



## **Project 2: Prototype IR System**

**By**

<b>Miss Kanyavee</b>	<b>Likitwattanakij</b>	<b>6488038</b>
<b>Miss Arada</b>	<b>Raksapatcharawong</b>	<b>6488066</b>
<b>Mr. Sarttra</b>	<b>Prasongtichol</b>	<b>6488124</b>

**A Report Submitted in Partial Fulfillment of  
the Requirements for  
ITCS414 Information Retrieval and Storage  
Faculty of Information and Communication Technology  
Mahidol University  
2023**

**a. Introduction**

In the vast expanse of the digital realm, the search for spine-chilling and hair-raising horror movies can be as daunting as navigating a haunted maze. Recognizing the harrowing challenges faced by enthusiasts in discovering the most bone-chilling cinematic experiences, we introduce a groundbreaking solution - the Horror Movie Search Engine. This innovative platform aims to address prevalent issues within the realm of horror movie exploration, presenting a user-centric approach that transcends the limitations of existing systems.

**b. Problem(s) that you are trying to solve.**

- 1. Obscure Discoverability:** The sheer volume of horror movies, spanning various sub-genres and eras, often leaves horror enthusiasts overwhelmed and struggling to unearth hidden gems. Our Horror Movie Search Engine is designed to combat this obscurity, offering a tailored search experience that illuminates even the darkest corners of horror cinema.
- 2. Subjective Fear Thresholds:** Horror is a deeply personal experience, and individuals possess varying thresholds for fear. Conventional movie databases fail to capture the nuanced preferences of users, leading to recommendations that may not align with their unique tastes. Our solution employs advanced algorithms to understand and adapt to users' subjective fear thresholds, providing a personalized horror journey.
- 3. Ineffective Categorization:** Existing systems often rely on generic categorizations that fall short in encapsulating the diverse spectrum of horror sub-genres. From psychological thrillers to supernatural horrors, our search engine employs a refined categorization system that ensures users can pinpoint their preferred brand of terror with precision.

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

**A Glimpse into the Shadows:**

While several movie databases and streaming platforms exist, few cater specifically to the distinct needs of horror aficionados. Conventional algorithms lack the depth needed to understand the intricacies of fear, often leading to hit-or-miss recommendations. The Horror Movie Search Engine distinguishes itself by delving into the unique nuances of horror, ensuring that users are presented with selections that resonate with their deepest fears and desires.

## d. Implementation

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

#### Data collection

We use data from “horror\_movie.csv” but there is less text information in “plot” field, so we decide to add overview, starring, and director to increase more text information. Next, we change the format of csv into ndjson to make it be able to put in elasticsearch\_loader.py.

