



Report
Movie Search Engine

By

6388113	Poomrapee	Wareeboutr
6388133	Pitchaya	Teerawongpairoj
6388196	Sasima	Srijanya

Present to

Assoc. Prof. Dr. Suppawong Tuarob

**A Report Submitted in Partial Fulfillment of
the Requirements for**

**ITCS371 Introduction to Software Engineering
Faculty of Information and Communication Technology
Mahidol University
Semester 1/2022**

Table Of Contents

Introduction.....	2
Problems.....	2-3
Existing relevant systems.....	4
Data collection, example documents, and data statistics.....	4-5
Tools and software.....	6
System diagram.....	7
Snapshots of the system.....	8-9
Example step-by-step.....	9-13
Discussion.....	14
Limitations of your system.....	14
Technical difficulties, challenges, and lessons learned.....	14
Opportunities for future improvements.....	15
Conclusion.....	15
References.....	16

Introduction

According to our group members who like watching movies, our group decided to do 'Movie Search Engine' which is a web search engine for searching movies.

This report explains the details of movie search engines including the problem of the movie search engine, existing relevant system, discussion, and conclusion.

Problems

1. Dataset

Our group tried to find data from Kaggle that contain all information we needed including:

- Movie's title
- Movies' plot
- Movies' genre
- Original language of movies
- Actors, director
- Movies' rating
- Release date
- Run time of movies
- IMDb id
- Link of movies' poster
- Link to movies's IMDb homepage

But we cannot find all data in one dataset. So, our group decided to use four datasets (four resources) and merge them together into one dataset. After merging all datasets, we have around 1,500 data. You can see all the dataset that our group used in the references.

2. Elastic

The problem we found when implementing Elasticsearch is we cannot build the inverted index for our dataset. Elastics trigger the "File structure cannot be determined" error as shown in the below picture.

More ways to add data

In addition to adding [integrations](#), you can try our sample data or upload your own data.

[Sample data](#) [Upload file](#)

× File structure cannot be determined

[illegal_argument_exception: [illegal_argument_exception] Reason: Cannot create a multi-line start pattern. No suitable column to match exists before the first column whose values contain line breaks or delimiters [title]. If the timestamp format was not identified correctly adding an override for this may help.]: Cannot create a multi-line start pattern. No suitable column to match exists before the first column whose values contain line breaks or delimiters [title]. If the timestamp format was not identified correctly adding an override for this may help.

Figure 1 - Elastic error message

We find the issue following the instruction Elastic told us “if the time timestamp format was not identified correctly adding an override for this may help”. Since Elastic does not provide a time format that matches the dataset, we try to override the format of the time field with our custom format (dd/MM/yyyy). Finally, the override setting works.

movies_4.2.csv

File contents

First 1,592 lines

```
1 release_date,title,overview,popularity,vote_count,vote_average,original_language,genre,poster_url,imdb_id,cast,director,runtime
2 30/10/1995,Toy Story,"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
3 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
4 differences.",171.714,15127,8,en,"Animation, Adventure, Family, Comedy",https://image.tmdb.org/t/p/original/uXDfj3bD4ijH5hW5B8rPr1Kpxab.jpg,tt0114709,Tom
5 Hanks|Tim Allen|Don Rickles|Jim Varney|Wallace Shawn|John Lasseter,81,,
6 15/12/1995,Jumanji,"When siblings Judy and Peter discover an enchanted board game that opens the door to a magical world, they unwittingly invite Alan -- an adult
7 who's been trapped inside the game for 26 years -- into their living room. Alan's only hope for freedom is to finish the game, which proves risky as all three
8 find themselves running from giant rhinoceroses, evil monkeys and other terrifying creatures.",16.597,8826,7.2,en,"Adventure, Fantasy, Family",https://image
9 .tmdb.org/t/p/original/6eGn2X51bahFo0I8wE1h2VGtGch.jpg,tt0113497,Robin Williams|Jonathan Hyde|Kirsten Dunst|Bradley Pierce|Bonnie Hunt|Joe Johnston,104,,
10 8/12/1995,Father of the Bride Part II,"Just when George Banks has recovered from his daughter's wedding, he receives the news that she's pregnant ... and that
11 George's wife, Nina, is expecting too. He was planning on selling their home, but that's a plan that -- like George -- will have to change with the arrival of
12 both a grandchild and a kid of his own.",17.343,559,6.2,en,"Comedy, Family",https://image.tmdb.org/t/p/original/rj4L8twQ0uGrp8nCELr716Qo3mw.jpg,tt0113041,Steve
13 Martin|Diane Keaton|Martin Short|Kimberly Williams-Paisley|George Newbern|Charles Shyer,106,,
14 15/12/1995,Heat,"Obsessive master thief, Neil McCauley leads a top-notch crew on various insane heists throughout Los Angeles while a mentally unstable detective,
```

Summary

Number of lines analyzed	1592
Format	delimited
Delimiter	,
Has header row	true
Time field	release_date
Time formats	dd/MM/yyyy, M/dd/yyyy, MM/d/yyyy, M/d/yyyy

[Override settings](#)

[Analysis explanation](#)

Figure 2 - Elastic with override setting (Time formats)

Existing relevant systems

Nowadays, there are many movie search engines because people like to search movies to select movies that they are interested in to find more information about the movie. The most popular movie search engines are IMDb and RottenTomatoes.

For example:

- https://www.rottentomatoes.com/m/search_engines
- <https://www.imdb.com/search/>
- <https://www.allmovie.com/advanced-search>
- <https://www.justwatch.com/uk>
- <https://reelgood.com/>
- <https://www.whatismymovie.com/>

