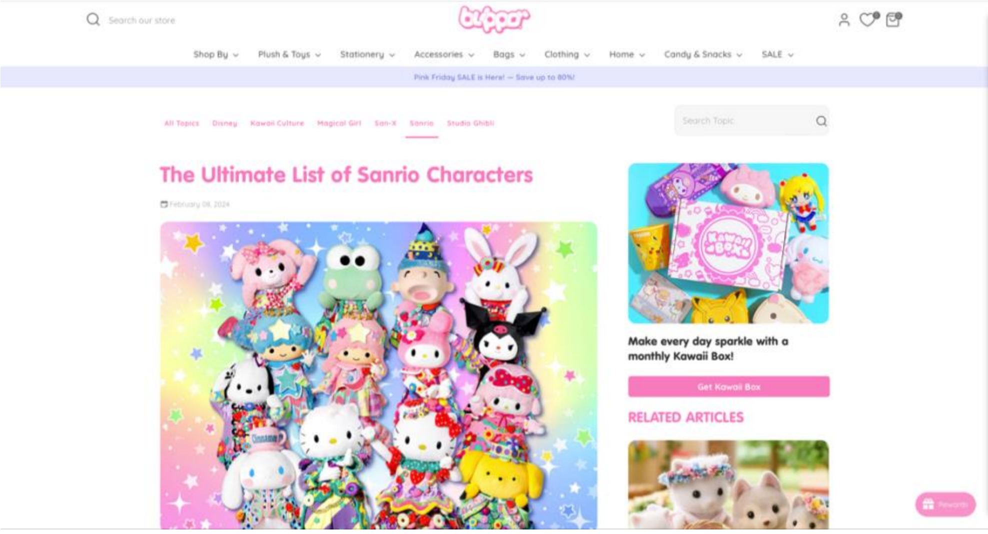
Sanriot

Category page

There is currently no text in this page. You can search for this page title in other pages, or search the related logs, but you do not have permission to create this page.

3. **Blippo Sanrio character pages (colorful but limited information and no advanced search)**
   Blippo offers a browsing page that lists Sanrio characters with images and short descriptions. This page does not include an actual search engine. Users cannot filter by traits such as species, colors, or debut year. It is mainly a visual catalog with limited information and without ranking or advanced search features.

   Website Link: [https://www.blippo.com/blogs/characters/sanrio-characters](https://www.blippo.com/blogs/characters/sanrio-characters)



Pink Friday SALE is Here! — Save up to 80%!

The Ultimate List of Sanrio Characters

February 08, 2024

Make every day sparkle with a monthly Kawaii Box!

RELATED ARTICLES

## Implementation

### Data collection, example documents, and data statistics

The dataset was collected from the Sanrio Japan website, Blippo’s Sanrio , and the Hello Kitty Fandom website. It contains 250-character entries in JSON format, each including details such as name, picture, description, first appearance, creator name, species, and gender. The dataset supports single-word, multi-word, and partial-match queries, providing a structured foundation for a Sanrio character search engine.

### Link of website

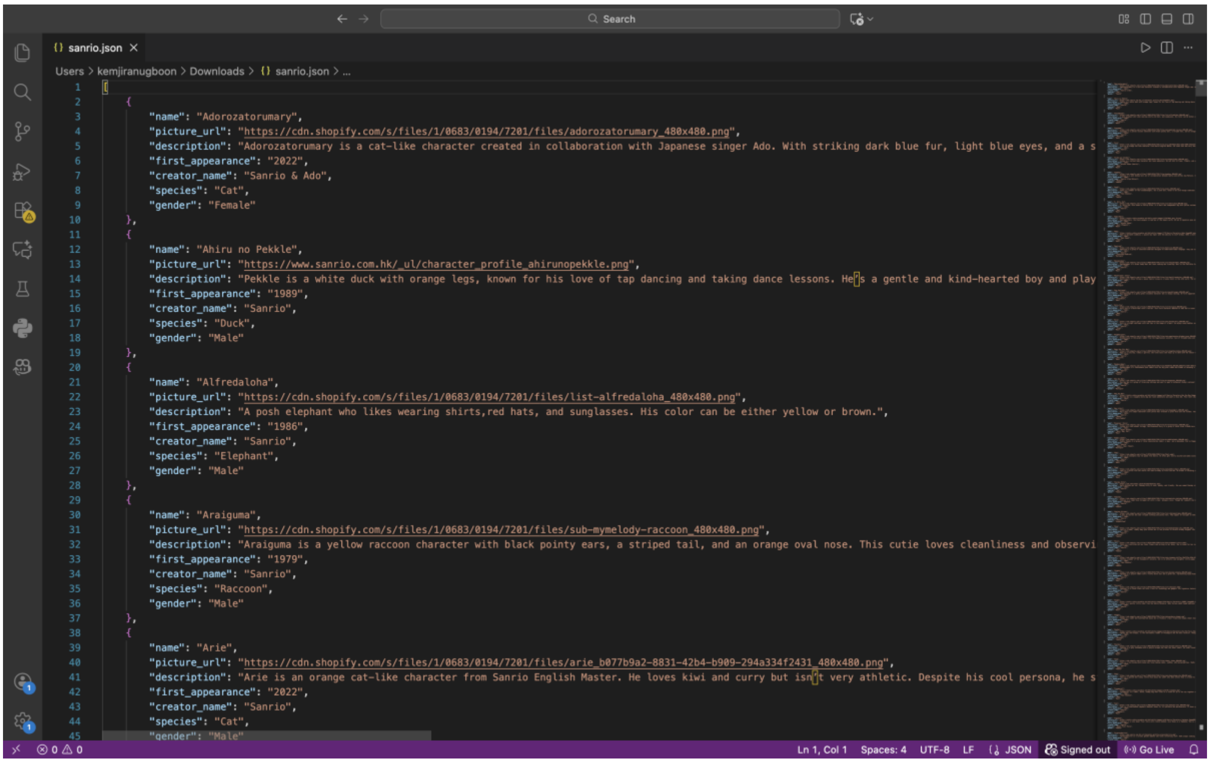
- https://www.sanrio.co.jp/characters/?search-category=debut

- https://www.blippo.com/blogs/characters/sanrio-characters?srsltid=AfmBOor4vN8EOxv-KEuQYbo2uVY61V778h3ivXm4LOQO5VFenRpZA4xP

-https://hellokitty.fandom.com/wiki/Category:Sanriot

### Example of documents

Example of a character document stored in **JSON format**, including attributes such as name, picture_url, description, first_appearance, creator_name,species and gender.



```
[
    {
        "name": "Adorozatorumary",
        "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/adorozatorumary_480x480.png",
        "description": "Adorozatorumary is a cat-like character created in collaboration with Japanese singer Ado. With striking dark blue fur, light blue eyes, and a s
        "first_appearance": "2022",
        "creator_name": "Sanrio & Ado",
        "species": "Cat",
        "gender": "Female"
    },
    {
        "name": "Ahiru no Pekkle",
        "picture_url": "https://www.sanrio.com.hk/_ul/character_profile_ahirunopekkle.png",
        "description": "Pekkle is a white duck with orange legs, known for his love of tap dancing and taking dance lessons. He’s a gentle and kind-hearted boy and play
        "first_appearance": "1989",
        "creator_name": "Sanrio",
        "species": "Duck",
        "gender": "Male"
    },
    {
        "name": "Alfredaloha",
        "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/list-alfredaloha_480x480.png",
        "description": "A posh elephant who likes wearing shirts,red hats, and sunglasses. His color can be either yellow or brown.",
        "first_appearance": "1986",
        "creator_name": "Sanrio",
        "species": "Elephant",
        "gender": "Male"
    },
    {
        "name": "Araiguma",
        "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/sub-mymelody-raccoon_480x480.png",
        "description": "Araiguma is a yellow raccoon character with black pointy ears, a striped tail, and an orange oval nose. This cutie loves cleanliness and observi
        "first_appearance": "1979",
        "creator_name": "Sanrio",
        "species": "Raccoon",
        "gender": "Male"
    },
    {
        "name": "Arie",
        "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/arie_b077b9a2-8831-42b4-b909-294a334f2431_480x480.png",
        "description": "Arie is an orange cat-like character from Sanrio English Master. He loves kiwi and curry but isn’t very athletic. Despite his cool persona, he s
        "first_appearance": "2022",
        "creator_name": "Sanrio",
        "species": "Cat",
        "gender": "Male"
```

## Tools and software

### 1. Python

Python is used extensively for data collection, cleaning, and processing. It handles tasks such as parsing character information from multiple websites and organizing the already-existing JSON files into a clean and consistent structure. Python also prepares the data for indexing in the search engine and powers the backend logic of the system, including query handling and integration with Elasticsearch.

### 2. Flask

Flask serves as the web framework for the front-end interface. It enables users to interact with the search engine through a web browser, providing search forms, displaying results, and supporting various types of queries such as single-word, multi-word, and partial matches. Flask ensures that the system is accessible and user-friendly.