#### Example documents

```
("Number":1,"original_title":"Orphan: First Kill","title":"Orphan: First Kill","original_language":"en","overview":"Orphan: First Kill is a 2022 American ps
("Number":2,"original_title":"Beast","title":"Beast","original_language":"en","overview":"Beast is a 2022 survival action horror-thriller film directed by B
("Number":3,"original_title":"Smile","title":"Smile","original_language":"en","overview":"Smile is a 2022 American psychological supernatural horror film wr
("Number":4,"original_title":"The Black Phone","title":"The Black Phone","original_language":"en","overview":"The Black Phone is a 2021 American supernatura
("Number":5,"original_title":"Presencias","title":"Presences","original_language":"es","overview":"Presencias (lit.'Presences') is a 2022 Mexican horror th
("Number":6,"original_title":"Jeepers Creepers: Reborn","title":"Jeepers Creepers: Reborn","original_language":"en","overview":"Jeepers Creepers: Reborn is
("Number":7,"original_title":"Nope","title":"Nope","original_language":"en","overview":"Nope (stylized in all caps) is a 2022 American science fiction horro
("Number":8,"original_title":"X","title":"X","original_language":"en","overview":"X is a 2022 American slasher film written, directed, produced and edited b
("Number":9,"original_title":"The Cellar","title":"The Cellar","original_language":"en","overview":"The Cellar is a 2022 supernatural horror film written an
("Number":10,"original_title":"My Best Friend's Exorcism","title":"My Best Friend's Exorcism","original_language":"en","overview":"My Best Friend's Exorcism
("Number":11,"original_title":"Day Shift","title":"Day Shift","original_language":"en","overview":"Day Shift is a 2022 American action comedy film directed
("Number":12,"original_title":"Shut In","title":"Shut In","original_language":"en","overview":"Shut In is a 2022 American thriller film directed by D.J. Car
("Number":13,"original_title":"La abuela","title":"The Grandmother","original_language":"es","overview":"The Grandmother (Spanish: La abuela) is a 2021 horr
("Number":14,"original_title":"Dahmer","title":"Dahmer","original_language":"en","overview":"Dahmer is a 2002 American drama film written and directed by Da
("Number":15,"original_title":"Halloween Kills","title":"Halloween Kills","original_language":"en","overview":"Halloween Kills is a 2021 American slasher fi
("Number":16,"original_title":"Veneciafrenia","title":"Veneciafrenia","original_language":"es","overview":"Venicephrenia (Spanish: Veneciafrenia) is a 2021
("Number":17,"original_title":"Texas Chainsaw Massacre","title":"Texas Chainsaw Massacre","original_language":"en","overview":"Texas Chainsaw Massacre is a
("Number":18,"original_title":"Scream","title":"Scream","original_language":"en","overview":"Scream is a 1996 American slasher film directed by Wes Craven a
("Number":19,"original_title":"Malnazidos","title":"Valley of the Dead","original_language":"es","overview":"Valley of the Dead (Spanish: Malnazidos) is a 2
("Number":20,"original_title":"Halloween Ends","title":"Halloween Ends","original_language":"en","overview":"Halloween Ends is a 2022 American slasher film
("Number":21,"original_title":"Smile","title":"Smile","original_language":"it","overview":"Smile is a 2009 English language Italian horror/thriller film sta
```

```
{
  "Number": 160,
  "original_title": "Scream",
  "title": "Scream",
  "original_language": "en",
  "overview": "Scream is a 1996 American slasher film directed by Wes Craven and written by Kevin Williamson. It stars David Arquette, Neve
  "plot": "A killer known as Ghostface begins killing off teenagers, and as the body count begins rising, one girl and her friends find the
  "tagline": "Someone has taken their love of scary movies one step too far.",
  "Starring": "David Arquette,Neve Campbell,Courteney Cox,Matthew Lillard,Rose McGowan,Skeet Ulrich,Drew Barrymore",
  "Director": "Wes Craven",
  "Writer": "Kevin Williamson",
  "poster_path": "https://image.tmdb.org/t/p/original/303klyyPaZBBE4n7IngzTomRDp.jpg",
  "film_year": "1996 film",
  "runtime": 112,
  "genre_names": "Crime, Horror, Mystery"
}

{
  "Number": 161,
  "original_title": "Hellraiser",
  "title": "Hellraiser",
  "original_language": "en",
  "overview": "Hellraiser is a 1987 British supernatural horror film written and directed by Clive Barker, and produced by Christopher Figg
  "plot": "When he tinkers with a box he bought while abroad, sexual deviant Frank inadvertently opens a portal to hell, where fetish-demon
  "tagline": "He'll tear your soul apart.",
  "Starring": "Andrew Robinson, Clare Higgins, Ashley Laurence,",
  "Director": "Clive Barker",
  "Writer": "Clive Barker",
  "poster_path": "https://image.tmdb.org/t/p/original/nnYHvA2tSVdtr8KxiY8F6YndCOy.jpg",
  "film_year": "1987 film",
  "runtime": 93,
  "genre_names": "Fantasy, Horror"
```

## Data Statistics

### File stats

All fields

14

of 14 total

Number fields

2

of 2 total

Field name

14

Field type

3

>	Type	Name ↑	Documents (%) ⓘ	Distinct values	Distributions ⓘ
>	k	Director	161 (100%)	138	<div> <div></div> <div>top 10 of 138 categories</div> </div>
>	#	Number	161 (100%)	161	<div> <div>min</div> <div>median</div> <div>max</div> <div>1</div> <div>81</div> <div>161</div> </div>
>	f	Starring	161 (100%)	160	
>	f	Writer	161 (100%)	136	
>	k	film_year	161 (100%)	34	<div> <div></div> <div>top 10 of 34 categories</div> </div>
>	k	genre_names	161 (100%)	44	<div> <div></div> <div>top 10 of 44 categories</div> </div>
>	k	original_language	161 (100%)	8	<div> <div></div> <div>8 categories</div> </div>
>	f	original_title	161 (100%)	151	
>	f	overview	161 (100%)	159	
>	f	plot	161 (100%)	161	
>	k	poster_path	161 (100%)	161	<div> <div></div> <div>top 10 of 161 categories</div> </div>
>	#	runtime	161 (100%)	60	<div> <div>min</div> <div>median</div> <div>max</div> <div>0</div> <div>97</div> <div>170</div> </div>
>	f	tagline	161 (100%)	146	
>	f	title	161 (100%)	151	

