Project FilmCatcher

Data Matching: Global Film Titles

Blue Jeans & **Bloody Tears**

OpusData · comscore · usheru · INTERNATIONAL SHOWTIMES API · AccuWeather · ni · gruvi

Identifying the Problem

Understanding the Problem

Same film but different titles

The Approach

FAST & FURIOUS 9

Fast and Furious 9

Fast & Furious 9

F9 - Fast & Furious 9

API
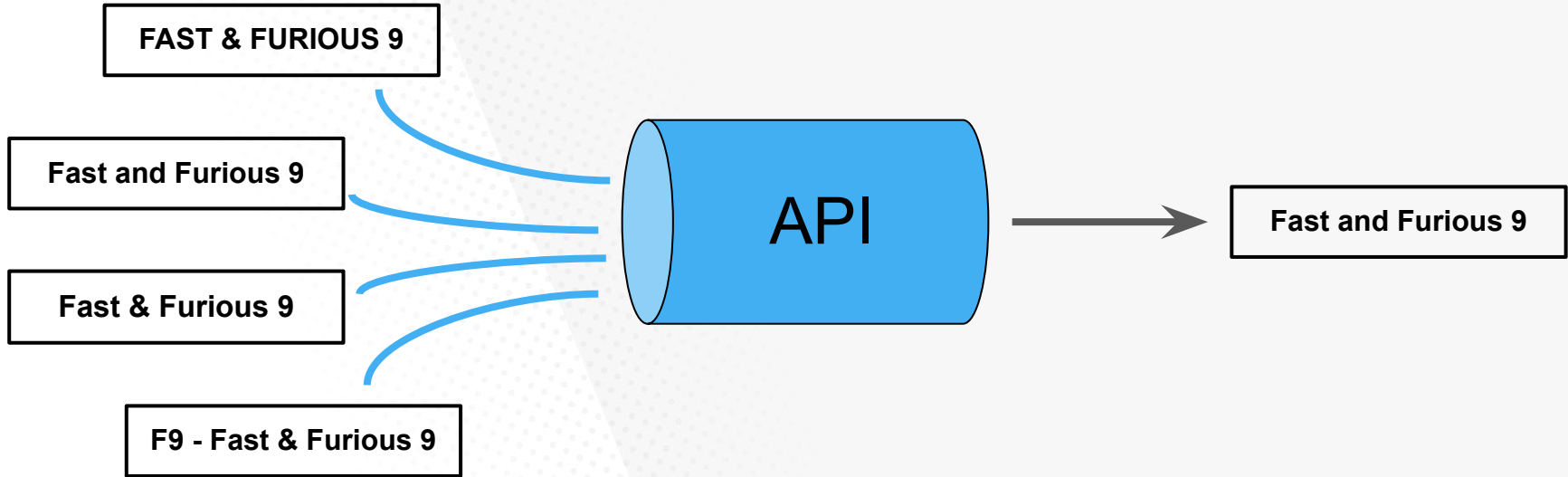
Fast and Furious 9

Implementation

Fast & Furiouse - 9 (2019) →

Cleanup:
- **White space**
- **Parenthesis**
- **Special Characters**

Parenthesis cleanup
Number Mutation
(...)
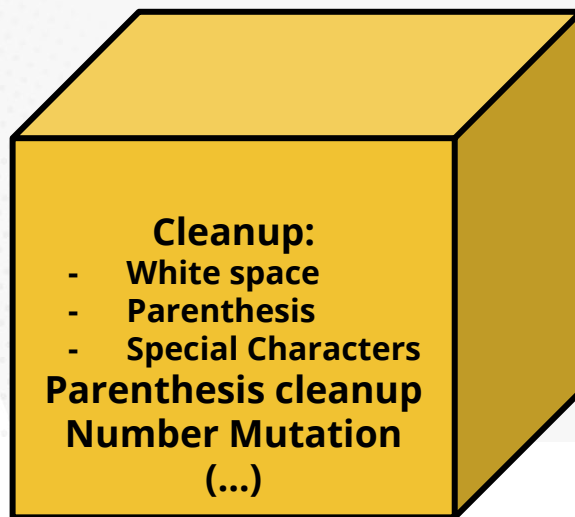
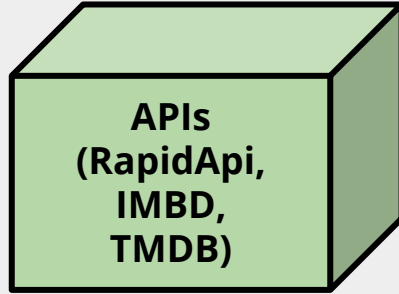| Name _____	Weight
Fast and Furiouse 9 (2019)	1
Fast and Furious 9 (2019)	0.95
Fast and Furious 9	0.9
Fast and Furious nine (2019)	0.9
Fast and Furious IX (2019)	0.7
Fast and Furious nine	0.7
Fast and Furious IX	0.6
Fast and Furious	0.4
Fast and 9	0.3
...	