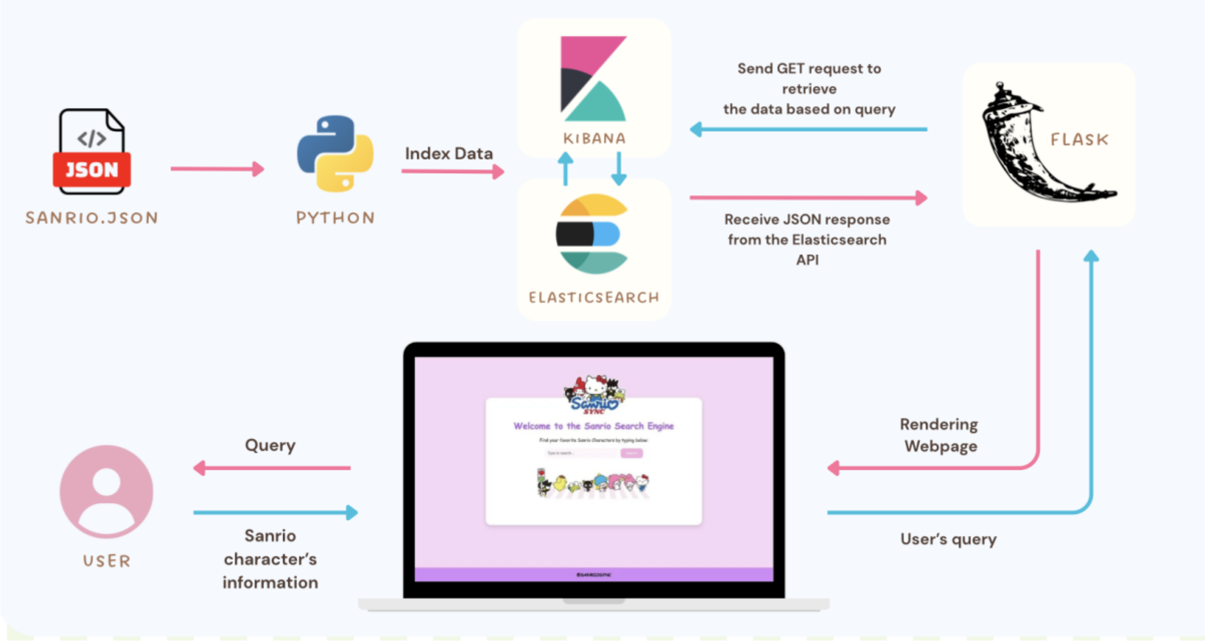
### 3. Elasticsearch

Elasticsearch is used for indexing the character dataset and providing fast, efficient search capabilities. It supports advanced search features, including full-text search, relevance ranking, and partial matches, making it ideal for retrieving relevant Sanrio character entries quickly.

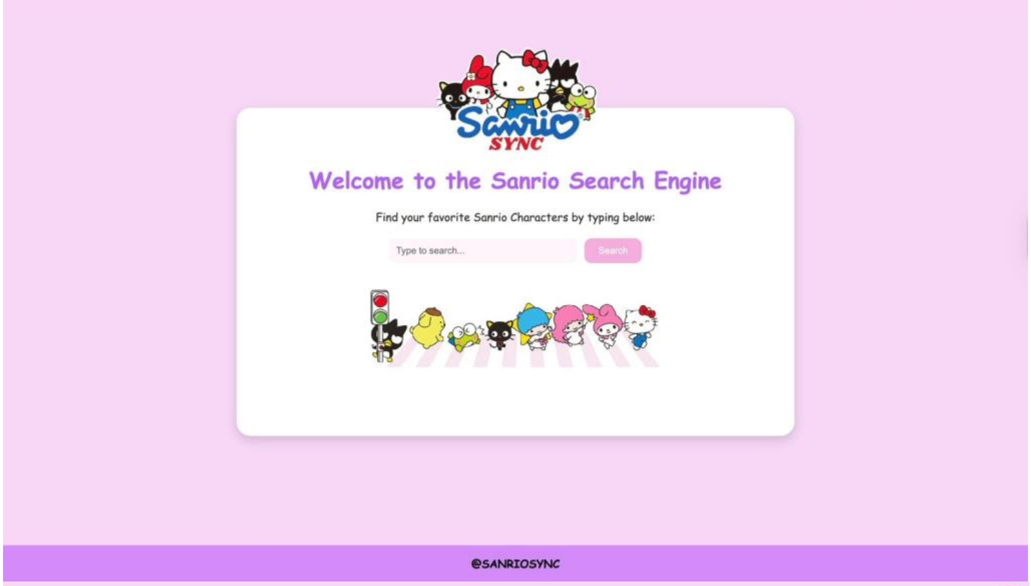
### 4. Kibana

Kibana is employed to visualize the indexed data and monitor the search engine's performance. It allows the team to analyze query patterns, detect potential issues, and gain insights into user interactions with the system. Kibana dashboards help in understanding how the data is being searched and accessed, which is valuable for further optimization.

## System diagram

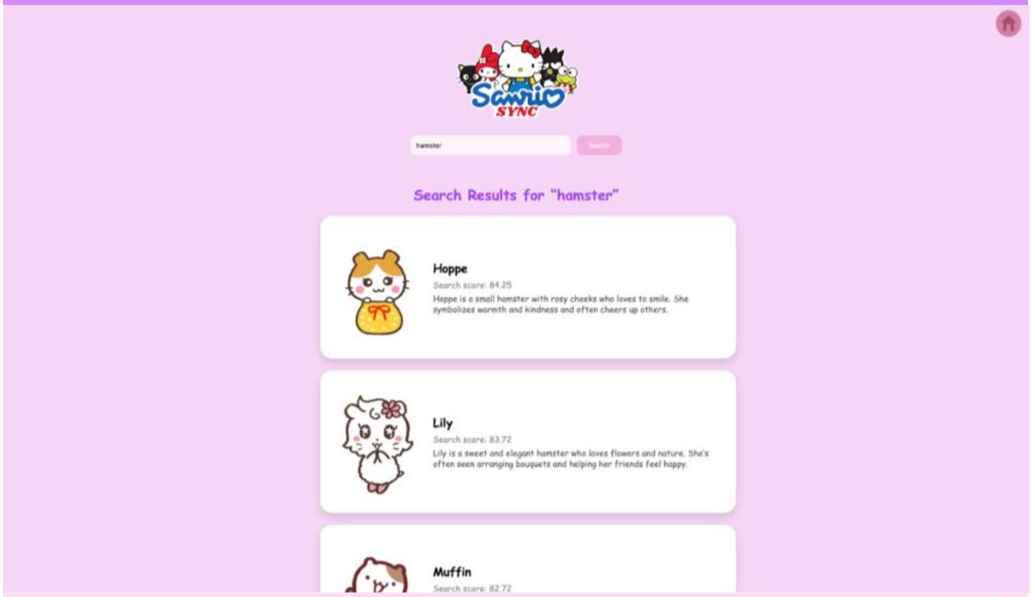

JSON
SANRIO.JSON
PYTHON
Index Data
KIBANA
ELASTICSEARCH
Send GET request to retrieve the data based on query
Receive JSON response from the Elasticsearch API
FLASK
Rendering Webpage
User's query
Query
Sanrio character's information
USER
Welcome to the Sanrio Search Engine

## Screenshots of the system



Welcome to the Sanrio Search Engine

Find your favorite Sanrio Characters by typing below:



Search Results for "hamster"

Hoppe

Search score: 84.25

Hoppe is a small hamster with rosy cheeks who loves to smile. She symbolizes warmth and kindness and often cheers up others.

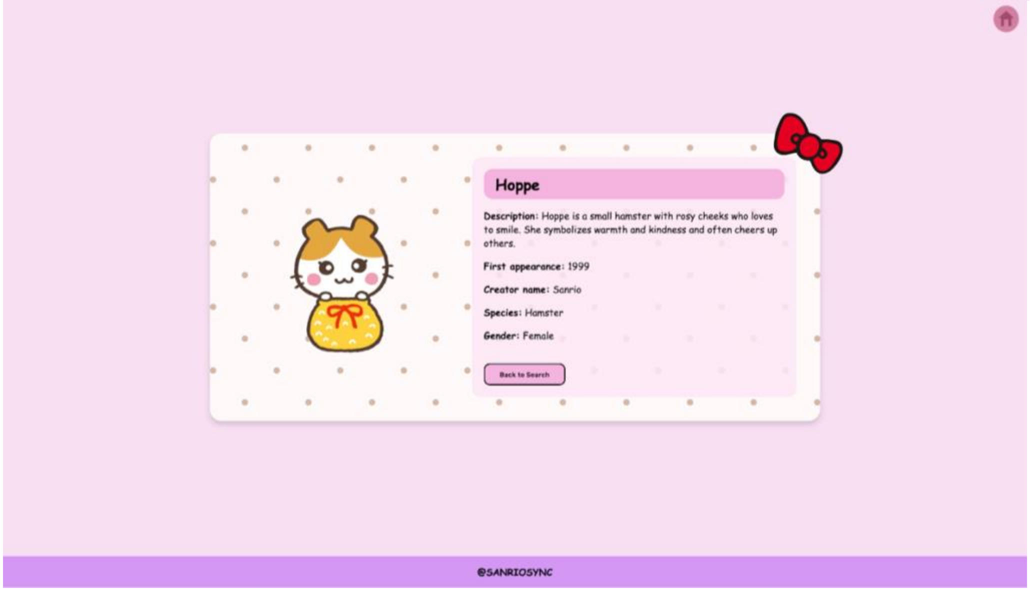
Lily

Search score: 83.72

Lily is a sweet and elegant hamster who loves flowers and nature. She's often seen arranging bouquets and helping her friends feel happy.

Muffin

Search score: 82.72



Hoppe

**Description**: Hoppe is a small hamster with rosy cheeks who loves to smile. She symbolizes warmth and kindness and often cheers up others.