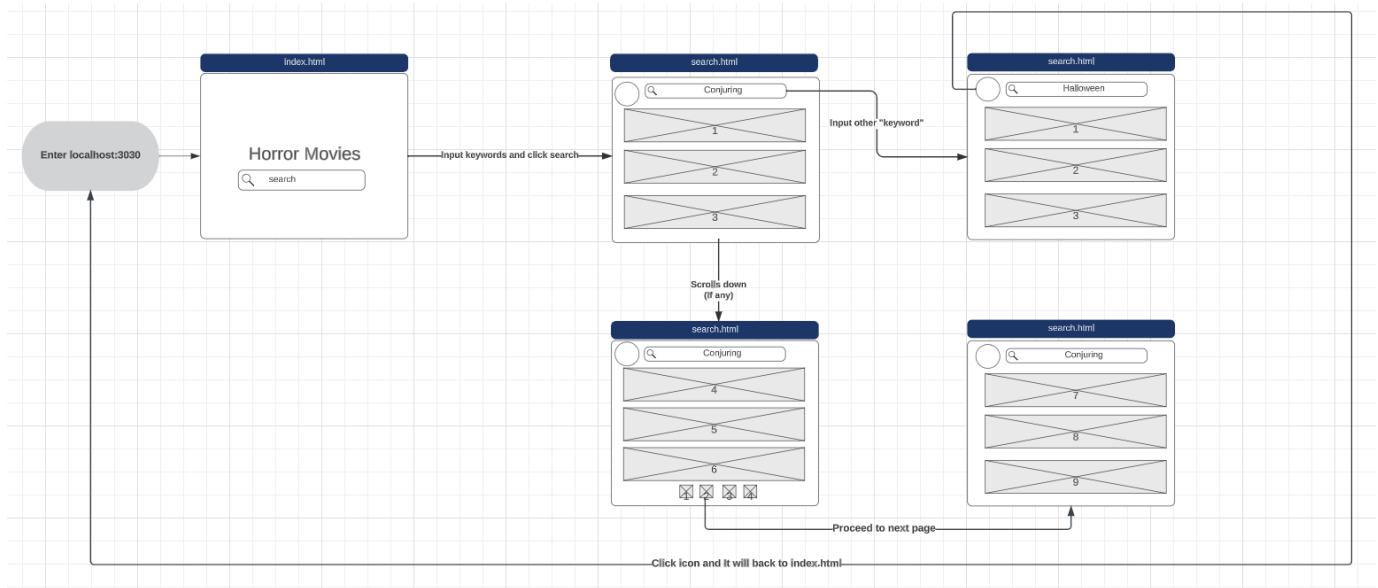
Rows per page: 25

< 1 >

### ii. Tools and software

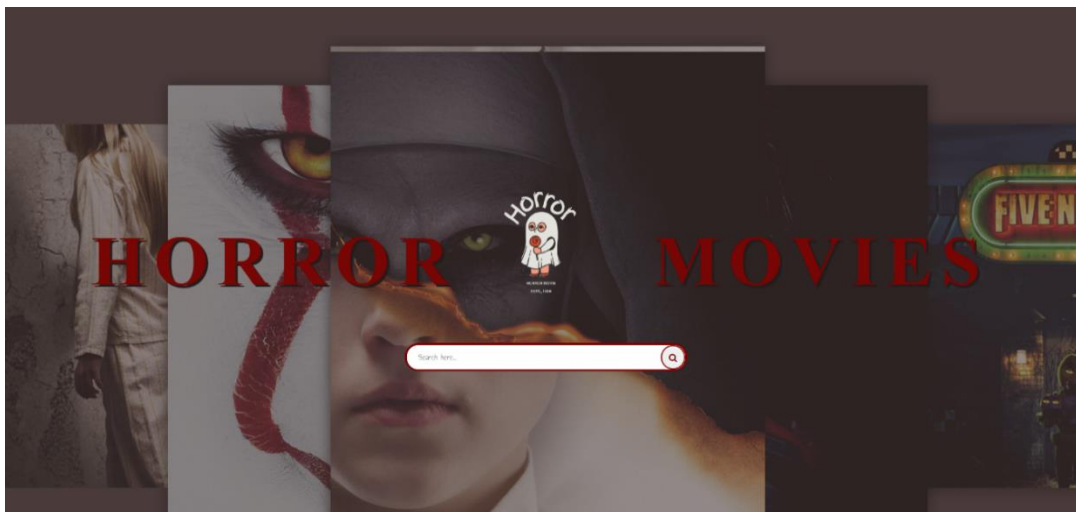
- **Visual Studio Code:** Use to modify and customize every source code in this project.
- **Window Power Shell:** Use to run Elasticsearch and kibana server.
- **Anaconda Prompt:** Use to create virtual environment and run flask server.
- **Python:** Use to create `horror_search_app.py` to act as flask app to run our engine and `elasticsearch_loader.py` to load our ndjson file to Elasticsearch server.
- **Elasticsearch:** Use to store index and documents and check for completeness of documents.
- **Microsoft Excel:** use to create modify `horror_movie.csv`.
- **JSON Converter:** We use the converter from <https://konbert.com/convert/json/to/ndjson> to convert our csv file to ndjson.
- **Web Browser:** Use to display interface of Search Engines