OpusData   comscore   usheru   INTERNATIONAL SHOWTIMES API   AccuWeather   ni   gruvi

Implementation

For each variation:

All Titles → APIs (RapidApi, IMBD, TMDB) ✕ Variation weights + combined API rank → Fast and Furious 9
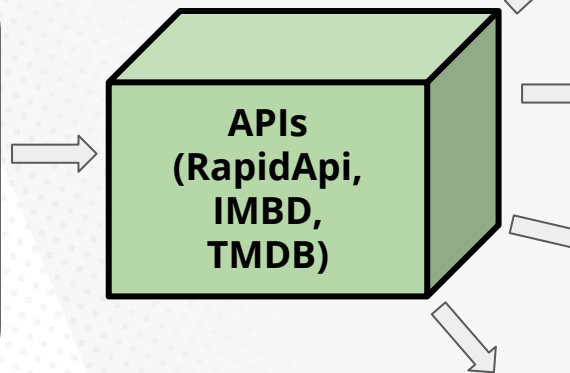
https://www.imdb.com/title/tt5433138/

How can we determine the confidence level?

Mutated Titles from "Fast and Furious 9 (2019)"

Name	Weight
Fast and Furious 9 (2019)	1
Fast & Furious 9 (2019)	0.95
Fast and Furious 9	0.9
Fast and Furious nine (2019)	0.9
Fast & Furious IX (2019)	0.7
Fast and Furious nine	0.7
Fast and Furious IX	0.6
Fast & Furious	0.4
Fast and 9	0.3

APIs (RapidApi, IMBD, TMDB)

Fast & Furious 9
Year: 2021
Cast: Vin Diesel, John Cena, Michelle Rodriguez, (...)
IMDB_ID: tt1234123

Confidence Factor: **0.934**

Furious 9
Year:
Cast:
IMDB_ID: tt412341

Confidence Factor: **0.732**

Nine Furious Men
Year: 1875
Cast:
IMDB_ID: tt0000123

Confidence Factor: **0.663**

Fast and Furious
Year: 2001
Cast: Vin Diesel, (...)
IMDB_ID: tt0040103

Confidence Factor: **0.535**

OpusData   comscore   usheru   INTERNATIONAL SHOWTIMES API   AccuWeather   ni   gruvi
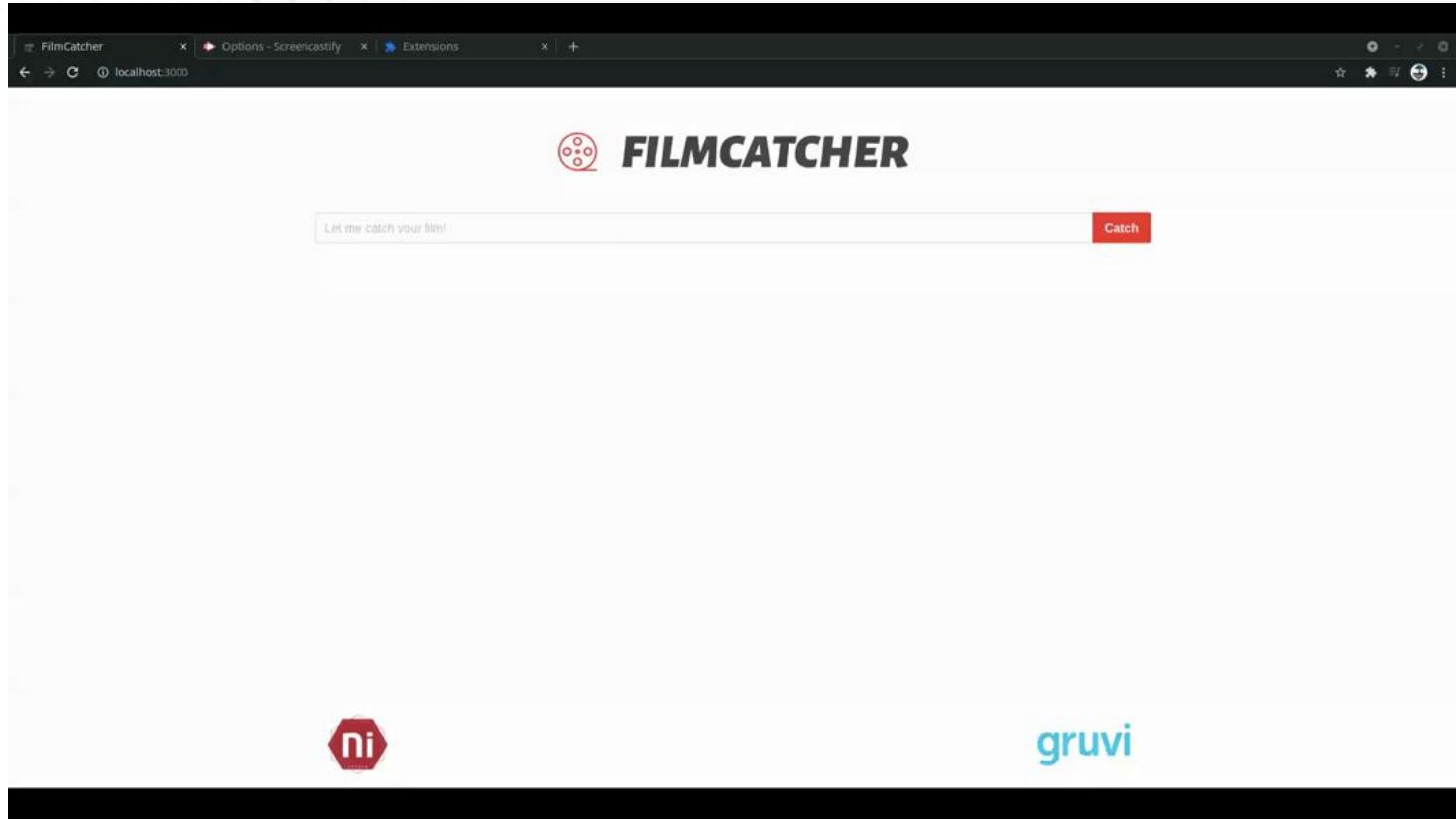
Practical Use Case