**First appearance**: 1999

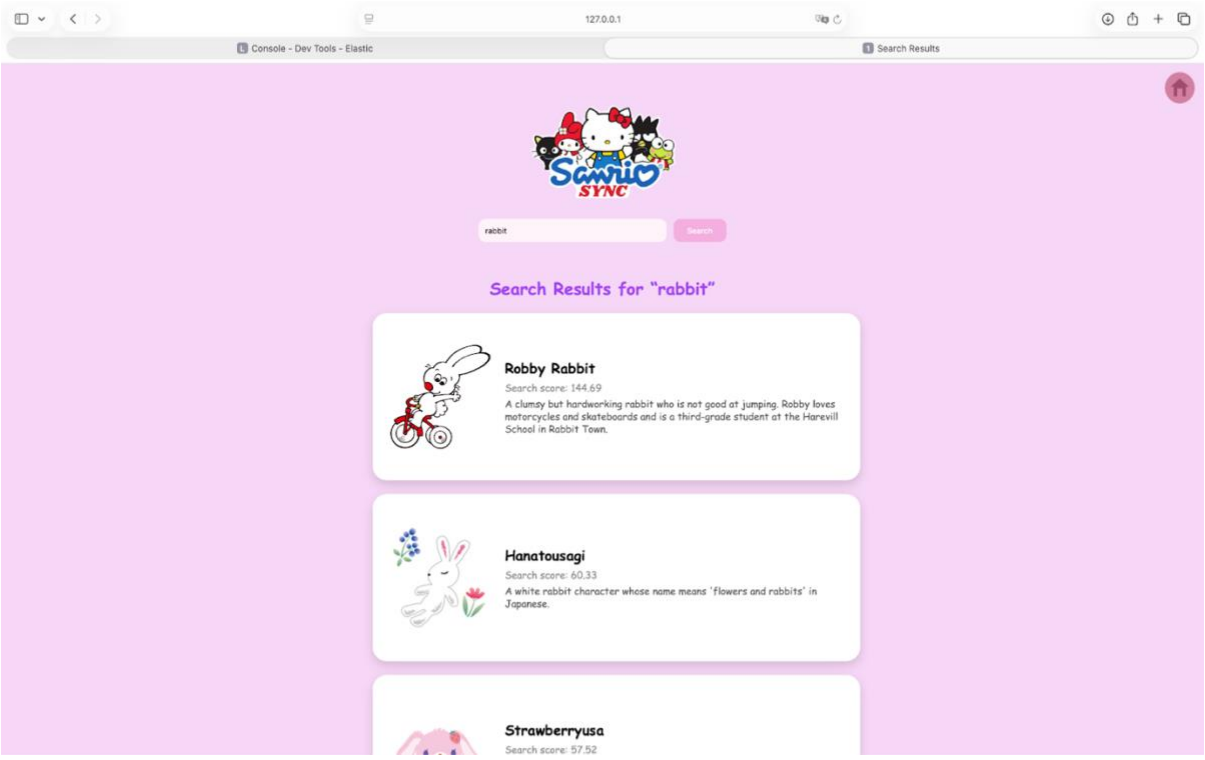
**Creator name**: Sanrio

**Species**: Hamster

**Gender**: Female

### Example step-by-step search sessions

#### 1. One word query



Search Results for "rabbit"

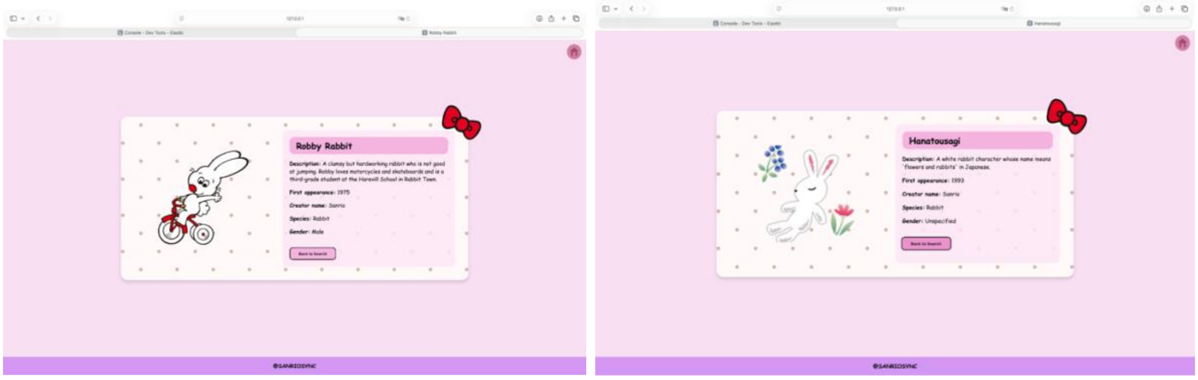
**Robby Rabbit**
Search score: 144.69
A clumsy but hardworking rabbit who is not good at jumping. Robby loves motorcycles and skateboards and is a third-grade student at the Harevill School in Rabbit Town.

**Hanatousagi**
Search score: 60.33
A white rabbit character whose name means 'flowers and rabbits' in Japanese.

**Strawberryusa**
Search score: 57.52

Search results for the one-word query "rabbit", showing relevant characters associated with rabbit-related traits. The system returns matches such as Robby Rabbit and other characters based on name, species, and attribute similarity.

#### - Character Detail Page



**Robby Rabbit**

**Description:** A clumsy but hardworking rabbit who is not good at jumping. Robby loves motorcycles and skateboards and is a third-grade student at the Harevill School in Rabbit Town.

**First appearance:** 1975

**Creator name:** Sanrio

**Species:** Rabbit

**Gender:** Male

**Hanatousagi**

**Description:** A white rabbit character whose name means 'flowers and rabbits' in Japanese.

**First appearance:** 1993

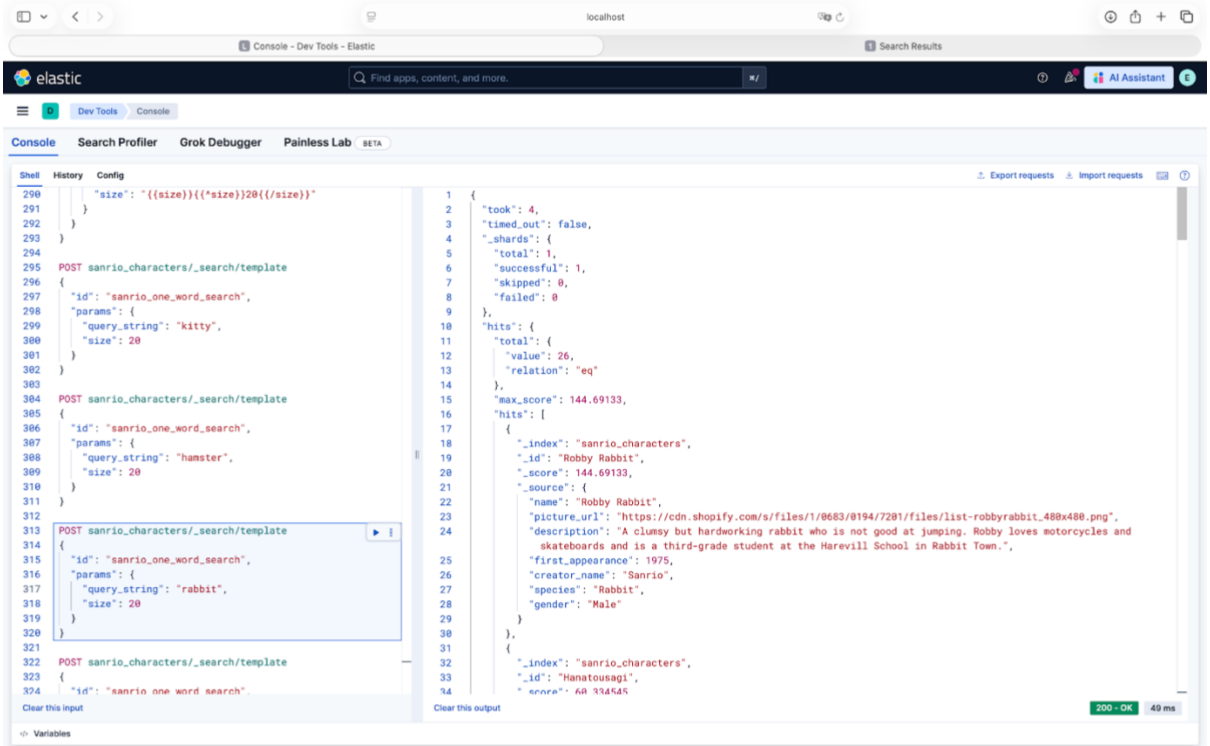
**Creator name:** Sanrio

**Species:** Rabbit

**Gender:** Unspecified

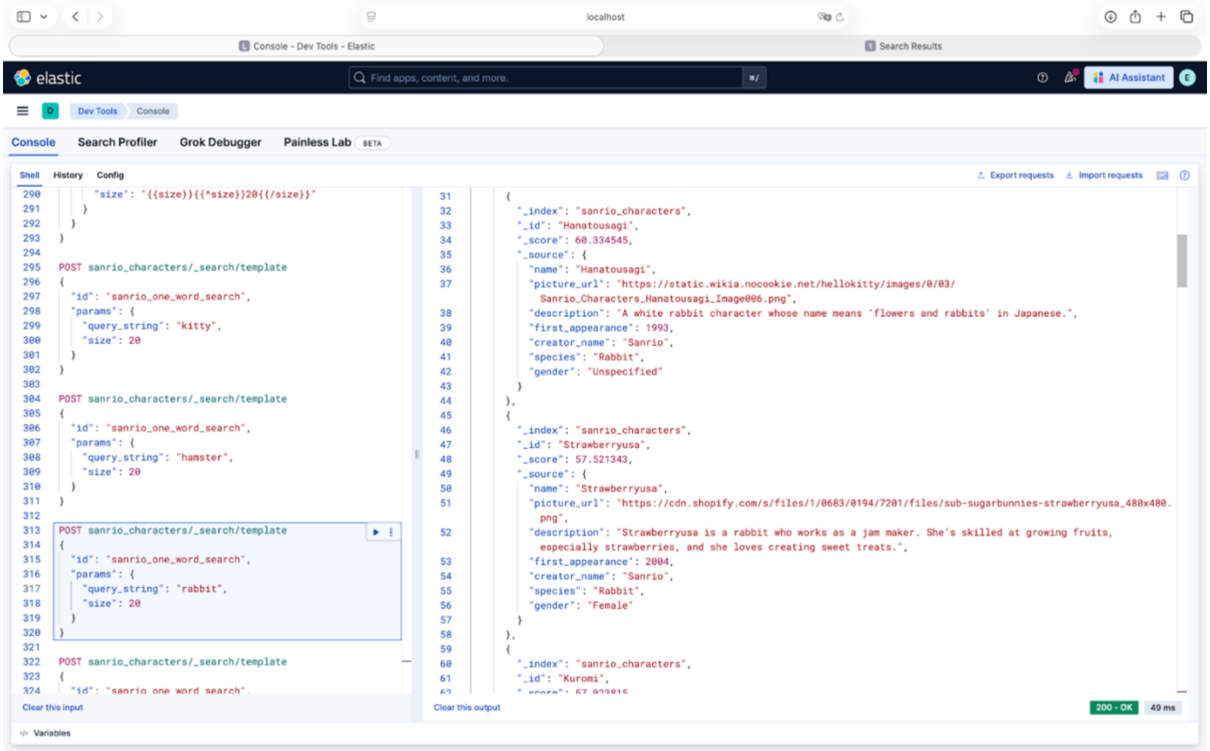
This figure illustrates the character detail page. When a user selects a specific character, the system presents the full character information such as description, debut year, creator name, species, and gender along with a larger character image. This page provides users with a more complete and organized view of each Sanrio character.