Data collection, example documents, and data statistics

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Release_Date	Title	Overview	Popularity	Vote_Count	Vote_Average	Original_Language	Genre	Poster_Url						
2	12/15/2021	Spider-Man: No Way Home	Peter Park	5083.954	8940	8.3	en	Action, Ad	https://image.tmdb.org/t/p/original/1g0dhYtq4irTY1GPXvft6k4YlJm.jpg						
3	3/1/2022	The Batman	In his seco	3827.658	1151	8.1	en	Crime, My	https://image.tmdb.org/t/p/original/74xTEgt7R36Fpoo050r9T25onhq.jpg						
4	2/25/2022	No Exit	Stranded a	2618.087	122	6.3	en	Thriller	https://image.tmdb.org/t/p/original/vDHsLnOWKIPGmW50kGfuhNF4w5l.jpg						
5	11/24/2021	Encanto	The tale of	2402.201	5076	7.7	en	Animation	https://image.tmdb.org/t/p/original/4j0PNHkMr5ax3IA8tjxcmlPU3QT.jpg						
6	12/22/2021	The King's Man	As a collec	1895.511	1793	7	en	Action, Ad	https://image.tmdb.org/t/p/original/aq4Pww5Xeuvi6HZKbyd23e6bE9.jpg						
7	1/7/2022	The Commando	An elite DE	1750.484	33	6.6	en	Action, Cri	https://image.tmdb.org/t/p/original/pSh8MyYu5CmfyWEHzy8FEARH2zq.jpg						
8	1/12/2022	Scream	Twenty-fiv	1675.161	821	6.8	en	Horror, My	https://image.tmdb.org/t/p/original/kzNHR1upJKF3eTzdg15V8s8a4C3.jpg						
9	2/10/2022	Kimi	A tech woi	1601.782	206	6.3	en	Thriller	https://image.tmdb.org/t/p/original/okNgwtxIWzGsNIR3GsOS0i0Qgbn.jpg						
10	2/17/2022	Fistful of Vengeance	A revenge	1594.013	114	5.3	en	Action, Cri	https://image.tmdb.org/t/p/original/3cccEF9QZgV9bLWwupJO41HSrOV.jpg						
11	11/3/2021	Eternals	The Eterns	1537.406	4726	7.2	en	Science Fic	https://image.tmdb.org/t/p/original/zByhtBvX99ZiCQhac1sh9d9r6nb.jpg						
12	2/18/2022	Pursuit	Detective	1500.523	16	5.9	en	Action, Cri	https://image.tmdb.org/t/p/original/wYlhSXWYqN8Ejsdut2P1P0c97N0.jpg						
13	8/6/2021	My Hero Academia: World H	A mysterie	1485.064	100	7.3	ja	Animation	https://image.tmdb.org/t/p/original/4NUzcKtYPKfTwKslJwNt8nRlXV.jpg						
14	2/25/2022	Restless	After going	1468.377	107	5.9	fr	Action, Thi	https://image.tmdb.org/t/p/original/aw4GGsRwhQyLsjzC7dsAahfCDY.jpg						
15	12/2/2021	Nightmare Alley	An ambitio	1455.144	952	7.1	en	Crime, Dra	https://image.tmdb.org/t/p/original/680klE0dlreQQOyWkFgNnCAJtw.jpg						
16	1/28/2022	The Ice Age Adventures of B	The fearle:	1431.307	737	7.1	en	Animation	https://image.tmdb.org/t/p/original/zzXFM4FKDG71ufrAkwQYv2xvnh.jpg						
17	2/25/2022	Hotel Transylvania: Transfor	When Van	1373.778	288	7	en	Animation	https://image.tmdb.org/t/p/original/teCy1egGQa0y8ULVlrDHQKnxBL.jpg						
18	2/18/2022	Texas Chainsaw Massacre	In this seq	1312.79	521	5.1	en	Horror	https://image.tmdb.org/t/p/original/meRIRfADEGV065xgP06eZVJ0CRS.jpg						
19	1/28/2022	The Requin	A couple o	1252.317	65	4.6	en	Thriller	https://image.tmdb.org/t/p/original/i0z8g2VRZP3dhVwv5MllbOZMKqR.jpg						
20	2/4/2022	Looop Lapeta	When her	1240.946	31	6	hi	Action, Co	https://image.tmdb.org/t/p/original/onGdTS8yYI89drvSjyElnf97rOq.jpg						
21	11/4/2021	Red Notice	An Interpo	1178.544	3193	6.8	en	Action, Co	https://image.tmdb.org/t/p/original/fAXONugq41NwUMuzMiFvicDET9Y.jpg						
22	12/1/2021	Sing 2	Buster anc	1112.9	2172	8.2	en	Animation	https://image.tmdb.org/t/p/original/aWwEkiTRFbbwY8txG5uJC4rMCFSP.jpg						
23	1/3/2022	The Jack in the Box: Awakeni	When a vir	1096.79	43	5.4	en	Horror	https://image.tmdb.org/t/p/original/3lb8vlWTrAKRrTWUrTrZPOMW4jp.jpg						
24	9/30/2021	Venom: Let There Be Carnag	After findi	1053.615	6539	7.1	en	Science Fic	https://image.tmdb.org/t/p/original/rjkmN1dnlUHVYAtwvV3Tj7FsDO.jpg						
25	12/16/2021	The Matrix Resurrections	Plagued by	941.024	3160	6.8	en	Science Fic	https://image.tmdb.org/t/p/original/8c4a8kE7PzaGQqndtMm11xbRp.jpg						
26	11/24/2021	Resident Evil: Welcome to R	Once the t	881.086	1312	6.1	en	Horror, Ac	https://image.tmdb.org/t/p/original/7uRbW0XxpWDMtnsd2PF3clu65jc.jpg						
27	10/19/2021	Last Man Down	After civili	869.605	318	6.3	en	Action, Thi	https://image.tmdb.org/t/p/original/4B7liCxCNZIGONmAMKcNcxVIZQV.jpg						
28	1/7/2022	American Siege	An ex-NYP	867.326	49	5.4	en	Action, Thi	https://image.tmdb.org/t/p/original/daeVrgyjd0ue8qb3AHyU3UeCwoZz.jpg						

