# Assignment 4: Vector_Space Information Retrieval System

**Description:**

Supported by the 2018_movie corpus, the program generate vsm_index (vector space model inverted index) for title and text fields of the movie page, offers disjunctive queries, returns search results with snippets and movie descriptions, ranks the results using cosine_similarity score and display the top k results to the user.

**Dependencies:**

MacOS High Sierra version 10.13.2
Python 3.6
nltk==3.3
Flask==1.0.2

**Run Instructions:**

Execute the following command to generate the vsm_index shelve, len_normalizatio shelve, movie description shelve, and test_corpus.json:

```
python3  vsm_index.py
```

Execute the following command to run the Boolean information retrieval system:

```
python3  vsm_query.py
```

**Modules and methods:**

preprocessing.py

**PreProcessing**

A class contains the methods to preprocess the text loaded from corpus, which is later used for building inverted index.

**__init__(self):**

contains nltk stopwords list and most frequent and unhelpful terms from the text

**flatten(self, x):**

leave 1D list unchanged, strings to a list, multi-D list to 1D list
:param x:
:return: 1D list

**normalize(self, token):**

do case-folding, removing stopwords, and stemming

**test_corpus(self, filename='test_corpus.json')**

Create a test containing 10 hand tailored documents corpus in json file

vsm_index.py

This module 1. Generate corpus shelve. .

**inverted_index(self, index_shelve_name, len_normalization_shelve_name, corpus_name='test_corpus.json'):**
create vsm_index.db with key_value pairs: {'term':[(docID, tf), ...]}
create term_normalization.db with key_value pairs: {'docID': the length of the vector,...}

**corpus_shelve(self, shelvename, corpus_name='test_corpus.json'):**
store the info from the corpus json file to corpus shelve file for easy access

**cosine_score_disjunct(self, query, k, inverted_index='vsm_index.db', doc_normalize ='len_normalization.db'):**
:param query: query term
:param k: top k results that the user wants
:param inverted_index: vsm_index.db
:param doc_normalize: len_normalization.db
:return: a list of tuples (doc's cosine_similarity score, docID), a list of stop words, and a list of unknown words, compute the cosine scores of a disjunctive query for each document and return top k documents

vsm_query.py

**dummy_movie_data(docID, shelvename='corpus_shelve'):**
Return data fields for a movie.

**dummy_movie_snippet(scores_pair):**
:param score_queue returned from cosine_score_disjunct
Return a snippet for the results page.

**query():**
generate the welcome page

**results(page_num):**
Generate a result set for a query and present the 10 results starting with <page_num>

**movie_data(film_id):**
Given the doc_id for a movie, present the title and text and structured fields for the movie

**more_data(page_num):**
use the selected current document's title and text as a query to search for similar documents

**Files in the folder:**

2018_movies.json: the movie corpus that supported the boolean IR system
test_corpus.json: a hand-made ten documents corpus

corpus.db: the movie corpus that is transferred from the json file

vsm_index.db: inverted index with term as key and a list of tuples (‘docID’, tf) and df as value

len_normalization.db: the vector length of each document vector

templates/query_page.html, templates/results_page.html, templates/error_page.html, templates/doc_data_page.html: html files used to create the web page

**Text normalization details:**

nltk.stop_words, most frequent words that potentially appear in every movie and not helpful with queries are removed from the tokens. The tokens are converted into lowercase and stemmed.