### iii. System diagram

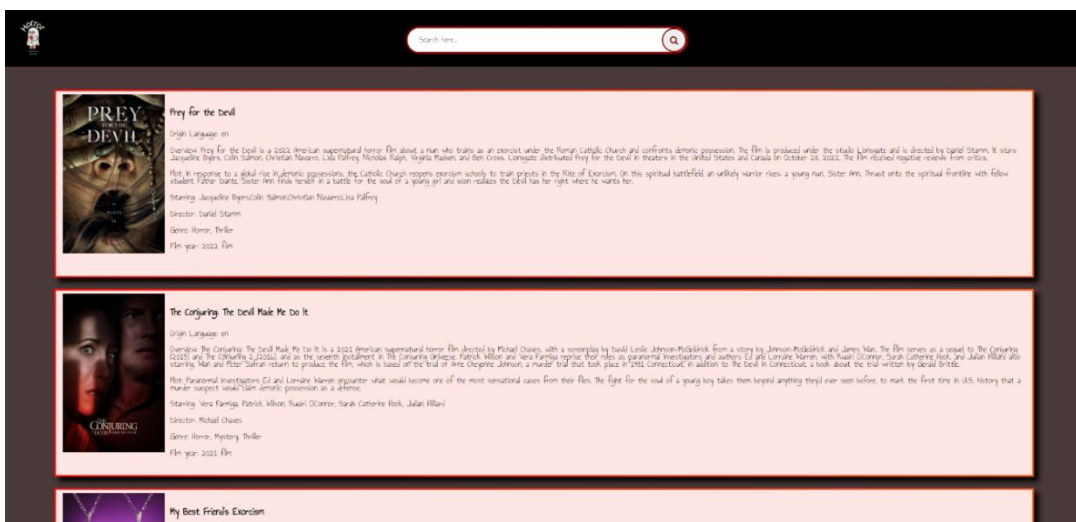


### iv. Snapshots of the system

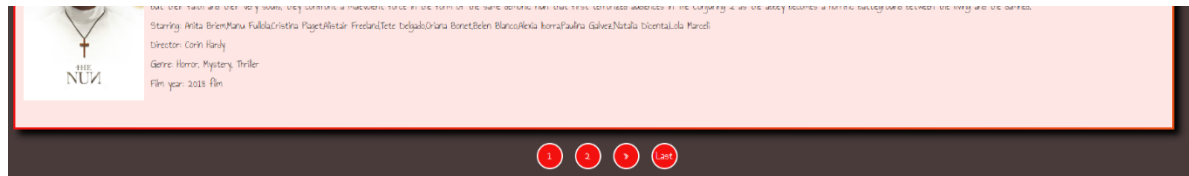
#### Index page



#### Search Result Page



## Page Navigation

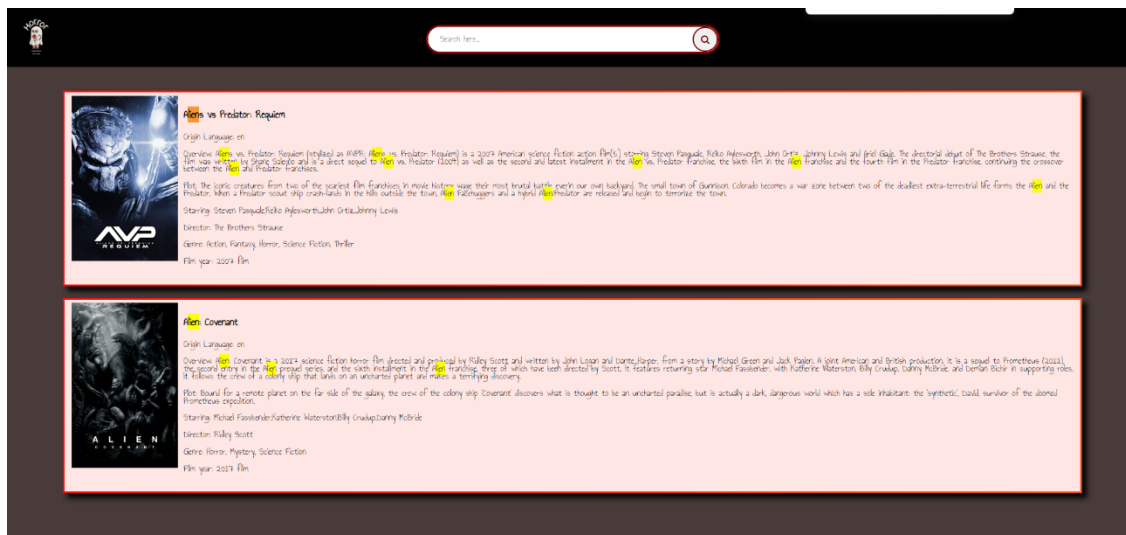


**v. Example step-by-step search sessions that highlight the following functionality:**

### 1. One word query

1.1 Input keyword “Alien” then click search