- **Elasticsearch Query and Results**



```
        "size": "{{size}}{{^size}}20{{/size}}"
      }
    }
  }

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "kitty",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "hamster",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
```

```
{
  "took": 4,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 26,
      "relation": "eq"
    },
    "max_score": 144.69133,
    "hits": [
      {
        "_index": "sanrio_characters",
        "_id": "Robby Rabbit",
        "_score": 144.69133,
        "_source": {
          "name": "Robby Rabbit",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/list-robbyrabbit_480x480.png",
          "description": "A clumsy but hardworking rabbit who is not good at jumping. Robby loves motorcycles and
            skateboards and is a third-grade student at the Harevill School in Rabbit Town.",
          "first_appearance": 1975,
          "creator_name": "Sanrio",
          "species": "Rabbit",
          "gender": "Male"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Hanatousagi",
        "_score": 60.334545
```

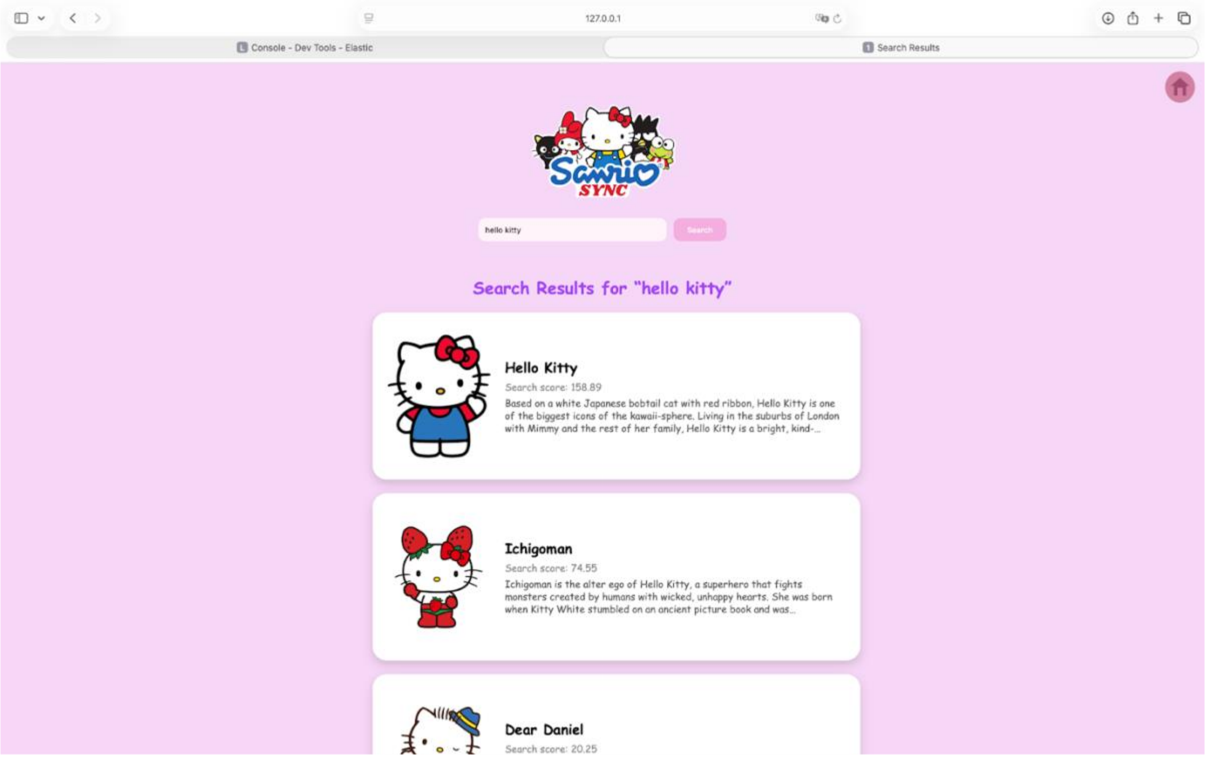


```
        "size": "{{size}}{{^size}}20{{/size}}"
      }
    }
  }

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "kitty",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "hamster",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
```

```
      {
        "_index": "sanrio_characters",
        "_id": "Hanatousagi",
        "_score": 60.334545,
        "_source": {
          "name": "Hanatousagi",
          "picture_url": "https://static.wikia.nocookie.net/hellokitty/images/0/03/
            Sanrio_Characters_Hanatousagi_Image006.png",
          "description": "A white rabbit character whose name means 'flowers and rabbits' in Japanese.",
          "first_appearance": 1993,
          "creator_name": "Sanrio",
          "species": "Rabbit",
          "gender": "Unspecified"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Strawberryusa",
        "_score": 57.521343,
        "_source": {
          "name": "Strawberryusa",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/sub-sugarbunnies-strawberryusa_480x480.
            png",
          "description": "Strawberryusa is a rabbit who works as a jam maker. She's skilled at growing fruits,
            especially strawberries, and she loves creating sweet treats.",
          "first_appearance": 2004,
          "creator_name": "Sanrio",
          "species": "Rabbit",
          "gender": "Female"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Kuromi",
        "_score": 57.023815
```

This figure displays the Elasticsearch Dev Tools interface used during development. It shows the search template query executed for the keyword “rabbit” and the corresponding ranked results returned by the Elasticsearch engine. This demonstrates how the backend processes and retrieves character data according to relevance scoring.

## 2. Multiple word query



Search Results for "hello kitty"

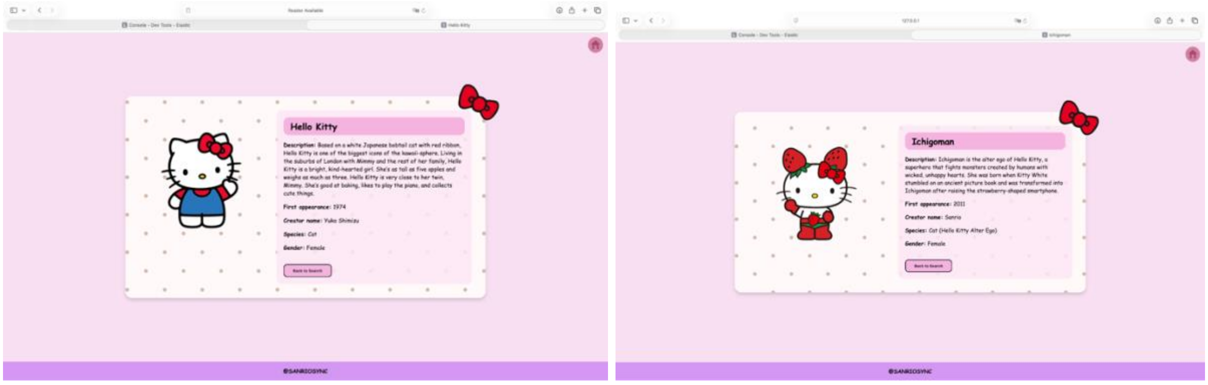
Hello Kitty
Search score: 158.89
Based on a white Japanese bobtail cat with red ribbon, Hello Kitty is one of the biggest icons of the kawaii-sphere. Living in the suburbs of London with Mimmy and the rest of her family, Hello Kitty is a bright, kind-...

Ichigoman
Search score: 74.55
Ichigoman is the alter ego of Hello Kitty, a superhero that fights monsters created by humans with wicked, unhappy hearts. She was born when Kitty White stumbled on an ancient picture book and was...

Dear Daniel
Search score: 20.25

Example of a multi-word query "hello kitty", where the system correctly identifies Hello Kitty as the top-ranked result through direct lexical matching and ranking boosts.

## - Character Detail Page



Hello Kitty

**Description:** Based on a white Japanese bobtail cat with red ribbon, Hello Kitty is one of the biggest icons of the kawaii-sphere. Living in the suburbs of London with Mimmy and the rest of her family, Hello Kitty is a bright, kind-hearted girl. She's as tall as five apples and weighs as much as three. Hello Kitty is very close to her twin, Mimmy. She's good at baking, likes to play the piano, and collects cute things.

**First appearance:** 1974

**Creator name:** Yuko Shimizu

**Species:** Cat

**Gender:** Female

Ichigoman

**Description:** Ichigoman is the alter ego of Hello Kitty, a superhero that fights monsters created by humans with wicked, unhappy hearts. She was born when Kitty White stumbled on an ancient picture book and was transformed into Ichigoman after raising the strawberry-shaped smartphone.