Figure 3 - Dataset [1]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage	director	tagline	keywords	overview	runtime	genres	production	release_date	vote_count	average	release_year	budget_ad	revenue_ad
2	135397	tt0369610	32.98576	1.5E+08	1.51E+09	Jurassic W	Chris Pratt	http://www.colin-trew.com	Colin Trew	The park is monster	Jurassic	Twenty-tw	124	Action	Ad Universal	6/9/2015	5562	6.5	2015	1.38E+08	1.39E+09
3	76341	tt1392190	28.41994	1.5E+08	3.78E+08	Mad Max: Tom Hardy	http://www.george-mil.com	George Mil	What a Lo future	che An apocal			120	Action	Ad Village Ro	#####	6185	7.1	2015	1.38E+08	3.48E+08
4	262500	tt2909446	13.11251	1.1E+08	2.95E+08	Insurgent	Shailene W	http://www.robert-sch-one.com	Robert Sch One	Choi: based on r	Beatrice P		119	Adventure	Summit En	#####	2480	6.3	2015	1.01E+08	2.72E+08
5	140607	tt2488496	11.1731	2E+08	2.07E+09	Star Wars: Harrison F	http://www.j.j.abram.com	J.J. Abram	Every gene	android	sg Thirty year		136	Action	Ad Lucasfilm	#####	5292	7.5	2015	1.84E+08	1.9E+09
6	168259	tt2820852	9.335014	1.9E+08	1.51E+09	Furious 7	Vin Diesel	http://www.james-wai.com	James Wai	Vengeance car race	sg Deckard St		137	Action	Cr Universal I	4/1/2015	2947	7.3	2015	1.75E+08	1.63E+09
7	281957	tt1663202	9.1107	1.35E+08	5.33E+08	The Reven	Leonardo	http://www.alejandrorodriguez.com	Al Alejandro	(In. One wh father-son in the 182)			156	Western	Cr Regency Ei	#####	3929	7.2	2015	1.24E+08	4.9E+08
8	87101	tt1340138	8.654359	1.55E+08	4.41E+08	Terminato	Arnold Sch	http://www.alan-taylor.com	Alan Taylo	Reset the I saving	the The year is		125	Science Fic	Paramoun	#####	2598	5.8	2015	1.43E+08	4.05E+08
9	286217	tt3659388	7.6674	1.08E+08	5.95E+08	The Martie	Matt Dam	http://www.ridley-scott.com	Ridley Sco	Bring Him based on r	During a m		141	Drama	Ad Twentieth	#####	4572	7.6	2015	99359956	5.48E+08
10	211672	tt2293640	7.404165	74000000	1.16E+09	Minions	Sandra Bul	http://www.kyle-balda.com	Kyle Balda	Before Gr: assistant	sg Minions St		91	Family	Ani Universal I	#####	2893	6.5	2015	68079970	1.06E+09
11	150540	tt2096673	6.326804	1.75E+08	8.54E+08	Inside Out	Amey Poehl	http://mon-pete-doct.com	Pete Doct	Meet the I dream	car Growing u		94	Comedy	Ad Walt Dsne	6/9/2015	3935	8	2015	1.61E+08	7.85E+08
12	206647	tt2379713	6.200282	2.45E+08	8.81E+08	Spectre	Daniel Crai	http://www.sam-mend.com	Sam Mend	A Plan No	spy based A cryptic n		148	Action	Ad Columbia I	#####	3254	6.2	2015	2.25E+08	8.1E+08
13	76757	tt1617661	6.189369	1.76E+08	1.84E+08	Jupiter Asc	Mila Kunis	http://www.lana-wad.com	Lana Wad	Expand yo Jupiter	aps In a univer		124	Science Fic	Village Ro	2/4/2015	1937	5.2	2015	1.62E+08	1.69E+08
14	264660	tt0470752	6.118947	15000000	36869414	Ex Machin	Domhnaill	http://www.alex-garlat.com	Alex Garlat	There is nc dancing	ai Caleb, a 2f		108	Drama	Sci DNA Films	#####	2854	7.6	2015	13799994	33919846
15	257344	tt2120120	5.984995	88000000	2.44E+08	Pixels	Adam San	http://www.chris-colar.com	Chris Colar	Game On. video gam	Video gam		105	Action	Co Columbia I	#####	1575	5.8	2015	80959964	2.24E+08
16	99861	tt2395427	5.944927	2.8E+08	1.41E+09	Avengers: Robert Do	http://mar-joss-whed.com	Joss Whed	A New Agem	marvel cor When Ton		141	Action	Ad Marvel Stu	#####	4304	7.4	2015	2.58E+08	1.29E+09	
17	273248	tt3460252	5.8984	44000000	1.56E+08	The Hatef	Samuel L.	http://thel-quentin.com	Te No one co bounty hui	Bounty hui		167	Crime	Dra Double Fe	#####	2389	7.4	2015	40479982	1.43E+08	
18	260346	tt2446042	5.749758	48000000	3.26E+08	Taken 3	Liam Nees	http://www.olivier-mei.com	Olivier Mei	It Ends Her revenge	In Ex-govern		109	Crime	Act Twentieth	1/1/2015	1578	6.1	2015	44159981	3E+08
19	102899	tt0478970	5.573184	1.3E+08	5.19E+08	Ant-Man	Paul Rudd	http://mar-peyton.com	Rei Heroes Do	marvel cor Armed wit		115	Science Fic	Marvel Stu	#####	3779	7	2015	1.2E+08	4.77E+08	
20	150689	tt1661199	5.556818	95000000	5.42E+08	Cinderella	Lily James	http://www.cate-blanc.com	Kenneth B	Midnight is cinderella	When her		112	Romance	Ad Walt Dsne	#####	1495	6.8	2015	87399962	4.99E+08
21	131634	tt1951266	5.476958	1.6E+08	6.51E+08	The Hunge	Jennifer La	http://www.francis-lax.com	Francis Lax	The fire w revolution	With the n		136	War	Adve Studio Bab	#####	2380	6.5	2015	1.47E+08	5.98E+08
22	158852	tt1964418	5.462138	1.9E+08	2.09E+08	Tomorrow	Britt Robei	http://mon-brad-bird.com	Brad Bird	Imagine a inventor	le Bound by t		130	Action	Far Walt Dsne	#####	1899	6.2	2015	1.75E+08	1.92E+08
23	307081	tt1798694	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal	http://rach-antoine.com	Rach Antoine F	Believe in sport	Billy "The c		123	Action	Dri Escape Art	#####	1386	7.3	2015	27599988	94373004
24	254128	tt2126355	4.907832	1.1E+08	4.7E+08	San Andre	Dwayne Jc	http://www.brad-ptyc.com	Brad Ptyc	A Rescue p californi	In the afte		114	Action	Dri New Line	#####	2060	6.1	2015	1.01E+08	4.33E+08
25	216015	tt2322441	4.710402	40000000	5.7E+08	Rifty Shad	Dakota Jol	https://www.seth-mac-f.com	Seth Mac F	Ted is Corr sperm ban	Newlywed		125	Drama	Ro Focus Feat	#####	1865	5.3	2015	36799984	5.24E+08
26	318846	tt1596363	6.448046	28000000	1.33E+08	The Big Sh	Christian E	http://www.adam-mck.com	Adam Mck	The men w		130	Comedy	Cr Paramoun	#####	1545	7.3	2015	25759989	1.23E+08	
27	177677	tt2381249	4.566713	1.5E+08	6.82E+08	Mission: Ir	Tom Cruis	http://www.christophe-desperate.com	Christophe Desperate	spy seque Ethan and		131	Action	Paramoun	#####	2349	7.1	2015	1.38E+08	6.28E+08	
28	214756	tt2637276	4.564549	68000000	2.16E+08	Ted 2	Mark Wahlberg	http://seth-mac-f.com	Seth Mac F	Ted is Corr sperm ban	Newlywed		115	Comedy	Universal I	#####	1666	6.3	2015	62559972	1.99E+08

Figure 4 - Dataset [2]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Title	Year	Summary	Short Sum	IMDB ID	Runtime	YouTube T Rating	Movie Pos	Director	Writers	Cast					
2	Patton Oswalt: Annihilation	2017	Patton Os Patton Os	tt7026230	66	4hZi5QaM	7.4	https://hyi.com	Bobcat Go	Patton Os Patton Oswalt						
3	New York Doll	2005	A recoveri A recoveri	tt0436629	75	hwD04Nsn	7.9	https://hyi.com	Greg Whit	Arthur Kan Sylvain Sylvain						
4	Mickey's Magical Christmas:	2001	After even Mickey an	tt0300195	65	uCKwHHft	6.8	https://hyi.com	Tony Craig	Thomas H: Carlos Alazraqui	Wayne Allwine					
5	Mickey's House of Villains	2001	The villain: The villain	tt0329374	0	JA03ciYt-E	6.6	https://hyi.com	Jamie Mit	Thomas H: Tony Anselmo	Wayne Allwine					
6	And Then I Go	2017	In the crue In the crue	tt2018111	99	8CdiD6-iF	7.6	https://hyi.com	Vincent Gr	Brett Hale Arman Darbo	Sawyer Barth					
7	An Extremely Goofy Movie	2000	It's a big ti Max goes	tt0208185	79	H8oSvldAG	6.4	https://hyi.com	Douglas M	Scott Spen Bill Farmer	Jason Marsden	Jeff Bennett				
8	Peter Rabbit	2018	Based on t Feature ac	tt5117670	95	7Pa_Weid	6.6	https://hyi.com	Will Gluck	Rob Lieber Fayssal Bazzi	James Corden					
9	Love Songs	2007	Julie's boy: A musical	tt0996605	91	s54vpKAfR	7.2	https://hyi.com	Christophe	Christophe Clotilde Hesme	Louis Garrel	Ludvine Sagnier				
10		89	2017 89 tells the The true st	tt7614404	91	5hFAExhHf	8.1	https://hyi.com	Dave Stew	Lee Dixon Ian Wright						
11	The Foster Boy	2011	The illegiti A Swiss-Ge	tt2057931	107	E9Qv_XVJ-	7.4	https://hyi.com	Markus Im	Plinio Bach Katja Riemann	Maximilian Simonischek	Stefan Kurt				
12	Forever My Girl	2018	After being After being	tt4103724	108	3vqcMr1q	6.4	https://hyi.com	Bethany A	Bethany A Abby Ryder Fortson	Alex Roe	Jessica Rothe				
13	Tom Segura: Disgraceful	2018	Comedian Comedian	tt7379330	0	kYYINIM3l	7.5	https://hyi.com	Jay Karas	Tom Segur Tom Segura						
14	The Secret Rules of Modern	2015	Professor Professor	tt5818010	58	hpyxmT6K	7.5	https://hyi.com	David Brigg	Marcus du Sautoy						
15	Secrets in the Fall	2015	Nothing is Nothing is	tt522908	0	eMVw1pH	7.3	https://hyi.com	Brittany G	Brittany G Hollie Shay	Kyle Kupecky	Vincent Seidle				
16	Silent Night	2017	Adam une: Adam une:	tt7133554	100	ca6BUYvK	7.5	https://hyi.com	Piotr Dom	Piotr Dom Agnieszka Suchora	Dawid Ogrodnik	Tomasz Zietek				
17	Suicide Squad: Hell to Pay	2018	Task Force Task Force	tt7167602	86	EPZVzk-wb	7.2	https://hyi.com	Sam Liu	Alan Burne Christian Slater	Vanessa Williams					
18	Wildling	2018	Anna spen A blossom	tt5085924	92	eyl1Wf90t	6.1	https://hyi.com	Fritz BÄ	h Fritz BÄ	h Bel Powley	Brad Dourif	Liv Tyler			
19	The Humanity Bureau	2017	A dystopia A dystopia	tt6143568	95	kUH8lGHr	6.1	https://hyi.com	Rob W. Kir	Dave Schu Jakob Davies	Nicolas Cage	Sarah Lind				
20	Farewell Ferris Wheel	2016	Farewell, F Farewell F	tt2349677	72	#NAME?	7.4	https://hyi.com	Miguel 'M. Miguel 'M.I.G.'	Martinez						
21	Don't Talk to Irene	2017	Irene must Irene must	tt5030452	90	IK48UxWu	6.4	https://hyi.com	Pat Mills	Pat Mills Anastasia Phillips	Bruce Gray	Michelle McLeod				
22	Blood Road	2017	The myste The myste	tt6900092	96	XCzVqIN95	6.9	https://hyi.com	Nicholas S	Mark Ande Huyen Nguyen	Ian Adamson	Rebecca Rusch				
23	Andre the Giant	2018	A look at t A look at t	tt6543420	85	f_JTeuajias	8.2	https://hyi.com	Jason Hehi	Robin Wri Cary Elwes						
24	Dead on Arrival	2017	A pharmac A pharmac	tt5903088	97	8Jl-kC-rq0t	6.9	https://hyi.com	Stephen C	Stephen C Christa B. Allen	D.B. Sweeney	Scottie Thompson				
25	Big Time	2017	Bjarke Ingx BIG TIME	tt7630164	93	VK0mGdM	6.2	https://hyi.com	Kaspar Ast	Bjarke Ingx Elisabet Ingels						
26	Adventures in Babysitting	2016	Two teen i Two teen i	tt4456850	105	BbTXXUM	6.1	https://hyi.com	John Schul	Tiffany Pai Nikki Hahn	Sabrina Carpenter	Sofia Carson				
27	Banana in a Nutshell	2005	An intimat An intimat	tt1537196	0	6vgD7826u	6.6	https://hyi.com	Roseanne	Roseanne Liang						
28	Hostiles	2017	In 1892. af In 1892. a	tt5478478	134	1M5ci4Un	7.3	https://hyi.com	Scott Cook	Scott Cook Ava Cooer	Rosamund Pike	Scott Sheehard				