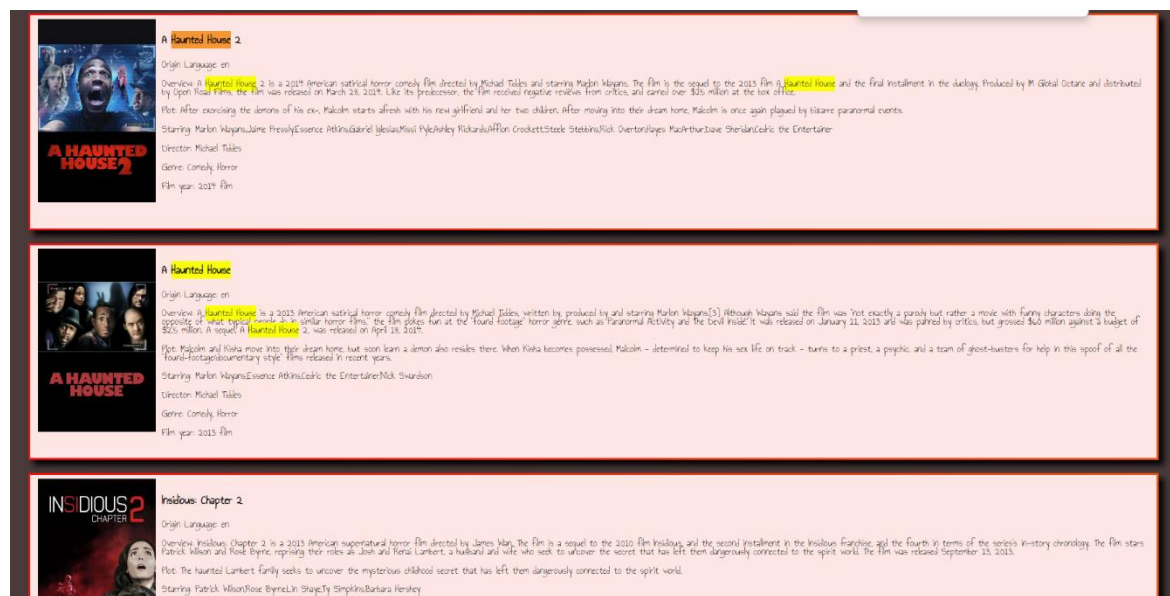
1.2 The result is shown as following:



## 2. *Multiple word query*

2.1 Input keyword” The Haunted House” then click search.

2.2 The result is shown as following:



### 3. Partial match

1.1 Input “..out..” then click search

1.2 The result will show the movie that have “out” only at the beginning of the result then show the other that contain “..out..”