**First appearance:** 2011

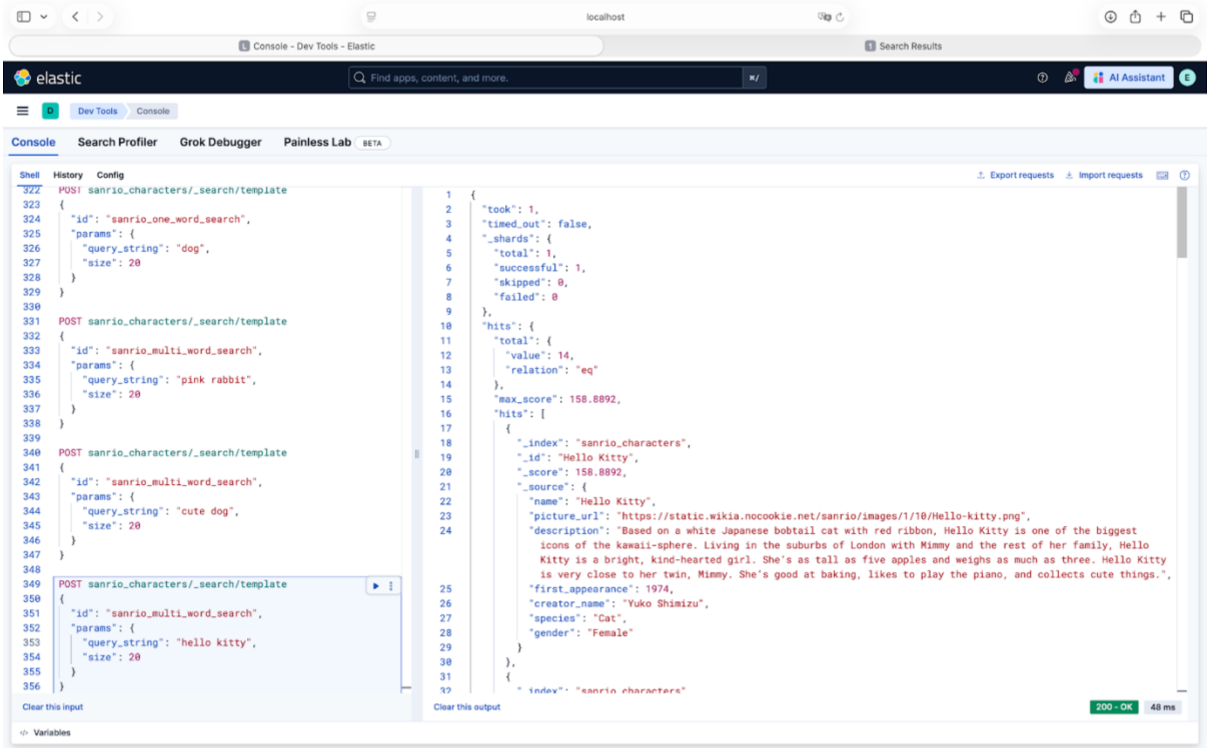
**Creator name:** Sanrio

**Species:** Cat (Hello Kitty Alter Ego)

**Gender:** Female

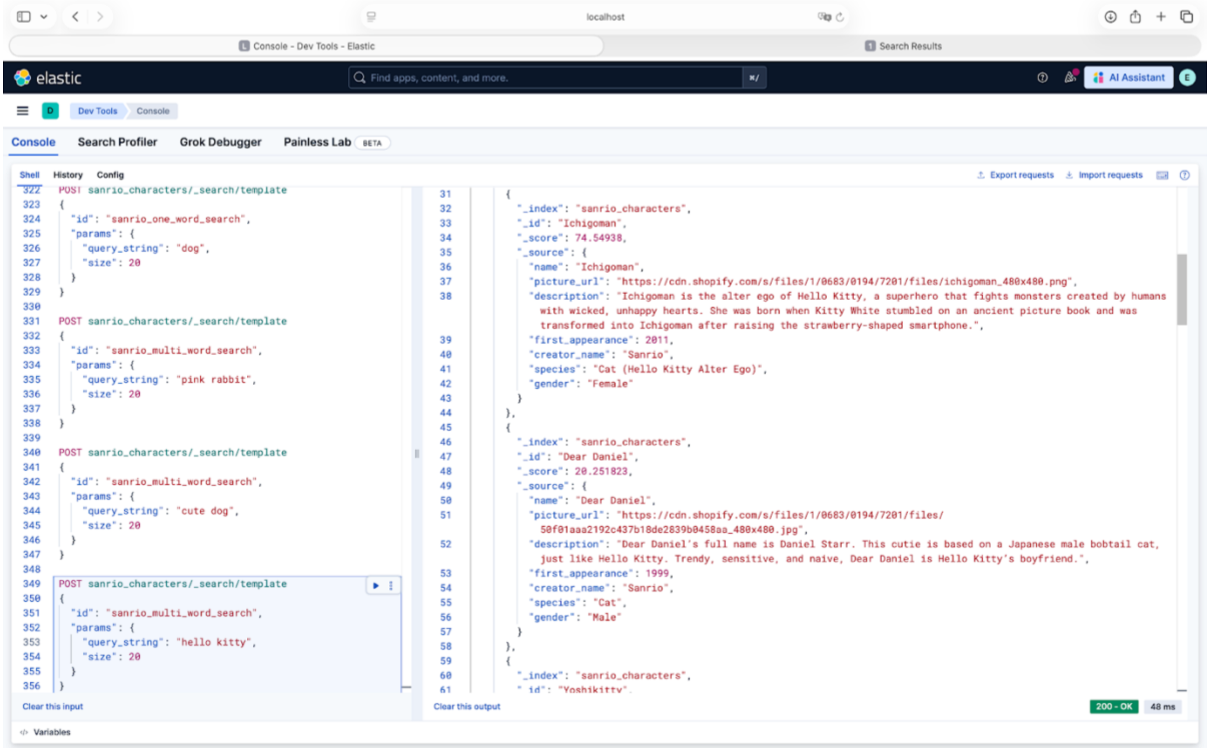
This figure illustrates the character detail page. When a user selects a specific character, the system presents the full character information such as description, debut year, creator name, species, and gender along with a larger character image. This page provides users with a more complete and organized view of each Sanrio character.

- **Elasticsearch Query and Results**



```
POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "cute dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "hello kitty",
    "size": 20
  }
}
```

```
{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 14,
      "relation": "eq"
    },
    "max_score": 158.8892,
    "hits": [
      {
        "_index": "sanrio_characters",
        "_id": "Hello Kitty",
        "_score": 158.8892,
        "_source": {
          "name": "Hello Kitty",
          "picture_url": "https://static.wikia.nocookie.net/sanrio/images/1/10/Hello-kitty.png",
          "description": "Based on a white Japanese bobtail cat with red ribbon, Hello Kitty is one of the biggest
            icons of the kawaii-sphere. Living in the suburbs of London with Mimmy and the rest of her family, Hello
            Kitty is a bright, kind-hearted girl. She's as tall as five apples and weighs as much as three. Hello Kitty
            is very close to her twin, Mimmy. She's good at baking, likes to play the piano, and collects cute things.",
          "first_appearance": 1974,
          "creator_name": "Yuko Shimizu",
          "species": "Cat",
          "gender": "Female"
        }
      },
      {
        "_index": "sanrio_characters"
```

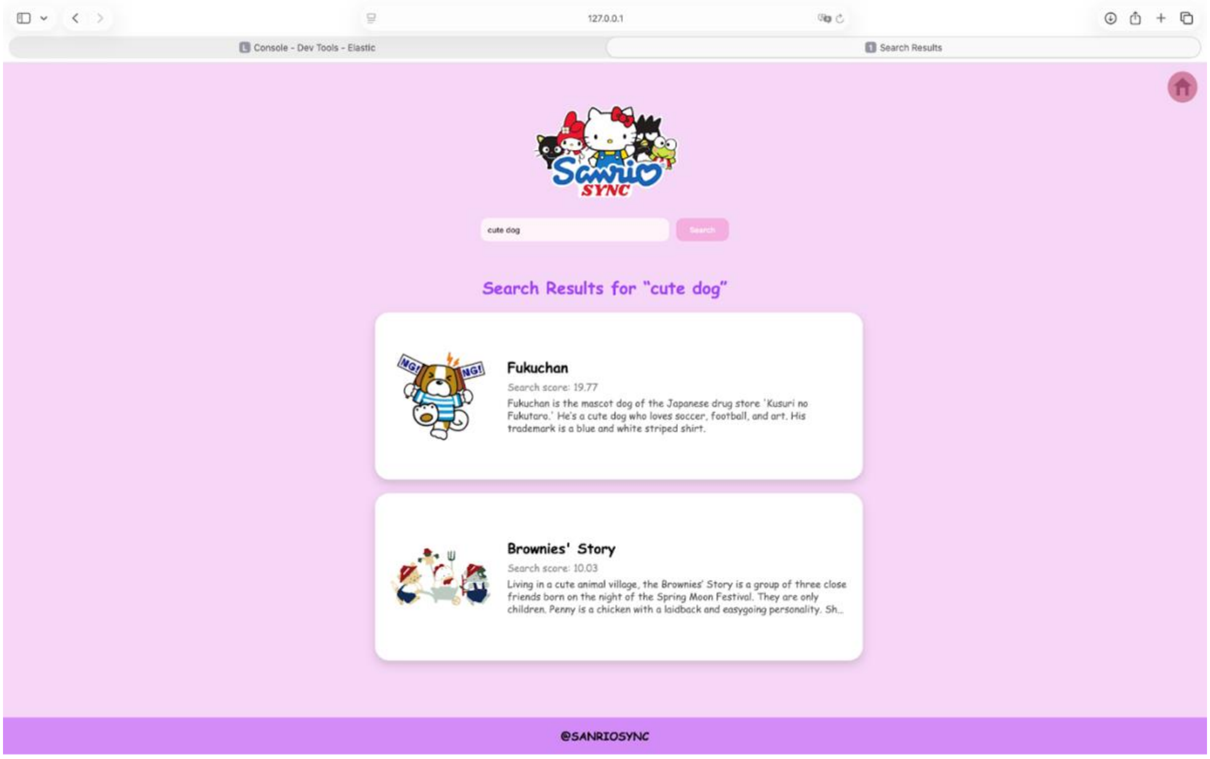


```
POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "cute dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "hello kitty",
    "size": 20
  }
}
```

```
      {
        "_index": "sanrio_characters",
        "_id": "Ichigoman",
        "_score": 74.54938,
        "_source": {
          "name": "Ichigoman",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/ichigoman_480x480.png",
          "description": "Ichigoman is the alter ego of Hello Kitty, a superhero that fights monsters created by humans
            with wicked, unhappy hearts. She was born when Kitty White stumbled on an ancient picture book and was
            transformed into Ichigoman after raising the strawberry-shaped smartphone.",
          "first_appearance": 2011,
          "creator_name": "Sanrio",
          "species": "Cat (Hello Kitty Alter Ego)",
          "gender": "Female"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Dear Daniel",
        "_score": 20.251823,
        "_source": {
          "name": "Dear Daniel",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/
            50f01aaa2192c437b18de2839b0458aa_480x480.jpg",
          "description": "Dear Daniel's full name is Daniel Starr. This cutie is based on a Japanese male bobtail cat,
            just like Hello Kitty. Trendy, sensitive, and naive, Dear Daniel is Hello Kitty's boyfriend.",
          "first_appearance": 1999,
          "creator_name": "Sanrio",
          "species": "Cat",
          "gender": "Male"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Yoshikitty",
```

This figure displays the Elasticsearch Dev Tools interface used during development. It shows the search template query executed for the keyword “hello kitty” and the corresponding ranked results returned by the Elasticsearch engine. This demonstrates how the backend processes and retrieves character data according to relevance scoring.