Figure 5 - Dataset [3]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Release_Date	Movies' Title	Movies' Plot	Movies' Popularity	Vote_Count	Vote_Average	Language	Genre	Poster_Url	match_imdb_id	match_cast	match_director	match_runtime	merge_summary	
2	5/13/2015	Mad Max: Fur	An apocalyptic story set in the	89,139	19090	7.5	en	Action, /	https://ima	tt1392190	Tom Hardy C George Miller		120	In a post-apocalyptic waste	
3	3/18/2015	Insurgent	Beatrice Prior must confront	78,099	8780	6.4	en	Action, /	https://ima	tt2908446	Shailene Woo Robert Schwentke		119	Beatrice Prior must confro	
4	12/15/2015	Star Wars: Th	Thirty years after defeating th	77,811	16772	7.3	en	Action, /	https://ima	tt2488496	Harrison Ford J.J. Abrams		136	Three decades after the En	
5	4/1/2015	Furious 7	Deckard Shaw seeks revenge	105,485	8973	7.3	en	Action, /	https://ima	tt2820852	Vin Diesel Pa James Wan		137	Deckard Shaw seeks reven	
6	12/25/2015	The Revenant	In the 1820s, a frontiersman,	53,111	15554	7.5	en	Westerr	https://ima	tt1663202	Leonardo DiC Alejandro Gonz		156	A frontiersman on a fur tra	
7	6/23/2015	Terminator G	The year is 2029. John Conno	108,488	7089	5.9	en	Science	https://ima	tt1340138	Arnold Schwar Alan Taylor		125	When John Connor, leader	
8	9/30/2015	The Martian	During a manned mission to M	68,531	16724	7.7	en	Drama, /	https://ima	tt3659388	Matt Damon Ridley Scott		141	An astronaut becomes stra	
9	6/17/2015	Minions	Minions Stuart, Kevin and Bol	45,208	9143	6.4	en	Family, /	https://ima	tt2293640	Sandra Bulloc Kyle Balda Pierre		91	Minions Stuart, Kevin, and	
10	6/9/2015	Inside Out	Growing up can be a bumpy r	108,817	17586	7.9	en	Animati	https://ima	tt2096673	Amy Poehler Pete Doctor		94	After young Riley is uproot	
11	10/26/2015	Spectre	A cryptic message from Bond	58,265	8944	6.5	en	Action, /	https://ima	tt2379713	Daniel Craig C Sam Mendes		148	A cryptic message from 00	
12	2/4/2015	Jupiter Ascen	In a universe where human ge	64,915	5811	5.4	en	Science	https://ima	tt1617661	Milla Kunis Ch Lana Wachowski		124	A young woman discovers	
13	12/16/2014	Ex Machina	Caleb, a coder at the world's l	27,676	11034	7.6	en	Drama, /	https://ima	tt0470752	Domhnall Gle Alex Garland		108	A young programmer is sel	
14	7/16/2015	Pixels	Video game experts are recru	79,688	6366	5.7	en	Action, /	https://ima	tt2120120	Adam Sandler Chris Columbus		105	When aliens misinterpret v	
15	4/22/2015	Avengers: Age	When Tony Stark tries to jum	163,611	19473	7.3	en	Action, /	https://ima	tt2395427	Robert Down Joss Whedon		141	When Tony Stark and Bruc	
16	12/25/2015	The Hateful E	Bounty hunters seek shelter fr	32,716	11871	7.7	en	Drama, /	https://ima	tt3460252	Samuel L. Jaci Quentin Tarantino		167	In the dead of a Wyoming	
17	12/16/2014	Taken 3	Ex-government operative Bry	119,377	4792	6.2	en	Thriller, /	https://ima	tt0446042	Liam Neeson Olivier Megaton		109	Ex-government operative B	
18	7/14/2015	Ant-Man	Armed with the astonishing a	85,255	16802	7.1	en	Science	https://ima	tt0478970	Paul Rudd M Peyton Reed		115	Armed with a super-suit wi	
19	11/18/2015	The Hunger G	With the nation of Panem in a	75,942	10412	6.9	en	Action, /	https://ima	tt1951266	Jennifer Lawri Francis Lawrence		136	As the war of Panem escal	
20	5/19/2015	Tomorrowlan	Bound by a shared destiny, a	47,761	6056	6.2	en	Advent	https://ima	tt1964418	Britt Roberts Brad Bird		130	Bound by a shared destiny,	
21	3/24/2015	Southpaw	Billy "The Great" Hope, the re	101,201	4644	7.4	en	Action, /	https://ima	tt1798684	Jake Gyllenha Antoine Fuqua		123	A champion boxer fights to	
22	5/27/2015	San Andreas	In the aftermath of a massive	101,043	7077	6.2	en	Action, /	https://ima	tt2126355	Dwayne John Brad Peyton		114	In the aftermath of a mass	
23	2/11/2015	Fifty Shades o	When college senior Anastasi	150,025	10087	5.9	en	Drama, /	https://ima	tt2322441	Dakota Johns Sam Taylor-Johns		125	Literature student Anast	
24	12/11/2015	The Big Short	The men who made millions f	29,765	7289	7.3	en	Comedy	https://ima	tt1596363	Christian Bale Adam McKay		130	In 2006-7 a group of invest	
25	7/23/2015	Mission: Impc	Ethan and team take on their	48,086	7178	7.2	en	Action, /	https://ima	tt2381249	Tom Cruise Jc Christopher McQu		131	Ethan and team take on th	

Figure 6 - Dataset [4]

From figure 1, 2, and 3 there are three dataset that our group found on Kaggle. Our group merged three dataset to one dataset which is figure 4. In figure 4, it has around 1,500 data in the dataset.

Tools and software

- **Excel (data)**

Our group uses Excel for collecting data and use in the Elasticsearch for calling data when the search engine is working.