The first Harry Potter movie was received greatly by the general public. (...) On a personal level, I find Harry Potter & the Philosopher's Stone to be the best of the series. (...) In America, Harry Potter and the Sorcerer's Stone was a box office hit!

Film Catcher API

Harry Potter and the Philosopher's Stone

imdb.com/title/tt0241527/

Results, Final Remarks and Q&A

Provided test dataset:
https://hackathon-datasets.s3.eu-west-2.amazonaws.com/mismatched_titles_sample.csv

Our detailed results:
https://drive.google.com/file/d/1AbscNbEcMkBtPoeoeWKH6hPe3T7-Qn0h/view?usp=sharing

Removing unreleased movies:
Ratio of correct results before
- 1st result - 0.875
- 2nd result - 0.975
- 3rd - 1.0

Not removing unreleased movies:
Ratio of correct results before
- 1st result - 0.791(6)
- 2nd result - 0.958(3)
- 3rd - 0.9791(6)

Space Jam 2	tt0117705	0.9458334492	tt3554046	0.9284825333	tt5832032	0.9260929	tt5804204	0.908897
Space Jam: A New Legacy	tt3554046	0.9482410597						
Doctor Strange: In the Multiverse of Madness	tt9419884	0.9416110861	tt12616980	0.91676825	tt1208230	0.901013		
Doctor Strange in the Multiverse of Madness	tt9419884	0.8971575556	tt12616980	0.86058	tt1208230	0.809771		
The Matrix 4	tt10838180	0.9315346444	tt0133093	0.9029809791	tt0439783	0.896142	tt0234215	0.886096
The Matrix Resurrections	tt10838180	0.9137466	tt9397836	0.550704	tt3525520	0.4768	tt1092775	0.4768
No Time To Die	tt10839440	0.99952776	tt11917048	0.9519312	tt2382320	0.921978	tt0899070	0.896638
Sin tiempo para morir	tt2382320	0.9137466	tt4799088	0.90889772				
007 No Time to Die	tt2382320	0.8989103778	tt0075614	0.8875518136	tt0068444	0.856081	tt6132690	0.856081
The Boss Baby: Family Business	tt6932874	0.9475649722	tt8111438	0.9161146988				
Baby Boss 2 : une affaire de famille	tt6932874	0.9432107667						
Ainbo: Hrdinka pralesa	tt6570098	0.9090565556						
Ainbo	tt0098165	0.949787492	tt6570098	0.92499312	tt1348808	0.871725	tt1079565	0.78161
Charlie's Angels (2020)	tt5033998	0.9244722101	tt0160127	0.912356901	tt0073977	0.894676	tt1760943	0.894676
Charlie's Angels	tt5033998	0.957915758	tt0160127	0.9545260645	tt0305357	0.947974	tt0383915	0.940787
Godzilla: King of the Monsters	tt3741700	0.9511649225	tt0197521	0.9479815375	tt1089398	0.943373	tt0390053	0.931572
Godzilla II - King of the Monsters	tt3741700	0.9538502185	tt0197521	0.9060581525	tt0390053	0.906058	tt1089398	0.869143
POKÃ‰MON Detective Pikachu	tt5884052	0.9561971404	tt11011018	0.924430575	tt1303394	0.916114	tt1031920	0.901365
PokÃ©mon: Detective Pikachu	tt5884052	0.9521316009	tt13033940	0.9331521038	tt1101101	0.908204	tt1293005	0.901365
The Lion King	tt6570098	0.945169508	tt6105098	0.937352425	tt0120131	0.937352	tt0318403	0.928369
The Lion King (2019)	tt6570098	0.9145597268	tt11279422	0.8717252	tt1195878	0.871725	tt0120131	0.839053

OpusData   comscore   usheru   INTERNATIONAL SHOWTIMES API   AccuWeather   ni   gruvi

Results, Final Remarks and Q&A

Features that could be further developed:

- Language detection

- The use of more API's to search movie databases

- Implementation of additional user prompts to narrow down searches

- Improvements to confidence factor calculations (possible use of ML)

Blue Jeans & Bloody Tears

- Eduardo Parracho
- João Barbosa
- João Macedo
- Tito Griné