### 3. Partial match



cute dog

Search Results for "cute dog"

**Fukuchan**

Search score: 19.77

Fukuchan is the mascot dog of the Japanese drug store 'Kusuri no Fukutaro.' He's a cute dog who loves soccer, football, and art. His trademark is a blue and white striped shirt.

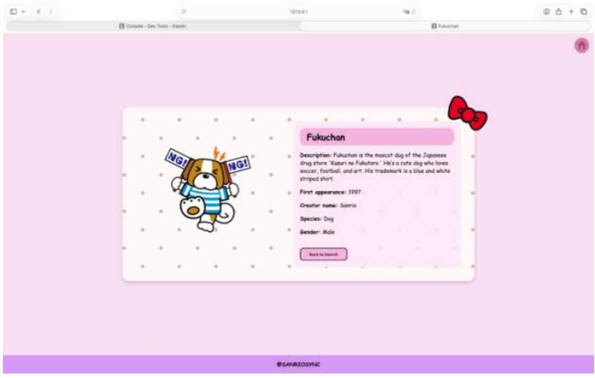
**Brownies' Story**

Search score: 10.03

Living in a cute animal village, the Brownies' Story is a group of three close friends born on the night of the Spring Moon Festival. They are only children. Penny is a chicken with a laidback and easygoing personality. Sh...

Partial and combined query example using the term "cute dog", showing the system's ability to infer the intended meaning and retrieve dog such as Fukuchan through fuzzy and semantic matching.

### - Character Detail Page



**Fukuchan**

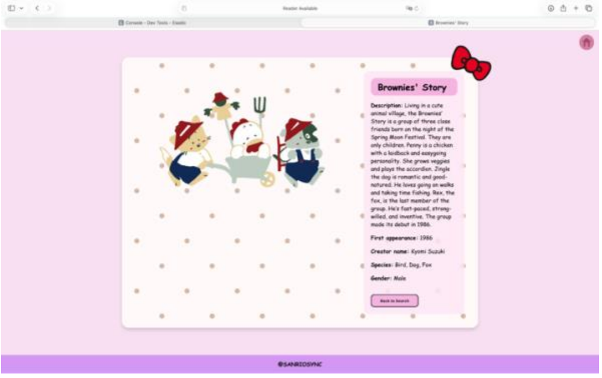
**Description:** Fukuchan is the mascot dog of the Japanese drug store 'Kusuri no Fukutaro.' He's a cute dog who loves soccer, football, and art. His trademark is a blue and white striped shirt.

**First appearance:** 1997

**Creator name:** Sanrio

**Species:** Dog

**Gender:** Male



**Brownies' Story**

**Description:** Living in a cute animal village, the Brownies' Story is a group of three close friends born on the night of the Spring Moon Festival. They are only children. Penny is a chicken with a laidback and easygoing personality. She grows veggies and plays the accordion. Jingle the dog is romantic and good-natured. He loves going on walks and taking time fishing. Rex, the fox, is the last member of the group. He's fast-paced, strong-willed, and inventive. The group made its debut in 1986.

**First appearance:** 1986

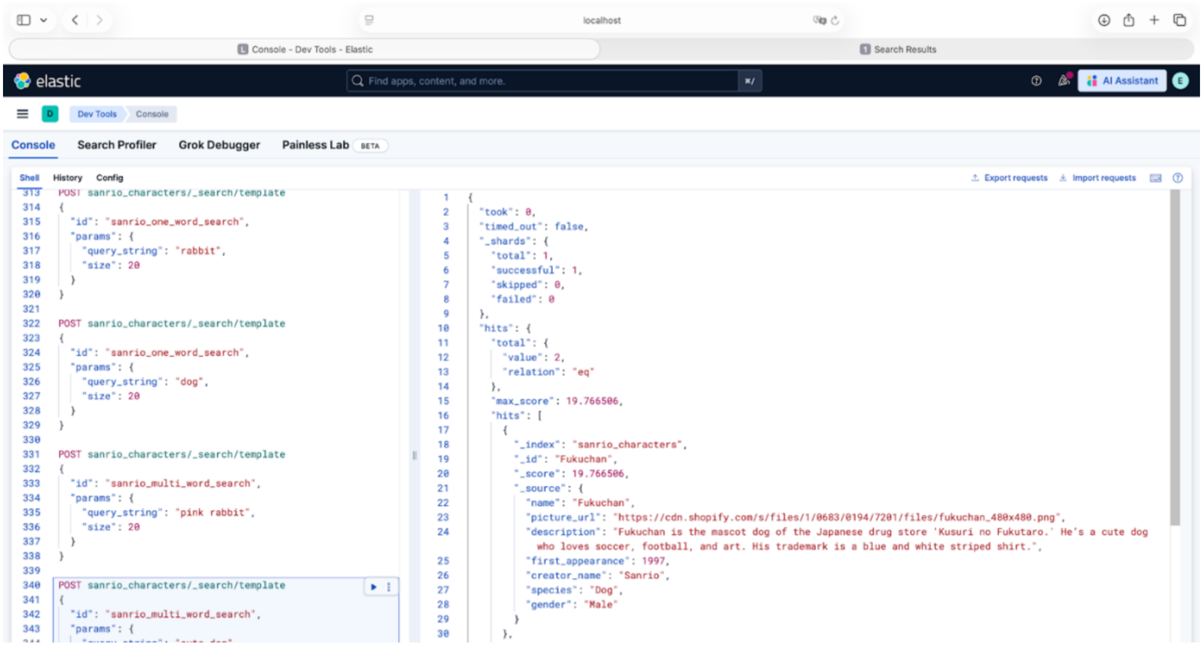
**Creator name:** Kyomi Suzuki

**Species:** Bird, Dog, Fox

**Gender:** Male

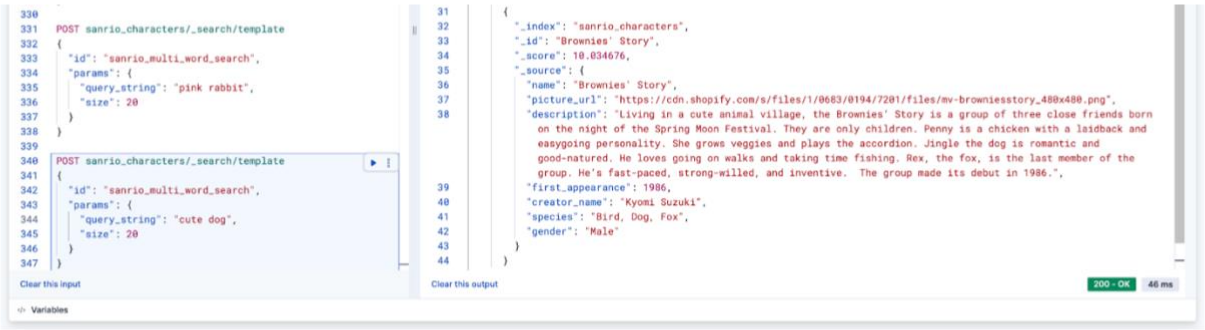
This figure illustrates the character detail page. When a user selects a specific character, the system presents the full character information such as description, debut year, creator name, species, and gender along with a larger character image. This page provides users with a more complete and organized view of each Sanrio character.

## - Elasticsearch Query and Results



```
POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
```

```
{
  "took": 0,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 2,
      "relation": "eq"
    },
    "max_score": 19.766506,
    "hits": [
      {
        "_index": "sanrio_characters",
        "_id": "Fukuchan",
        "_score": 19.766506,
        "_source": {
          "name": "Fukuchan",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/fukuchan_480x480.png",
          "description": "Fukuchan is the mascot dog of the Japanese drug store 'Kusuri no Fukutaro.' He's a cute dog
            who loves soccer, football, and art. His trademark is a blue and white striped shirt.",
          "first_appearance": 1997,
          "creator_name": "Sanrio",
          "species": "Dog",
          "gender": "Male"
        }
      },
```



```

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "cute dog",
    "size": 20
  }
}
```

```
      {
        "_index": "sanrio_characters",
        "_id": "Brownies' Story",
        "_score": 10.034676,
        "_source": {
          "name": "Brownies' Story",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/mv-browniesstory_480x480.png",
          "description": "Living in a cute animal village, the Brownies' Story is a group of three close friends born
            on the night of the Spring Moon Festival. They are only children. Penny is a chicken with a laidback and
            easygoing personality. She grows veggies and plays the accordion. Jingle the dog is romantic and
            good-natured. He loves going on walks and taking time fishing. Rex, the fox, is the last member of the
            group. He's fast-paced, strong-willed, and inventive.  The group made its debut in 1986.",
          "first_appearance": 1986,
          "creator_name": "Kyomi Suzuki",
          "species": "Bird, Dog, Fox",
          "gender": "Male"
        }
      }
```

This figure displays the Elasticsearch Dev Tools interface used during development. It shows the search template query executed for the keyword “cute dog” and the corresponding ranked results returned by the Elasticsearch engine. This demonstrates how the backend processes and retrieves character data according to relevance scoring.