- **Visual Studio Code (HTML and CSS)**

Our group uses HTML and CSS for the interface of our movie search engine. HTML is the standard markup language for Web pages [4]. CSS is used to format the layout of a webpage which our group uses both inner CSS and external CSS [5].

System diagram

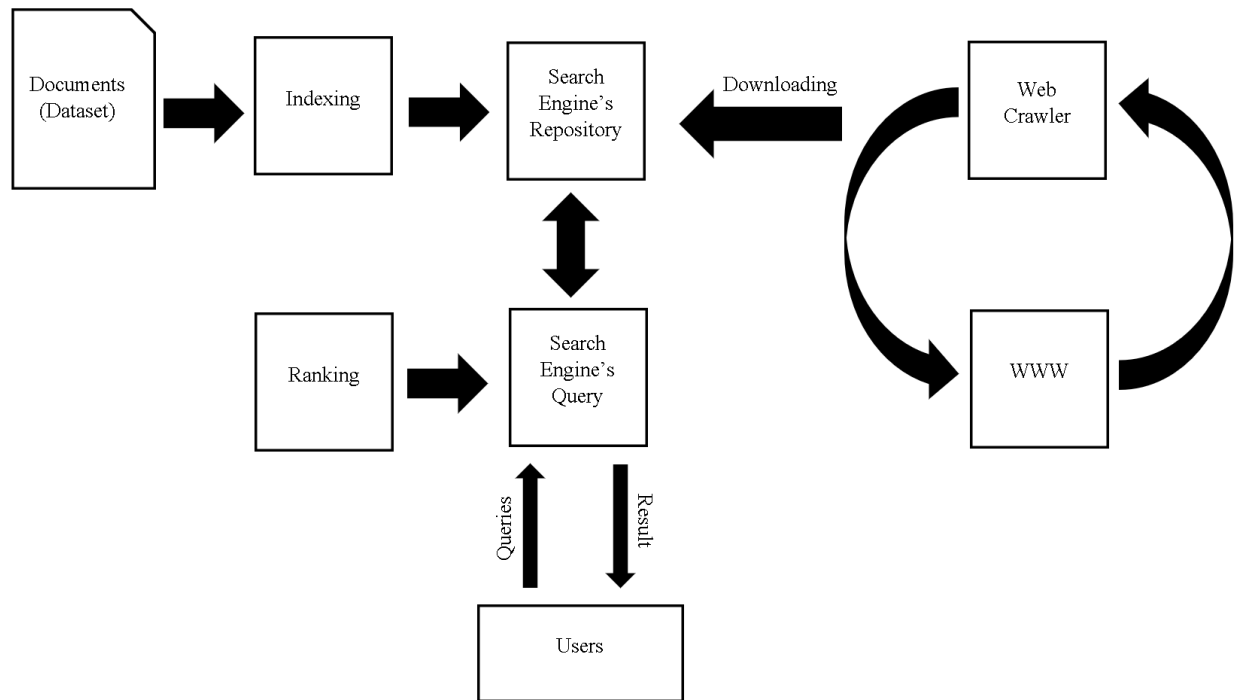


Figure 7 Movie Search System Diagram

Snapshots of the system

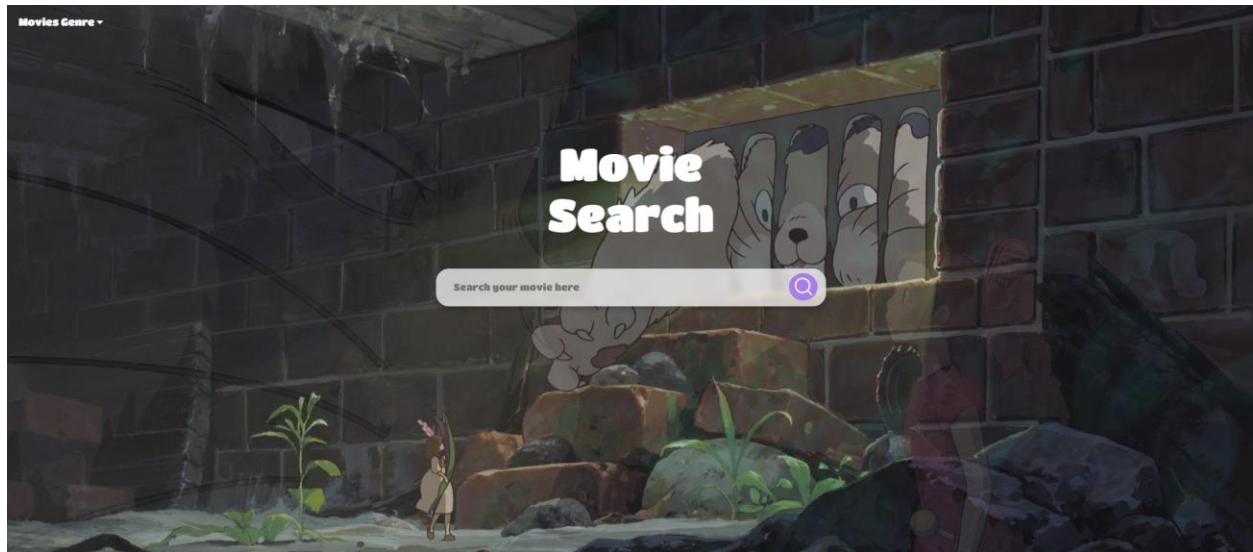


Figure 8 - Movie search page

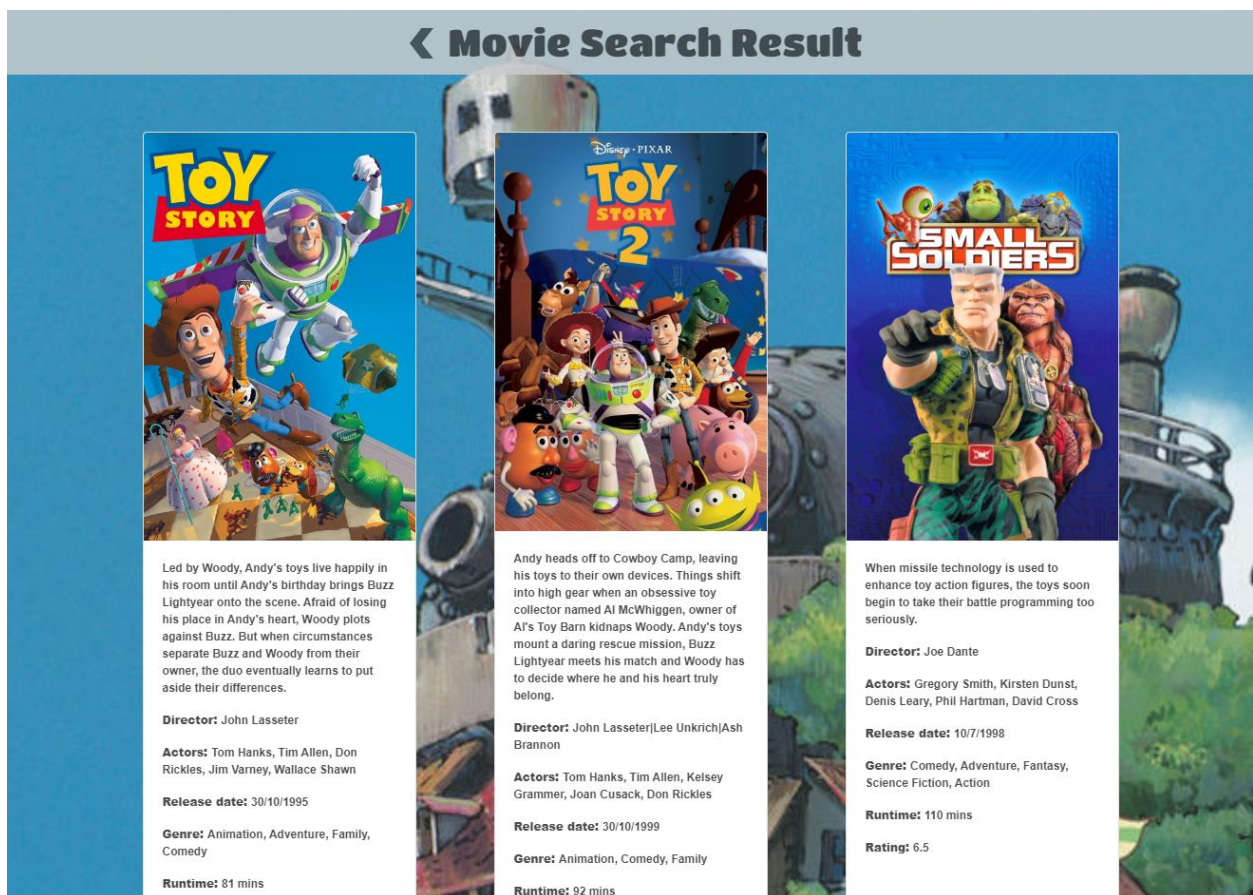