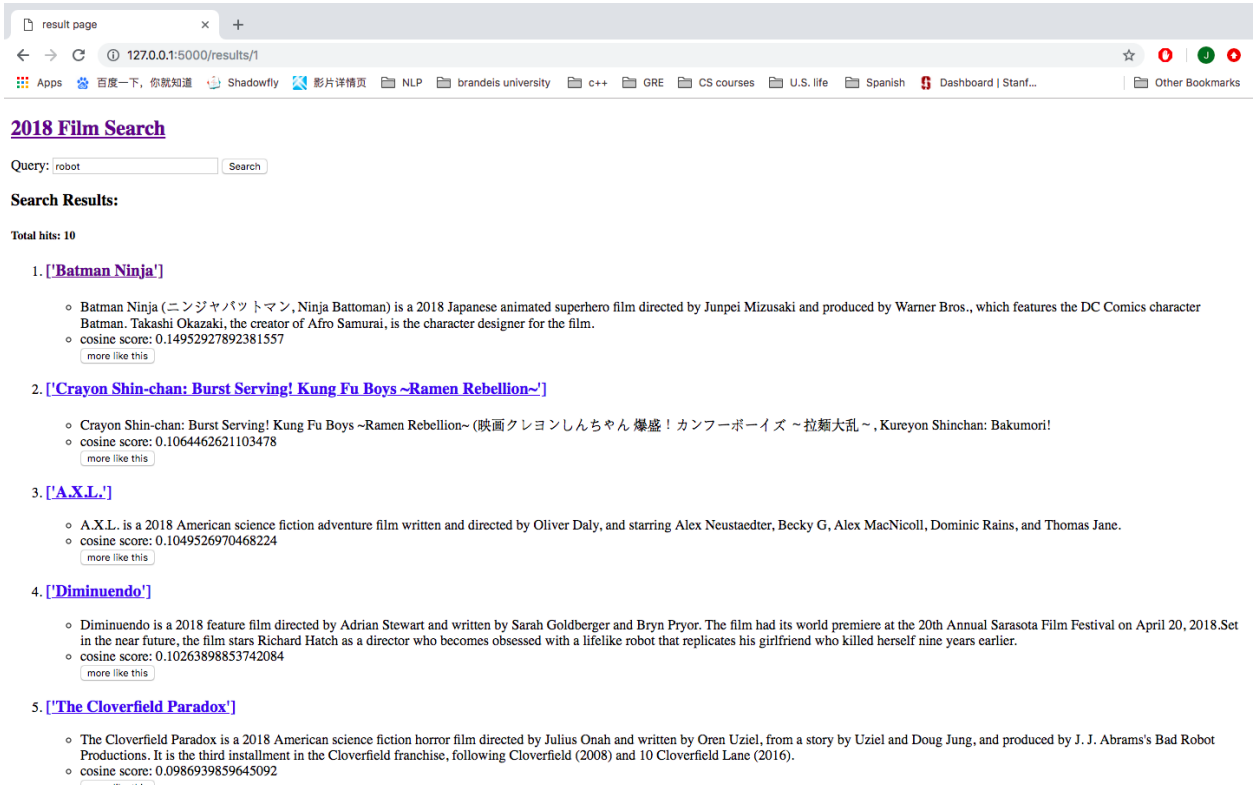
**Testing:**

Utilizing the test_copus created in proprocessing.py, different words with same stems are all matched. Only the matched movie titles and snippets are returned in the search result page.

**Example from 2018_movies:**

Test query: robot [('1', 11), ('34', 2), ('190', 7), ('265', 1), ('296', 2), ('357', 6), ('376', 1), ('402', 1), ('451', 1), ('591', 1), 10]

The first returning result has docID ‘190’

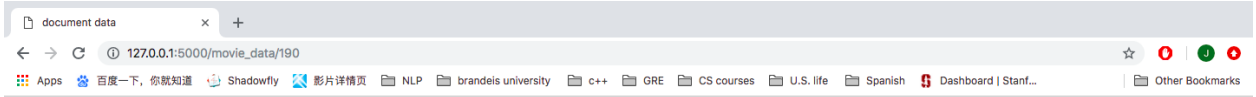



## 2018 Film Search

**Search Results:**

**Total hits: 20**

1. **['Batman Ninja']**
   - Batman Ninja (ニンジャバット トマン, Ninja Battoman) is a 2018 Japanese animated superhero film directed by Junpei Mizusaki and produced by Warner Bros., which features the DC Comics character Batman. Takashi Okazaki, the creator of Afro Samurai, is the character designer for the film.
   - cosine score: 2.3928950150418467
2. **['Commando Ninja']**
   - Commando Ninja is a 2018 English-language French martial arts action comedy film written and directed by Benjamin Combes. It pays homage to 1980s action films such as Commando, The Terminator, Rambo: First Blood Part II, Predator, and American Ninja.
   - cosine score: 0.8593611576272476
3. **['Batman: Gotham by Gaslight']**
   - Batman: Gotham by Gaslight is a 2018 American animated steampunk superhero alternate history action thriller film produced by Warner Bros. Animation and distributed by Warner Bros. Home Entertainment.
   - cosine score: 0.5807069894776745
4. **['The Death of Superman']**
   - The Death of Superman is a 2018 American animated direct-to-video superhero film produced by Warner Bros. Animation and DC Entertainment.
   - cosine score: 0.5773214808562652
5. **['DC Super Hero Girls: Legends of Atlantis']**
   - DC Super Hero Girls: Legends of Atlantis is a 2018 American animated film based on the DC Super Hero Girls franchise, produced by Warner Bros. Animation and distributed by Warner Bros. Home Entertainment.
   - cosine score: 0.404815865810322
6. **['The Crimes That Bind']**
   - The Crimes That Bind (祈りの幕が下りる時, Inori no Maku ga Oriru toki) is a 2018 Japanese film directed by Kyōichirō Kaga based on the novel by Keigo Higashino. Plot The film centers around the discovery of the body of Michiko Oshitani.
   - cosine score: 0.38762349851747013




# *['Batman Ninja']*

---

**Director:** ['Junpei Mizusaki']
**Starring:** Kōichi Yamadera, Wataru Takagi, Ai Kakuma, Rie Kugimiya, Hōchū Ōtsuka
**Location:** ['feudal Japan']