## 4. Ranking

The ranking system determines which Sanrio character should appear first on the search results page. This ranking is performed automatically by Elasticsearch, using its internal scoring model (BM25) with custom field boosting.

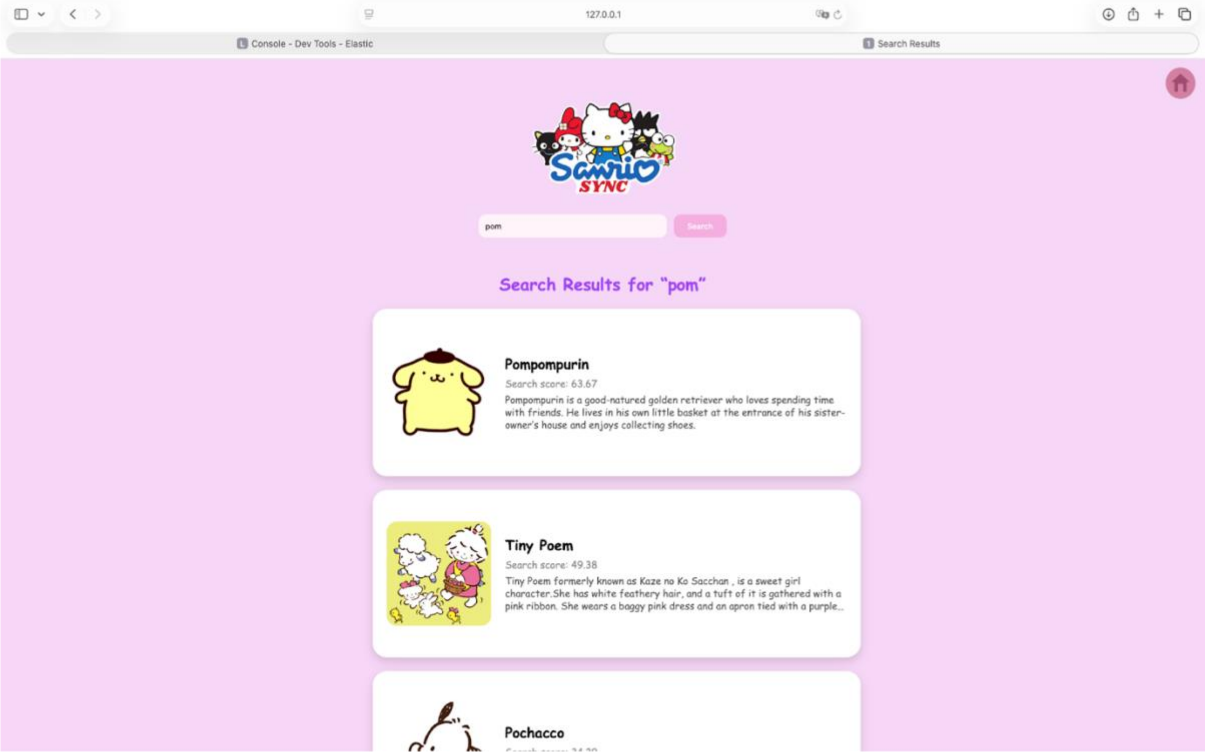
**There are six main factors:**

- Field Boost: Different fields have different weights. Name field gets 10x boost, species. Keyword gets 10x, species gets 8x, description. Raw gets 3x, and description with synonyms gets 2x.
- Term Frequency: Keywords appearing multiple times in the same document increase the score.
- Inverse Document Frequency: Rare terms are more valuable. e.g., "Golden retriever" appearing in few documents scores higher than "dog" appearing in many documents.
- Field Length Normalization: Shorter documents with keywords score higher than longer documents.
- Phrase Proximity: For multi-word queries, adjacent keywords score significantly higher.
- Match Type: Exact matches score highest, followed by fuzzy matches with typos, then synonym matches.

The final score calculation combines these factors: Field Boost multiplied by Term Frequency and Inverse Document Frequency, divided by Field Length, plus bonuses for phrase proximity and match type. Documents (characters) that match the user's search query more strongly based on their name, description receive a higher relevance score and are ranked at the top.

### Example

Search query: "pom", demonstrating the system's ability to retrieve Pom Pom Purin and other related characters through prefix matching and character-name similarity.



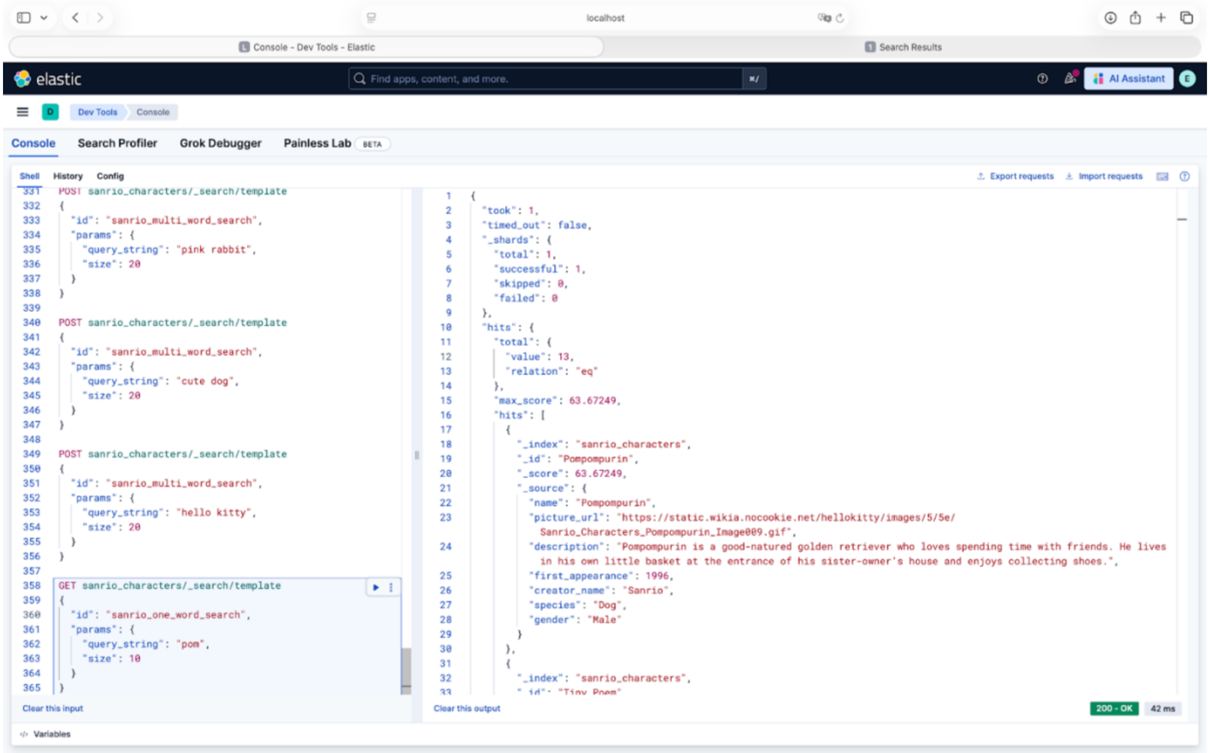
Search Results for "pom"

Pompompurin
Search score: 63.67
Pompompurin is a good-natured golden retriever who loves spending time with friends. He lives in his own little basket at the entrance of his sister-owner's house and enjoys collecting shoes.

Tiny Poem
Search score: 49.38
Tiny Poem formerly known as Kaze no Ko Sacchan , is a sweet girl character.She has white feathery hair, and a tuft of it is gathered with a pink ribbon. She wears a baggy pink dress and an apron tied with a purple...

Pochacco

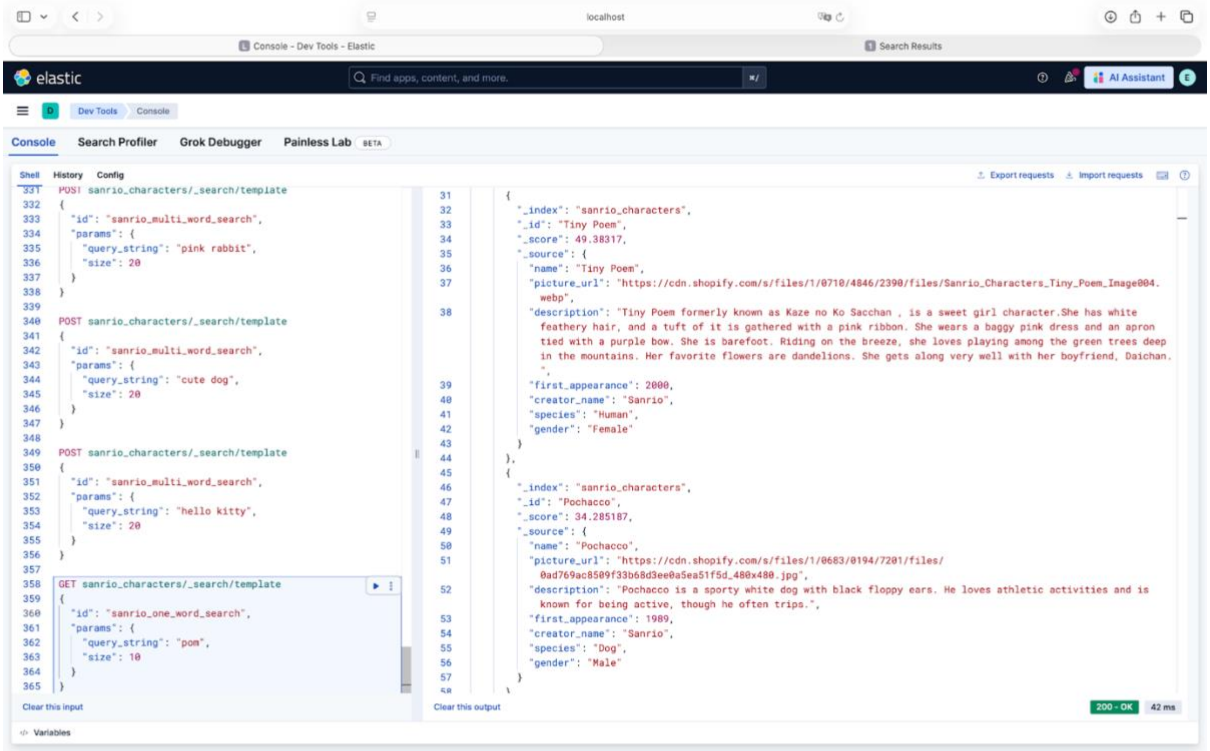
## - Elasticsearch Query and Results



```
POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "cute dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "hello kitty",
    "size": 20
  }
}

GET sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "pom",
    "size": 10
  }
}
```

```
{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 13,
      "relation": "eq"
    },
    "max_score": 63.67249,
    "hits": [
      {
        "_index": "sanrio_characters",
        "_id": "Pompompurin",
        "_score": 63.67249,
        "_source": {
          "name": "Pompompurin",
          "picture_url": "https://static.wikia.nocookie.net/hellokitty/images/5/5e/
            Sanrio_Characters_Pompompurin_Image009.gif",
          "description": "Pompompurin is a good-natured golden retriever who loves spending time with friends. He lives
            in his own little basket at the entrance of his sister-owner's house and enjoys collecting shoes.",
          "first_appearance": 1996,
          "creator_name": "Sanrio",
          "species": "Dog",
          "gender": "Male"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Tiny Poem"
```