Figure 9 - Search result page (Toy Story)

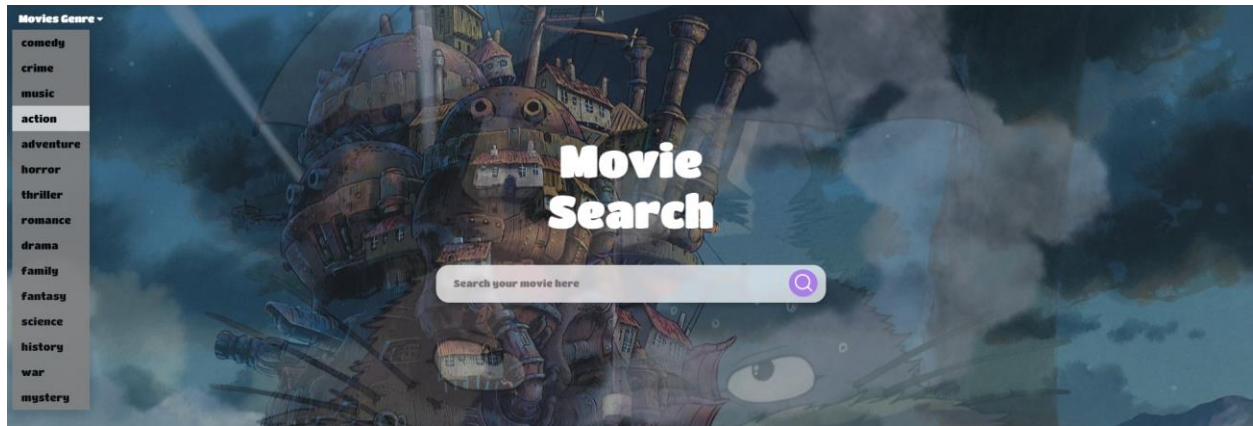


Figure 10 - Movie category selection

Example step-by-step search sessions

1. One word query

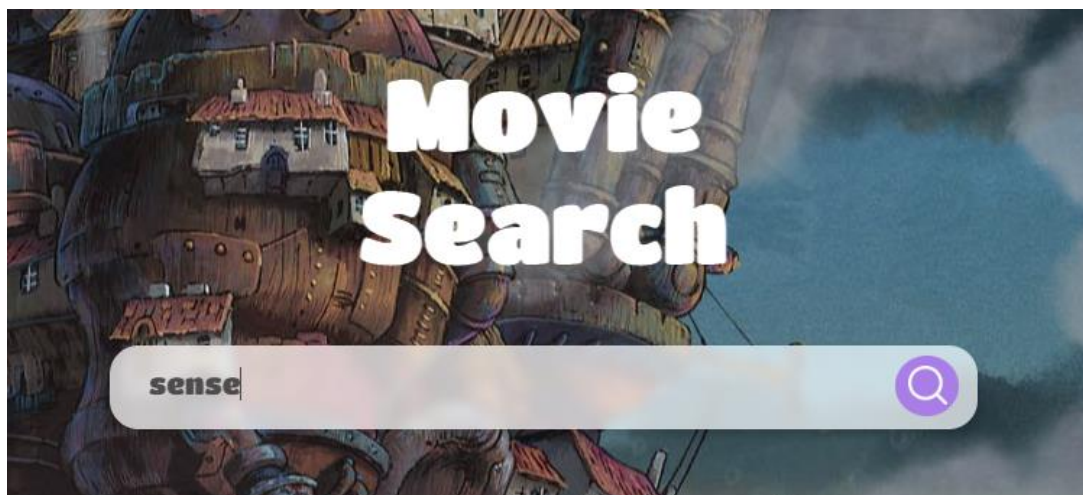


Figure 11 - Step 1, put a keyword (sense) in the search box and press enter



Figure 12 - Step 2, the results of the keyword (sense) has shown on the result page

2. Multiple words query

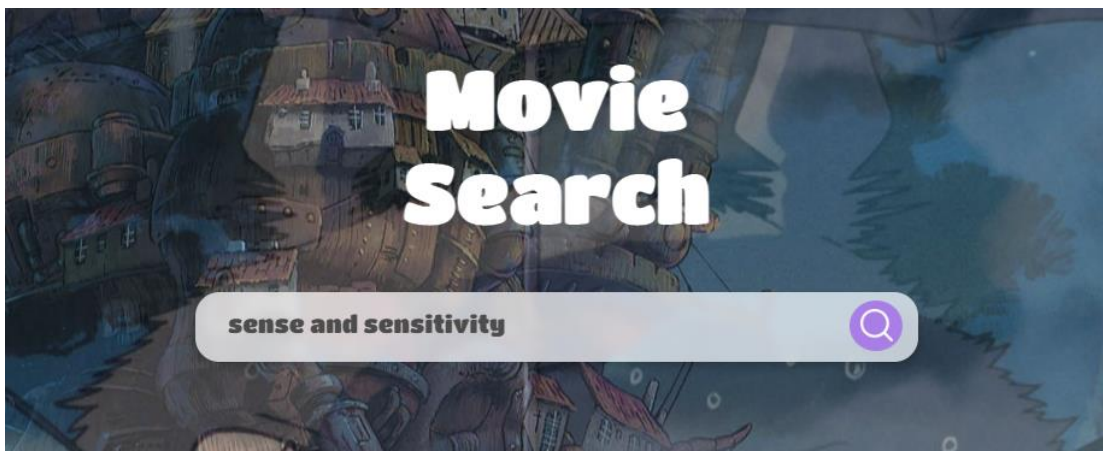


Figure 13 - Step 1, put multiple words (sense and sensitivity) in the search box and press enter

