

Popular Animation Movies Search System

João Pires

UP201806079

Mafalda Costa

UP202109478

Tomás Gonçalves

UP201806763

Agenda

- Collection used for indexing
- Indexing process
- Improvements introduced (new search system)
- Search results for an illustrative query
- Information Needs
- Evaluation of the new version of the search system
 - System Configuration
 - Example for IN. #1
 - Average-Precision and Mean Average-Precision
- Conclusions

Collection used for indexing

6227 documents with the same structure as the example on the right were uploaded to Solr, which represents the collection used in our search system.

```
{
  "id": "ToyStory",
  "Title": "Toy Story",
  "Rated": "G",
  "Release": null,
  "Runtime": "81 min",
  "Genre": "Animation, Adventure, Comedy",
  "Director": "John Lasseter",
  "Writer": "John Lasseter, Pete Docter, Andrew Stanton",
  "Actors": "Tom Hanks, Tim Allen, Don Rickles",
  "Plot": "A little boy named Andy loves to be in his room, playing with his toys, especially his doll named Woody. But, what do the toys do when Andy is not with them, they come to life. Woody believes that his life (as a toy) is good. However, he must worry about Andy's family moving, and what Woody does not know is about Andy's birthday party. Woody does not realize that Andy's mother gave him an action figure known as Buzz Lightyear, who does not believe that he is a toy, and quickly becomes Andy's new favorite toy. Woody, who is now consumed with jealousy, tries to get rid of Buzz. Then, both Woody and Buzz are now lost. They must find a way to get back to Andy before he moves without them, but they will have to pass through a ruthless toy killer, Sid Phillips.",
  "Language": "English",
  "Country": "United States",
  "Awards": "Won 1 Oscar. 27 wins & 23 nominations total",
  "Poster": "https://m.media-amazon.com/images/M/MV5BMDU2ZWU1MjktMTRhMy00ZTA5LWEzNDgtYmNmZTEwZTViZWUkXkEYXkFqcGdeQXVyNDQ2OTk4MzI0._V1_SX300.jpg",
  "Metascore": "95",
  "imdbRating": "8.3",
  "imdbVotes": "930,086",
  "DVD": "23 Mar 2010",
  "BoxOffice": "$223,225,679",
  "Production": null,
  "Website": null,
  "Popularity": 153.86,
```

```
"Poster_Path": "/uXDfjUbdP4i3W5hWSBrPr1Kpxab.jpg",
"Video": false,
"Vote_Average": 8,
"Vote_Count": 14635,
"Reviews": [
  "\nThis movie came out when I was three. Now I'm twenty seven and the goddamn thing still holds up.",
  "_Final rating:\\u2605\\u2605\\u2605\\u2605 - Very strong appeal. A personal favourite._\\n\\n\\nDecided to revisit this after many years and still holds up so well. Great movie for both kids and adults with wonderful teachable moments. Just a groundbreaking animated movie all around. **4.5/5**\\n\\n\\nA stunning feature film entrance from Pixar! <em>'Toy Story'</em> is a true delight, from the first second to the last.",
  "The CGI animation is excellent. All the toys look brilliant, as does the world itself - I love the feel of it. The music is very good, Randy Newman does a nice job - \\n\\n\\nYou've Got a Friend in Me\\n\\n is a cracker.",
  "Tom Hanks leads a strong cast. Hanks plays Woody, to fantastic effect. He is the best part of this film. Tim Allen is great, too, as Buzz Lightyear. Don Rickles (Mr. Potato Head), Wallace Shawn (Rex) and John Ratzenberger (Hamm) also bring fun.",
  "Everything else is just as terrific: the humour, the pacing, the plot - I enjoy it all. There are probably some flaws in parts, namely Buzz's supposed unawareness, but nothing impacts the viewing experience. Go watch!\\n\\n\\n"
],
"Overview": "Led by Woody, Andy's toys live happily in his room until Andy's birthday brings Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody plots against Buzz. But when circumstances separate Buzz and Woody from their owner, the duo eventually learns to put aside their differences."
```

```
}
```

Indexing process

- Indexation for the fields:
 - Title, Rate, Director(s), Writer(s), Actor(s), Plot, Language, Country, Awards, Production and Reviews.
- Use of Docker to accelerate the indexing process.
- Creation of four different custom field types: Text, Number, Date and Link.
- Use of standard and whitespace tokenizers.
- Use of lowercase and ASCII filters.

Improvements introduced (new search system)

- Creation of a Graphical User Interface to allow the user to interact with the search system using a user-friendly interface (use of Flask framework).
- Use of lexemes from Wikidata to overcome the problem of the user queries not matching the specific strings contained in the fields of each movie (in the collection itself).
- Support for cross-language using the Google Translator API for Python, allowing the user to interact with the search system using one of the 106 different languages supported.
- Correction of duplicate results.
- Indexation of the searchable fields, only.
- Use of a new Solr parameter to retrieve not only 10 results, but all possible. Value of 'row' set to 2147483647 (maximum int value).