Elasticsearch Dev Tools output for the query “pom”, showing how the system retrieves Pom Pom Purin as the top-ranked result using prefix matching and relevance scoring.



```
POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "pink rabbit",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "cute dog",
    "size": 20
  }
}

POST sanrio_characters/_search/template
{
  "id": "sanrio_multi_word_search",
  "params": {
    "query_string": "hello kitty",
    "size": 20
  }
}

GET sanrio_characters/_search/template
{
  "id": "sanrio_one_word_search",
  "params": {
    "query_string": "pom",
    "size": 10
  }
}
```

```
      {
        "_index": "sanrio_characters",
        "_id": "Tiny Poem",
        "_score": 49.38317,
        "_source": {
          "name": "Tiny Poem",
          "picture_url": "https://cdn.shopify.com/s/files/1/0710/4846/2390/files/Sanrio_Characters_Tiny_Poem_Image004.
            webp",
          "description": "Tiny Poem formerly known as Kaze no Ko Sacchan , is a sweet girl character.She has white
            feathery hair, and a tuft of it is gathered with a pink ribbon. She wears a baggy pink dress and an apron
            tied with a purple bow. She is barefoot. Riding on the breeze, she loves playing among the green trees deep
            in the mountains. Her favorite flowers are dandelions. She gets along very well with her boyfriend, Daichan.
            ",
          "first_appearance": 2000,
          "creator_name": "Sanrio",
          "species": "Human",
          "gender": "Female"
        }
      },
      {
        "_index": "sanrio_characters",
        "_id": "Pochacco",
        "_score": 34.285187,
        "_source": {
          "name": "Pochacco",
          "picture_url": "https://cdn.shopify.com/s/files/1/0683/0194/7201/files/
            0ad769ac8509f33b68d3ee0a5ea51f5d_480x480.jpg",
          "description": "Pochacco is a sporty white dog with black floppy ears. He loves athletic activities and is
            known for being active, though he often trips.",
          "first_appearance": 1989,
          "creator_name": "Sanrio",
          "species": "Dog",
          "gender": "Male"
        }
      }
```

Showing how the system retrieves characters whose names or attributes partially match the prefix, such as Tiny Poem and Pochacco, through relevance scoring and substring matching.

# Discussion

## Limitations of my system

1. Small Dataset and Limited Character Information

The current dataset contains only about 250 entries, which significantly limits the scalability of the system. Although Sanrio characters have extensive official information, only a portion of it could be collected, and some characters lack complete details such as descriptions or first appearance dates. This constraint reduces the richness of each character page.

In the future, adding more context such as character backstories, relationships, and references to related media (TV shows, movies, or games) would greatly improve user engagement and create a more informative experience. Additionally, certain characters contain large amounts of detailed information, but due to time constraints and difficulty in finding reliable sources, some of this data had to be removed or simplified.

2. Limited Handling of Rare Synonyms

Rare or uncommon synonyms are not always recognized by the search engine. Users who search using less typical vocabulary may receive fewer or less relevant results. Further tuning of synonym lists or integrating semantic models could address this limitation.

3. Image Handling and Validation

The search engine retrieves image URLs directly from Elasticsearch. If an image link is broken or the file is missing, it negatively affects the overall experience. A more robust validation method is needed to verify URLs before rendering them. Implementing a fallback image would ensure that missing or invalid links do not disrupt the interface.

4. Absence of Caching Mechanisms for Frequent Queries

The application currently performs every search request in real time without caching. For commonly searched keywords, caching frequently accessed results would reduce the load on the Elasticsearch server, improve response time, and increase overall system efficiency. Implementing caching provides several additional benefits. First, it minimizes repeated computations by storing previously retrieved search results, allowing the application to return responses instantly without querying Elasticsearch again. This not only speeds up the user experience but also stabilizes the system during periods of high traffic. Furthermore, caching can act as a buffer when Elasticsearch momentarily slows down or experiences connectivity issues, ensuring the website remains responsive.

## Technical difficulties and challenges

A major challenge during development was managing Elasticsearch and Kibana. Because both tools stop running once the terminal session closes, shutting down or restarting the computer required reinstalling or reinitializing Kibana and generating a new authentication token before reconnecting to Elasticsearch. This repeated setup process was time-consuming and often disrupted the workflow. Moreover, since Elasticsearch and Kibana were entirely new technologies for the team, a significant amount of time was devoted to studying, configuring, and experimenting with them to understand indexing, querying, and cluster management concepts.

Data collection also presented notable difficulties. While certain Sanrio characters have extensive and well-documented information, others lack reliable sources or consistent descriptions. In some cases, the information available was too fragmented or incomplete to include with confidence, while a few characters had extremely complex backgrounds that required more effort than the project timeframe allowed. To maintain accuracy and consistency, characters with unverifiable or excessively detailed data were excluded, resulting in a simplified dataset.

Another issue arose from the behavior of the TF-IDF scoring mechanism used by Elasticsearch. Characters containing repeated or dense keywords in their descriptions often received disproportionately high scores, causing them to appear at the top of search results even when they were not the most relevant to the user's query. This occasionally led to search outputs that did not match user expectations. To improve result quality, we adjusted the boost values of key fields such as name and species which carry greater semantic importance. Enhancing the weight of these fields allowed the ranking system to better reflect user intent rather than simply keyword frequency, resulting in more accurate and meaningful search results.

## Lessons learned

This project deepened our understanding of how search engines operate and how information retrieval systems can be applied within real-world web applications. Working with Elasticsearch introduced us to essential concepts such as document indexing, query analysis, ranking techniques, and performance optimization. Throughout development, we realized that structured data and proper preprocessing are fundamental to achieving strong IR performance, as clean and well-organized data directly affects the accuracy and reliability of search results.

In addition, implementing semantic search highlighted how semantic understanding can greatly enhance the user experience, especially for descriptive or natural-language queries. Users often express intent in flexible helps bridge the gap between keyword-based queries and conceptual meaning. Another important takeaway was learning that boosting and ranking require careful tuning depending on the domain. Search relevance is not universal; what matters most varies across datasets. Adjusting field weights such as prioritizing names or species over repeated description keywords taught us how domain-specific tuning can significantly improve result quality.

Beyond technical insights, the project strengthened our ability to troubleshoot unfamiliar tools, manage incomplete datasets, and navigate a steep learning curve. These experiences will be valuable in future development work and are highly applicable in fields such as software engineering, information retrieval, and data science.

## Opportunities for future improvements

One important technical improvement is configuring Elasticsearch and Kibana to run as background services that automatically start when the machine boots. This would eliminate the repeated setup process required after each restart and ensure more reliable, continuous operation.

Another key direction for expansion is supporting multiple languages. Since Sanrio characters originate from Japanese sources and are popular worldwide, adding multilingual support such as Japanese and Thai would allow the search engine to serve a broader audience and provide more accessible information. This also aligns naturally with semantic search, as multilingual embeddings can further enhance query understanding across different languages.

In addition, a major future goal is expanding the dataset to include all Sanrio characters, not just the subset currently available. Many characters are still missing due to limited sources, inconsistent documentation, or time constraints during initial data collection. By continuously gathering and verifying information, the system can be updated to cover the full Sanrio character roster, providing a more complete and valuable resource for users. On the backend, improving query optimization and indexing strategies as our Elasticsearch expertise grows will enable the search engine to scale more effectively. Enhancements such as recommendation features for similar characters and custom embeddings for better semantic understanding can further improve relevance and search accuracy. Together, these improvements will make the system more comprehensive, scalable, and suitable for larger use cases in the future.

## Conclusion

This project successfully delivers a complete search engine system designed specifically for Sanrio characters, combining data preprocessing, indexing, ranking, and a user-friendly interface into one integrated platform. The system allows users to efficiently search through various character attributes, such as name, description, species, debut year, and personality, making it easier for fans to access accurate and well-organized information.

By addressing the lack of centralized and easy-to-use resources for Sanrio characters, the project provides a valuable solution that enhances the fan experience. The clean interface, structured data, and flexible search capabilities demonstrate the effectiveness of our information retrieval approach and reflect the goals of the ITCS441 course.

Overall, this project not only meets the academic requirements but also shows strong potential for real-world use. With future improvements such as expanding the character database, refining the ranking algorithm, or adding personalized recommendations the system could become an even more powerful tool for the Sanrio community. This work represents our dedication to improving user experience through thoughtful design, technical skills, and creative problem-solving.