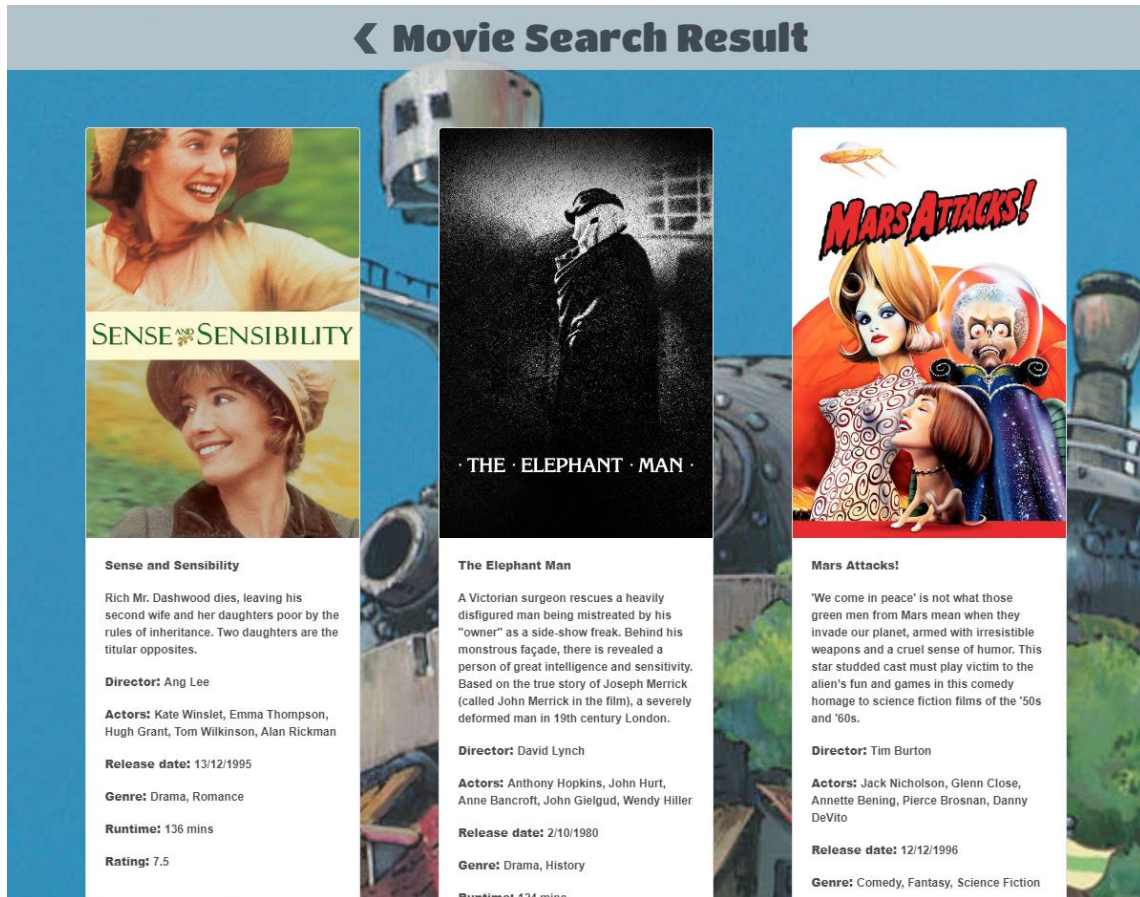


Figure 14 - Step 2, the results of the keyword (sense and sensitivity) has shown on the result page

3. Partial match

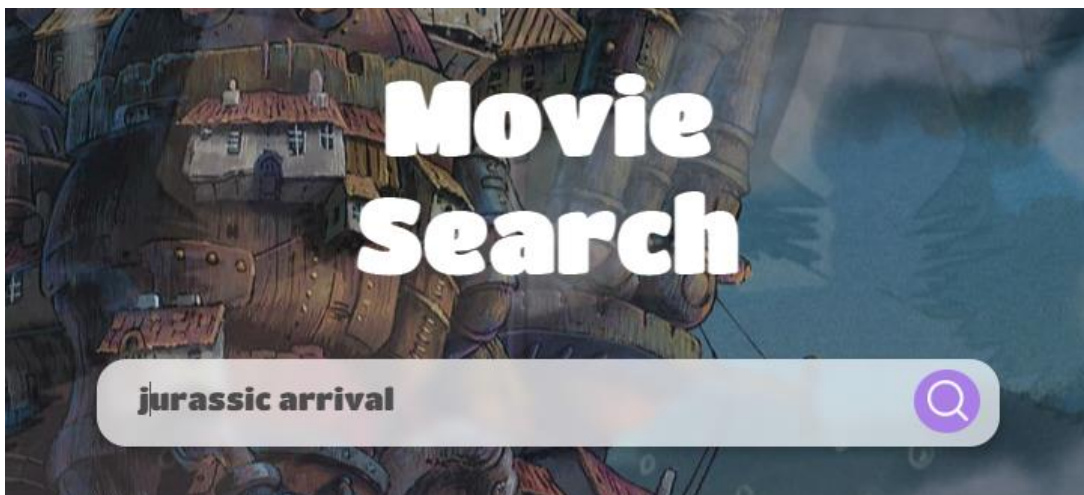


Figure 15 - Step 1, put multiple different words (jurassic arrival) in the search box and press enter

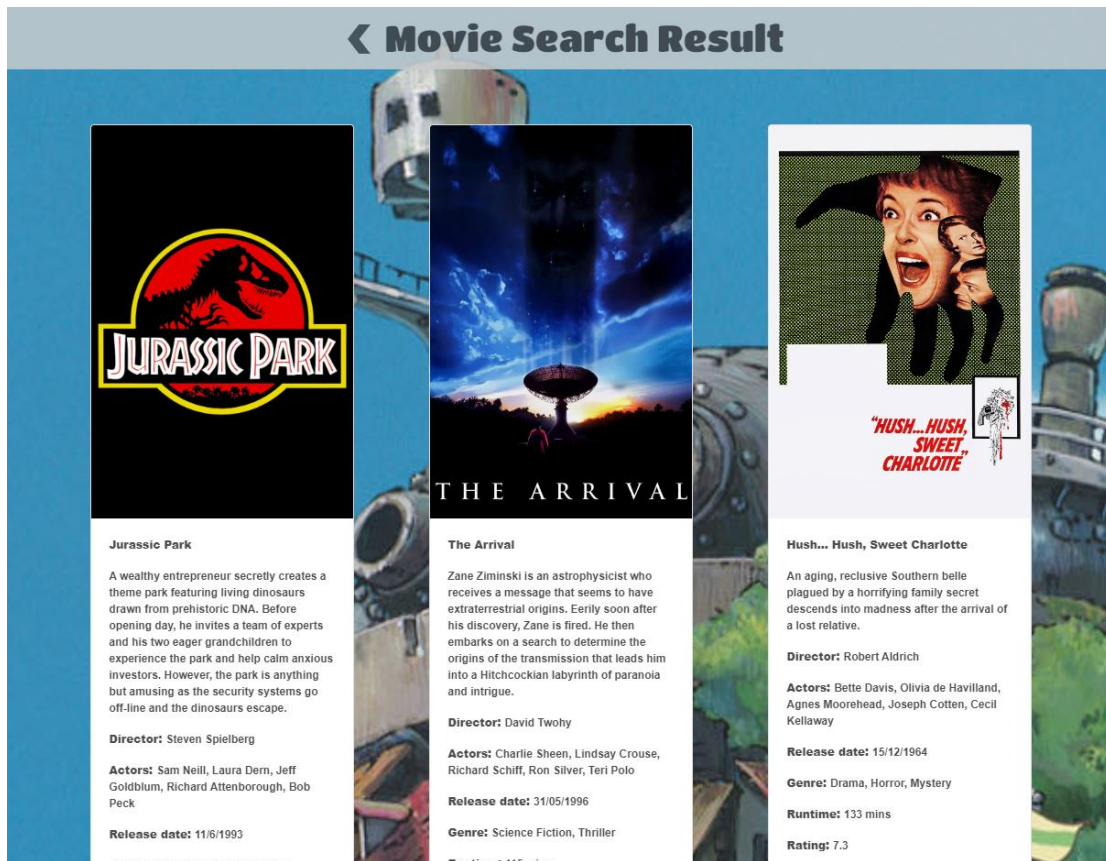
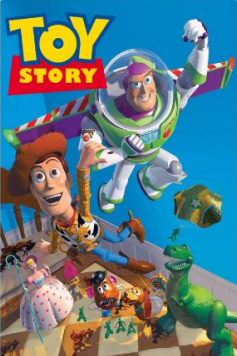


Figure 16 - Step 2, the results of the keyword (jurassic arrival) has shown on the result page

4. Ranking

Movie Search Result



Toy Story


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.