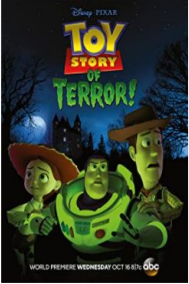




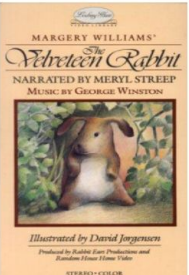
Search results for an illustrative query

Popular Animation Movies

Search

Search results for an illustrative query

Search results for: Toy Story

<p>Toy Story</p>  <p>Details</p>	<p>Lamp Life</p>  <p>Details</p>	<p>Toy Story of Te...</p>  <p>Details</p>	<p>Toy Story 4</p>  <p>Details</p>
<p>Toy Story 2</p>  <p>THE TOYS ARE BACK! THANKSGIVING</p>	<p>Toy Story That ...</p>  <p>TUESDAY DEC 2 8P</p>	<p>Toy Story 3</p>  <p>THE BREAKOUT COMEDY OF THE SUMMER</p>	<p>Little Ears: The ...</p>  <p>MARGERY WILLIAMS' <i>The Velveteen Rabbit</i> NARRATED BY MYRL STREEP MUSIC BY GEORGE WINSTON Illustrated by David Jergensen STEREO • COLOR</p>

Search results for an illustrative query

Search results for: Toy Story

Toy Story 4

Rated: G

Runtime: 100 min

Genre(s): Animation, Adventure, Comedy

Director(s): Josh Cooley

Writer(s): John Lasseter, Andrew Stanton, Josh Cooley

Actor(s): Tom Hanks, Tim Allen, Annie Potts

Plot: Woody, Buzz Lightyear and the rest of the gang embark on a road trip with Bonnie and a new toy named Forky. The adventurous journey turns into an unexpected reunion as Woody's slight detour leads him to his long-lost friend Bo Peep. As Woody and Bo discuss the old days, they soon start to realize that they're two worlds apart when it comes to what they want from life as a toy.

Language: English

Country: United States

Awards: Won 1 Oscar. 56 wins & 64 nominations total

Metascore: 84






imdbRating: 7.7

imdbVotes: 224159

DVD: 01 Oct 2019

BoxOffice: \$434,038,008

Close



Information Needs

1. IN."Which is the name of the movie where the toys gain life?"
Q."toys that have life"
2. IN."Who was the Hollywood actor that was part of a Christmas movie where a kid goes on a train to see Santa Claus?"
Q."kid on trip to see santa claus"
3. IN."What is the imdb score of the movie where an ogre who's voice was made by Mike Myers meets his wife's parents?"
Q."ogre voices by Mike Myers meets wifes parents"
4. IN."In which year was released the first Ice Age movie with Sid as a character?"
Q."animals ice sid"
5. IN."Was the animation movie where animals escape from the Central Park zoo awarded?"
Q."central park zoo animals"

Evaluation of the new version of the search system (System Configurations)

- System 1: System without Query Phrase Slop, Phrase Slop, Tie, Phrase Fields and Boost parameters.
- System 2: System without Phrase Fields and Boost parameters.
- System 3: System without Boost Fields.
- System 4: System with the standard query parameters, defined in the previous section.
- System 5: System with a higher weight on the Title (10) than the Plot (6) and Overview(6) fields, and a higher boost on the 'excellent' and 'recommend*' words.

Evaluation of the new version of the search system (Example for IN. #1)

	System 1	System 2	System 3	System 4	System 5
1	1	1	1	1	1
2	1	1	1	1	0
3	1	1	0	1	0
4	1	1	1	1	1
5	1	1	1	1	0
6	1	0	1	0	0
7	1	1	0	1	0
8	1	1	0	0	0
9	1	1	0	1	0
10	0	1	1	0	1

Table 5: Relevance table for the first Information Need.

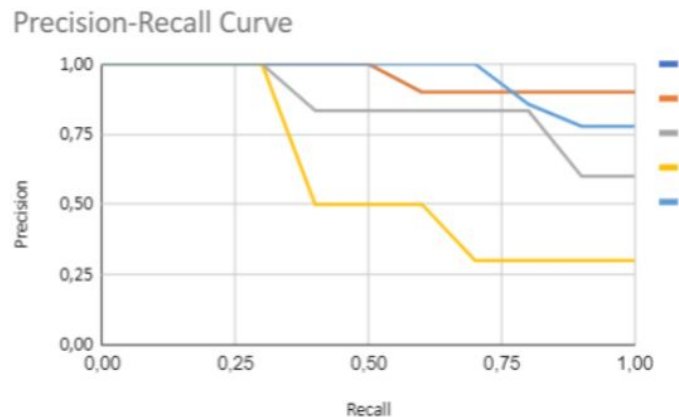


Figure 6: Precision-Recall Curve for the query "toys that have life".

Evaluation of the new version of the search system (Average-Precision and Mean Average-Precision)

	IN1	IN2	IN3	IN4	IN5
System 1	1	0,23	0,98	0,85	0,78
System 2	0,95	0,33	0,95	0,98	0,83
System 3	0,83	0,87	0,83	1	0,46
System 4	0,94	0,29	0,83	1	0,70
System 5	0,6	0,92	0,74	1	0,57

Table 3: Average-Precision table.

	System 1	System 2	System 3	System 4	System 5
MAP	0,77	0,80	0,80	0,75	0,76

Table 4: Mean Average-Precision table.

Conclusions

The objective of the project was achieved by developing the final version of the search system, starting from the data processing to data analysis and then finalizing with the information retrieval, which included the indexation, the queries, the evaluation and the development of additional features, such as the cross-language support and the GUI.

The current state of the search system allows the user to use the GUI to search for almost every animation movie in the world, using any language, obtaining the results in a displayed user-friendly interface.

Eventual future work would include the application of the same translation functions used for the user query to the translation of the results so that the GUI could be accessed and displayed in a language chosen by the user.