#### GET OUT

Get **out**

Origin Language: en

Overview: **Get Out** is a 2017 American psychological horror film written, co-produced and directed by Jordan Peele. It stars Daniel Kaluuya, Allison Williams, Lili Finiucci, Lakeith Stanfield, Bradley Whitford, Caleb Landry Jones, Stephen Root, and Catherine Keener. The plot follows a young black man (Kaluuya), who answers his dating app's request when he meets the family of his white girlfriend (Williams).

Plot: Chris and his girlfriend Rose go upstate to visit her parents for the weekend. At first, Chris masks the family's eerily accommodating behavior as nervous attempts to deal with their daughter's interracial relationship, but as the weekend progresses, a series of increasingly disturbing discoveries lead him to a truth that he never could have imagined.

Starring: Daniel Kaluuya, Allison Williams, Bradley Whitford, Lakeith Stanfield, Caleb Landry Jones, Stephen Root, Catherine Keener

Director: Jordan Peele

Genre: Horror, Mystery, Thriller

Film year: 2017 Film



#### X

X

Origin Language: en

Overview: **X** is a 2022 American slasher film written, directed, produced and edited by Ti West. It stars Mia Goth in dual roles as a young woman named Maxine and an elderly woman named Pearl. The film also stars Jenna Ortega, Martin Henderson, Brittany Snow, Owen Campbell, Stephen Ure and Scott Mescudi appearing in supporting roles. The film follows a cult and crew who gather to make a pornographic film on an elderly couple's rural Texas property, but find themselves threatened by the homicidal couple.

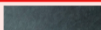
Plot: In 1978, a group of young filmmakers set **out** to make an adult film in rural Texas, but when their reclusive, elderly hosts catch them in the act, the cast find themselves fighting for their lives.

Starring: Mia Goth, Jenna Ortega, Martin Henderson, Brittany Snow, Owen Campbell, Stephen Ure, Scott Mescudi


Director: Ti West

Genre: Horror, Mystery, Thriller

Film year: 2022 Film



#### The Medium



#### The Medium

Origin Language: th

Overview: **The Medium** is a 2021 Thai-South Korean mediumistic supernatural folk horror film co-written and produced by No Vithairat and directed by Banjong Pisanthanakun. It is a co-production of Thailand's GMM 25 and South Korea's Showbox. The film premiered at the 26th Bucheon International Fantastic Film Festival on 11 July 2021. It was theatrically released in **South Korea** on 17 July 2021. It was selected as the Thai entry for the Best International Feature Film at the 94th Academy Awards but was not nominated.


Plot: A horrifying story of a shaman inheritance in the hazy region of Thailand. But the goddess that appears to have taken possession of a family member turns **out** not to be as benevolent as it first appears.

Starring: Naniha Gulmongkolpech, Sawanee Utoomma, Sirin Yankietkan, Yaaska Chaisorn

Director: Banjong Pisanthanakun

Genre: Horror, Thriller

Film year: 2021 Film



#### Freddy vs. Jason

Origin Language: en

Overview: **Freddy vs. Jason** is a 2003 American slasher film directed by Ronny Yu and written by Ianian Shannon and Mark Swift. It is a crossover between the A Nightmare on Elm Street and Friday the 13th franchises, being the eighth installment in the former and the seventh in the latter. The film joins the two series in a shared universe and pits their respective antagonists, Freddy Krueger and Jason Voorhees, against each other. Freddy is weakened and forgotten because the citizens of his home town Springwood have defeated him by using medications that suppress dreams. Freddy awakens Jason to stir up fear and grow his powers so that he may return and kill again. Jason turns **out** to not be as easily controlled as Freddy initially thought and the two supernatural mass murderers come into conflict. The film is chronologically set after Freddy's death in The Final Nightmare (1991) and Jason Goes to Hell: The Final Friday (1993) and is the last film in each franchise before their respective reboots.

Plot: In an attempt to free himself from a state of forgotten limbo, evil dream-demon Freddy Krueger devises a plan to manipulate undead mass murderer Jason Voorhees into slicing-and-dicing his way through the teenage population of Springwood. But when the master of dreams loses control of his monster, a brutal fight to the death is the only way **out**.