**Text:** Batman Ninja (ニンジャバット トマン, Ninja Battoman) is a 2018 Japanese animated superhero film directed by Junpei Mizusaki and produced by Warner Bros., which features the DC Comics character Batman. Takashi Okazaki, the creator of Afro Samurai, is the character designer for the film. The first poster was revealed on October 5, 2017, and the trailers were released later on December 1, 2017. The film was released in the United States in digital format on April 24, 2018; it was released in physical formats on May 8 and was released theatrically in Japan on June 15. In its American release, writers Leo Chu and Eric Garcia have admitted to rewriting the film from the original Japanese script written by Kazuki Nakashima, ultimately making two entirely different versions of the same film. Plot While battling Gorilla Grodd at Arkham Asylum, Batman is caught in Grodd's Quake Engine time displacement machine and sent to Feudal Japan. There, he is chased by samurai working for the villainous Joker. During his escape, Batman meets up with Catwoman, who reveals everyone else arrived two years earlier (due to Batman being in the outermost area affected by the Quake Engine). He learns from her that all of Gotham City's top criminals have become feudal lords after deceiving the Sengoku daimyō, battling each other until only one state remains. In order to stop the villains from changing history, Batman and Catwoman must get to the Quake Engine in Arkham Castle (formerly the asylum). Batman discovers that Alfred Pennyworth is also in the past and has built a Batcave outside Edo. When the Joker's troops ambush the hideout, Batman storms his way in his Batmobile towards Arkham Castle, which transforms into a giant robot fortress. Just as Batman confronts the Joker, he is forced to leave and save a mother and child below from being crushed by the robot's hand. He transforms his Batcycle into an armored suit to defeat a sumo Bane and stop the robot hand, only for the mother to reveal herself as Harley Quinn and knock him down. As Batman is surrounded by the Joker's minions, he is suddenly whisked away by ninjas led by Eian of the Bat Clan of Hida. He learns that the Bat Clan helped Nightwing, Red Hood, Robin, and Red Robin upon their arrival, and that the clan had followed a prophecy of a foreign bat ninja restoring order to the land. Robin gives Batman an invitation from Grodd to a nearby hot spring. There, Grodd explains that he intended to send the villains far away so he could take Gotham for himself, but Batman's interference sent them all to Feudal Japan instead. Batman and Grodd agree to work together to return to Gotham. Batman, Grodd, Catwoman, the Bat Family, and the Bat Clan battle the Joker and his forces by the river. They defeat the Joker and Harley, but Grodd turns on Batman, revealing his alliance with Two-Face before the Joker and Harley escape and blow up their own ship, taking Batman down with it. Having captured a power converter from Harley, Catwoman attempts to bargain with Grodd in bringing her back to Gotham; however, they need to obtain other power converters from Penguin, Poison Ivy, and Deathstroke to complete the Quake Engine. Two days later, Batman recovers from his wounds and encourages the Bat Family to learn the ways of the ninja in order to defeat Grodd. Red Hood locates the Joker and Harley, but Batman discovers that they lost their memories from the explosion and are living their lives as farmers. A month later, the Gotham villains mobilize their castle robots for battle at Jigokukohara, the field of Hell. Batman leads the Bat Family and the Bat Clan into the battlefield. After defeating the other villains, Grodd puts them under his mind control, with the intent of ruling the country himself. The Joker and Harley, however, crash his party from above, reclaiming their castle from Grodd. The Bat Family saves Catwoman and Grodd before the Joker merges all of the castles into the super robot Lord Joker. An injured Grodd gives Batman control of his army of monkeys; Robin enables them to merge into one giant samurai monkey to battle the Joker's robot. The samurai monkey then combines with a swarm of bats to form the Batgod to defeat Lord Joker before the Bat Family storm into the castle to battle the villains. The Joker reveals to Batman that as farmers, he and Harley planted special flowers that triggered their memories back once they bloomed. As the castle falls, Batman and the Joker engage in a sword fight. Using his ninjutsu skills, Batman defeats the Joker. With the Joker and the Gotham villains defeated, Feudal Japan is restored to its original state and the Bat Family take the villains back to the present day. In a mid-credits sequence, Catwoman sells weapons and furniture from the castle robots to an antique shop while Bruce rides a horse-driven Batmobile to a party hosted by the mayor. Voice cast Marketing Bandai released S.H. Figuarts figures of Ninja Batman and Demon King Joker in mid-2018.The Nendoroid figures of Pop Team Epic characters Popuko and Pipimi, dressed as Batman and Joker respectively, were displayed at the Warner Bros. booth at AnimeJapan 2018. It was suggested by Junpei Mizusaki at Kamikaze Douga; the studio animated both this film and Pop Team Epic television series. The crossover figures were accompanied by a 15-second TV commercial, where Popuko and Pipimi (in the aforementioned costumes) re-enact a sketch from Pop Team Epic comics before it jumps to a Batman Ninja scene. Reception On review aggregator website Rotten Tomatoes, the film received an approval rating of 80% based on 15 reviews, with an average rating of 6.5/10.IGN awarded Batman Ninja a score of 9.7 out of 10, saying, "DC tried something new by bringing in visionary Japanese animators to offer a refreshing take on one of the company's most beloved characters, and the finished product not only built upon the great adaptations that have come before, but surpassed them."The film earned $403,930 from domestic DVD sales and $2,346,251 from domestic Blu-ray sales, bringing its total domestic home video earnings to $2,750,181. References External links Official website (Warner Bros.) Official website (Warner Bros. Japan) (in Japanese) Official website (DC Comics) Batman Ninja (anime) at Anime News Network's encyclopedia Batman Ninja on IMDb

(Use browser "back" button to return to search results.)