Director: John Lasseter

Actors: Tom Hanks, Tim Allen, Don Rickles, Jim Varney, Wallace Shawn

Release date: 30/10/1995

Genre: Animation, Adventure, Family, Comedy



Toy Story 2

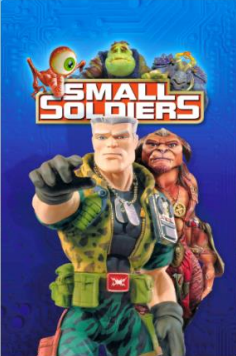
Andy heads off to Cowboy Camp, leaving his **toys** to their own devices. Things shift into high gear when an obsessive **toy** collector named Al McWhiggen, owner of Al's **Toy** Barn kidnaps Woody. Andy's **toys** mount a daring rescue mission, Buzz Lightyear meets his match and Woody has to decide where he and his heart truly belong.

Director: John Lasseter, Lee Unkrich, Ash Brannon

Actors: Tom Hanks, Tim Allen, Kelsey Grammer, Joan Cusack, Don Rickles

Release date: 30/10/1999

Genre: Animation, Comedy, Family



Small Soldiers

When missile technology is used to enhance **toy** action figures, the **toys** soon begin to take their battle programming too seriously.

Director: Joe Dante

Actors: Gregory Smith, Kirsten Dunst, Denis Leary, Phil Hartman, David Cross

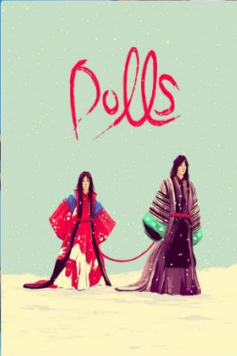
Release date: 10/7/1998

Genre: Comedy, Adventure, Fantasy, Science Fiction, Action

Runtime: 110 mins

Rating: 6.5

Figure 17 - the results of searching (toy story) with highlight the query word



Dolls

A group of people stop by a mansion during a storm and discover two magical **toy** makers, and their haunted collection of dolls.

Director: Stuart Gordon

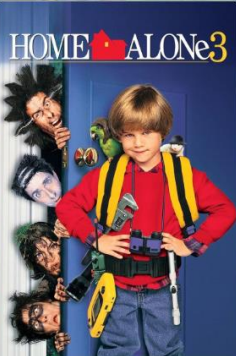
Actors: Ian Patrick Williams, Carolyn Purdy-Gordon, Carrie Lorraine, Guy Rolfe, Bunty Bailey

Release date: 5/9/2002

Genre: Drama, Romance

Runtime: 77 mins

Rating: 7.6



Home Alone 3

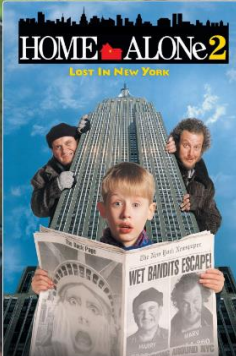
9-year-old Alex Pruitt is home alone with the chicken pox. Turns out, due to a mix-up among nefarious spies, Alex was given a **toy** car concealing a top-secret microchip. Now Alex must fend off the spies as they try break into his house to get it back.

Director: Raja Gosnell

Actors: Alex D. Linz, Olek Krupa, Rya Kihlstedt, Lenny Von Dohlen, Scarlett Johansson

Release date: 12/12/1997

Genre: Comedy, Family



Home Alone 2: Lost in New York

Instead of flying to Florida with his folks, Kevin ends up alone in New York, where he gets a hotel room with his dad's credit card—despite problems from a clerk and meddling bellboy. But when Kevin runs into his old nemeses, the Wet Bandits, he's determined to foil their plans to rob a **toy** store on Christmas eve.

Director: Chris Columbus

Actors: Macaulay Culkin, Joe Pesci, Catherine O'Hara, Daniel Stern, John Heard

Release date: 19/11/1992

Genre: Comedy, Family, Adventure

Figure 18 - the results of searching (toy story) with highlight the query word (continue)

Discussion

Limitations of your system

Since our movie search engine is the very first demo version, the result of the search is not efficient enough. Our search engine only looks up the title, overview, director, and cast fields of the dataset (also the movie's genre but we make it like a filter search, not a query search). The efficiency of the search is not good enough because we did not optimize the search threshold. We think we can improve search performance better than this in a later version. The last thing is the search engine does not provide spelling corrections. the user must type every word carefully.

Technical difficulties, challenges, and lessons learned

There are technical difficulties that our group met including:

- **Interface**

For the interface, it is difficult to create a web page by using HTML and CSS to have the same interface. When we open it on another laptop, it is always different. So, we have to create a web page to look the same as much as we can.

- **Server Downloading**

For the server downloading, it is really slow to download the image. You have to wait for the image download which takes a while. Also, background images of the website are not continuing from the previous image.

Our group learned many things from creating movie search engines. These are what we learn from create movie search engine:

- Learning how to prepare data by obtaining data from various sources and combining this together before we can process and retrieve our final dataset.
- To ensure that our dataset has high quality and data integrity, we learned how to clean and eliminate duplicate data, transform the data type, correct errors, and deal with missing values.
- To be able to become familiar with new tools like Elastic search, which enables us to quickly store, search, and analyze large amounts of data and return valuable results.
- The management of teams is the most significant lesson we have learned. Our team, which is made up of individuals with diverse areas of expertise, has divided the particular tasks

for each member, including front-end, back-end, and database management. As a result, our work was completed in a short period of time.

Opportunities for future improvements

Our group discussed about what we need to improve on our movie search engine. These are what we need to improve or create more functionalities:

- Old movies searching

Most movie search engines don't have many old movies below 2000. Some people like to watch old movies. If possible, it might be good for our movie search engine.

- Movies searching by photo

Some people might have some picture of the movie, but they cannot remember the movie's title. If we can add this feature to our movie search engine, it will be more convenient and a lot easier for finding movie's information.

Conclusion

In this project, we implemented a simple movie search (search engine) that can search a movie by title, storyline, director, and cast. We built the project with many tools: Microsoft Excel - used to handle records in the dataset, Elastic and Kibana - used to build and manipulate the inverted index via a beautiful configuration interface, and Flask - used to design and create the interface of our web-based search engine. In this project, we ran into many unexpected problems that we had to find a way to solve. For instance, the problem in the dataset preparation is that we cannot find the dataset that contains the information we want about movies, so we have to merge many datasets into the perfect one before we go to create the inverted index. Also the problem of building the inverted index, the dataset we have uploaded to Elastic triggers an error about the field determined, then we try to follow the instructions of its and try to customize the format of the time field after that, an error has gone. In this project, we have learned many things which can help us to do a job in the real world such as work breakdown structure, collaboration in the team, and problem-solving skills. Eventually, we accomplish many things more than just a search engine project. Thank you.

References

- [1] Hasan, D.I.R. (2022) *9000+ movies dataset, Kaggle*. Available at:
<https://www.kaggle.com/datasets/240bbdc8f00ac47a8b212c9815991ba2d1d9a382c8bc6a103b906f43e7aba303> (Accessed: November 23, 2022).
- [2] Tanyu, K. (2018) *Movie dataset, Kaggle*. Available at:
<https://www.kaggle.com/datasets/kayscrapes/movie-dataset> (Accessed: November 23, 2022).
- [3] Shakir, J. (2018) *TMDB Movies Dataset, Kaggle*. Available at:
<https://www.kaggle.com/datasets/juzershakir/tmdb-movies-dataset> (Accessed: November 23, 2022).
- [4] Hasan, D.I.R. (2022) *9000+ movies dataset, Kaggle*. Available at:
<https://www.kaggle.com/datasets/240bbdc8f00ac47a8b212c9815991ba2d1d9a382c8bc6a103b906f43e7aba303> (Accessed: November 23, 2022).
- [5] *HTML styles - CSS* (no date) *HTML Styles CSS*. Available at:
https://www.w3schools.com/html/html_css.asp (Accessed: November 23, 2022).