Starring: Monica Keaton, Kelly Rowland, Jason Ritter, Chris Marquette

Director: Ronny Yu

Genre: Horror


Film year: 2003 Film



#### 4. Ranking

4.1 Input keyword “Nun” then click search

4.2 The ranking will be weight on plot field then overview field. You can observe that Nun contain in both ‘overview’ and ‘plot’ field in “The Nun”  
So this is Rank 1 and descending as follow:



**The Nun**

Origin Language: en

Overview: The **Nun** (Spanish: La Monja) is an English-language 2018 Spanish horror film directed by Luis De La Madrid.


Plot: When a young **nun** at a disordered abbey in Romania takes her own life, a priest, with a haunted past and a nightmare on the threshold of her final vows are sent by the Vatican to investigate. Together they uncover the order's unholy secret. Risking not only their lives but their faith and their very souls, they confront a malevolent force in the form of the same demonic **nun** that first terrorized audiences in The Conjuring 2 as the abbey becomes a horrific battleground between the living and the damned.

Starring: Anita Briem,Manu Falcu,Cristina Păcurar,Victor Fredlund,Tete Bodalo,Orlana Boretti,Belen Blanco/Alexia Kora,Paulina Gálvez,Natalia Dicenta,Lola Marcel

Director: Corin Hardy

Genre: Horror, Mystery, Thriller

Film year: 2018 Film



**PREY FOR THE DEVIL**

Origin Language: en

Overview: Prey for the Devil is a 2022 American supernatural horror film about a **nun** who trains as an exorcist under the Roman Catholic Church and confronts demonic possession. The film is produced under the studio Lionsgate and is directed by Daniel Stamm. It stars Jacqueline Byers, Colin Salmon, Christian Navarro, Lida Palfrey, Nicholas Ralph, Virginia Madsen, and Ben Cross. Lionsgate distributed Prey for the Devil in theaters in the United States and Canada on October 28, 2022. The film received negative reviews from critics.


Plot: In response to a global rise in demonic possessions, the Catholic Church reopens exorcism schools to train priests in the Rite of Exorcism. On this spiritual battlefield, an unlikely warrior rises: a young **nun**, Sister Ann. Truist onto the spiritual frontline with fellow student Father Dante, Sister Ann finds herself in a battle for the soul of a young girl and soon realizes the devil has her right where he wants her.

Starring: Jacqueline Byers,Colin Salmon,Christian Navarro,Lida Palfrey

Director: Daniel Stamm

Genre: Horror, Thriller

Film year: 2022 Film



**A NUN'S CURSE**

Origin Language: en

Overview: A **Nun's** Curse is a 2019 film directed by Tommy Färrelth .


Plot: A group of travelers are forced to seek shelter inside an abandoned jail where a notorious **nun** named Sister Monday had once been assigned and was suspected of murdering prisoners.

Starring: Felissa Rose,Daniel Ruffei,Enka Arlec,Kristi Ray

Director: Tommy Färrelth

Genre: Horror

Film year: 2019 Film



**ANNABELLE: CREATION**

Origin Language: en

Overview: Annabelle: Creation is a 2017 American supernatural horror film directed by David F. Sandberg, written by Gary Dauberman and produced by Peter Safian and James Wan. It is a prequel to 2017's Annabelle and the fourth installment in The Conjuring Universe franchise. The film stars Stephanie Sigman, Talitha Bateman, Lulu Wilson, Anthony LaPaglia, and Miranda Otto, and depicts the possessed Annabelle doll's origin.

Plot: Several years after the tragic death of their little girl, a doll maker and his wife welcome a **nun** and several girls from a shuttered orphanage into their home, soon becoming the target of the doll maker's possessed creation,Annabelle.

Starring: Stephanie Sigman,Talitha Bateman,Lulu Wilson,Anthony LaPaglia,Miranda Otto

Director: David F. Sandberg

Genre: Horror, Mystery, Thriller

Film year: 2017 Film



## e. Discussion

### i. **Limitations of your system**

#### Limited Movies

Because there are only 161 documents or horror movies in the system so it could lead to incomplete movie sets such as alien movies that normally have 8 movies, but our system provides only 2 movies.

#### Irrelevant search result

When putting and searching for some keyword, the result could show some result that does not contain any relevant keywords which can make other people confused.

### ii. **Technical difficulties, challenges, and lessons learned.**

#### Technical difficulties

- **Document file completeness and format**

Since we use horror\_movies.csv and we have to transform the file into ndjson. However, after transformation, we have to delete numerous special character which have effect on error when use elasticsearch\_loader.py.

- **Flask Activation**

We have to manage the backend through flask and python. The difficulty is the setup part which have to configure virtual environment to make the search engines run as well as install necessary modules to the env.

- **Html & css customization**

since we didn't use this field of knowledge for long time, so it takes sus for a time to customize our search engine interface. Especially on search.html which display search result that have some complexity on display poster picture and description in the same line but at last we manage to accomplish this.

#### Challenges

- **Document preparation**

horror\_movies.csv have an issue in less text information so we have to add overview, starring, director from Wikipedia to the csv file before transform to ndjson file and have to decrease information to 161 documents to make it on time.

- **Poster URL**

since for field "poster\_path" have to use url for the picture to display in the search result so we have to use the template link url from the movie database or tmdb.org and apply to all data in the field of poster\_path.

- **Confusing Ranking**

We attempt to implement ranking source code in the system. However, after trying to search for the desired keyword, it appears that the result

does not arrange correctly with some part of result swap with the other ones that should be in first rank.

### **Lessons learned**

- **Recap knowledge on making website.**

We are able to use old knowledge from web programming to create our own search engines web platform which include both index and search page that we apply html and css knowledge for this.

- **Learn how to make search engine using Elasticsearch, Flask**

We learn another way to run the website on Flask which uses programming language. we can use this to develop our search engine in the future. Moreover, we realized the necessary field of information in document that we have to include in document and learnt about elasticsearch function that make us be able to solve some problem like duplicated index which can use DELETE command to delete them.

### **iii. Opportunities for future improvements**

#### **Expand Movie Database:**

- Import more movies from diverse sources, focusing on popular and classic horror films.
- Consider including essential metadata like release date, director, and cast.

#### **Enhance Search Options:**

- Implement advanced search filters such as genre, release year, and ratings.
- Include sorting options for relevance, ratings, and popularity.

#### **User-Friendly Interface:**

- Optimize the interface for both desktop and mobile users.
- Incorporate visually appealing elements, like movie posters, in the search results.

#### **Additional Information Tab:**

- Create a separate information tab for each movie with a brief synopsis and related recommendations.

#### **Personalized User Accounts:**

- Introduce user accounts for saving preferences and bookmarking movies.
- Consider a simple rating system for personalized recommendations.

### **Feedback Mechanism:**

- Include a feedback option for users to provide suggestions or report issues.

### **f. Conclusion**

In conclusion, our endeavor to create the Horror Movie Search Engine represents a significant stride in addressing the challenges faced by horror enthusiasts in navigating the vast landscape of cinematic fear. Through meticulous data collection, thoughtful implementation, and the utilization of advanced algorithms, we have crafted a platform that goes beyond conventional movie databases. Our solution not only tackles the obscurity of discovering hidden gems but also recognizes the subjective nature of fear, adapting recommendations to each user's unique preferences. By refining categorization and delving into the intricate nuances of horror, our search engine stands out as a tailored and comprehensive tool for horror aficionados.

The implementation phase, involving data enrichment and system development, showcased our commitment to creating a user-friendly experience. Utilizing tools such as Visual Studio Code, Elasticsearch, and Python, we meticulously designed a system that not only meets technical requirements but also ensures a seamless interaction for users. The system snapshots and step-by-step search sessions provided a glimpse into the user interface, emphasizing its simplicity and effectiveness.

Despite our achievements, we acknowledge certain limitations. Technical challenges, such as the scarcity of text information in some movie plots, presented hurdles that influenced our data collection strategy. Additionally, lessons learned from these challenges underscore the importance of continuous refinement and adaptation in the dynamic field of information retrieval.

Looking ahead, there are ample opportunities for future improvements. User feedback and evolving trends in horror cinema can guide enhancements to our algorithms, ensuring they remain attuned to the ever-changing landscape of fear. Collaborations with streaming platforms and expanded data sources can further enrich the content available on our platform, enhancing the overall user experience.

In essence, our prototype IR system marks the beginning of a journey to redefine the way horror enthusiasts engage with their favorite genre. As we navigate the shadows of technical challenges and embrace opportunities for improvement, our commitment to providing a curated and personalized horror movie exploration experience remains unwavering. We believe that the Horror Movie Search Engine has the potential to become a beacon for horror enthusiasts, guiding them through the darkness to discover the films that truly resonate with their deepest fears and desires.

