



KPMG NLP Skill Assessment

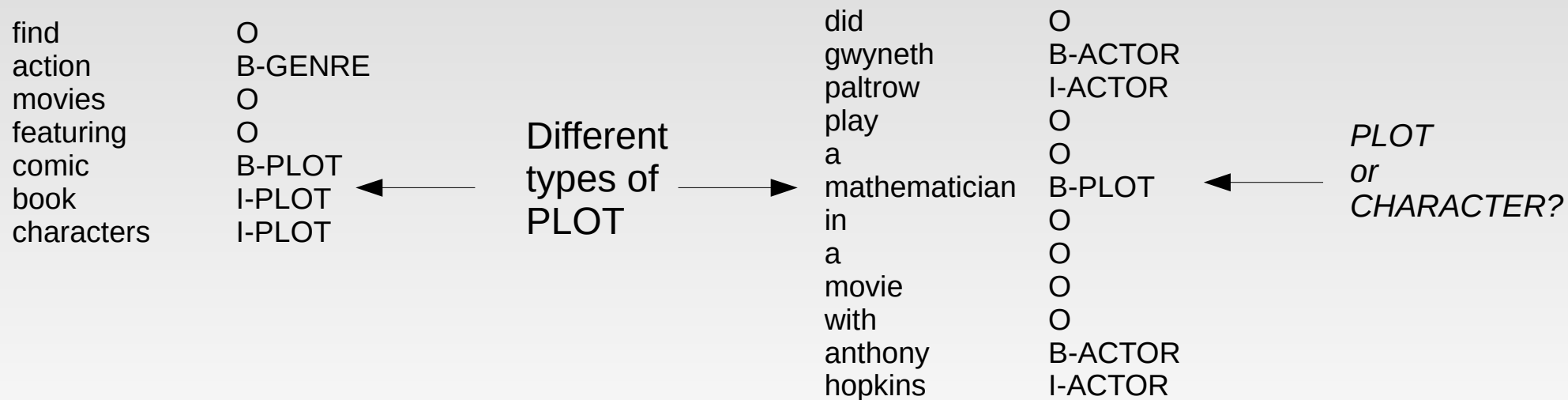
Marcos Martínez Galindo

- Data Insights
- Approach Overview
- Results & Metrics
- Next Steps
- Demo



Data Insights

- Queries about movies - BIO Labelled
- Complex queries & entities



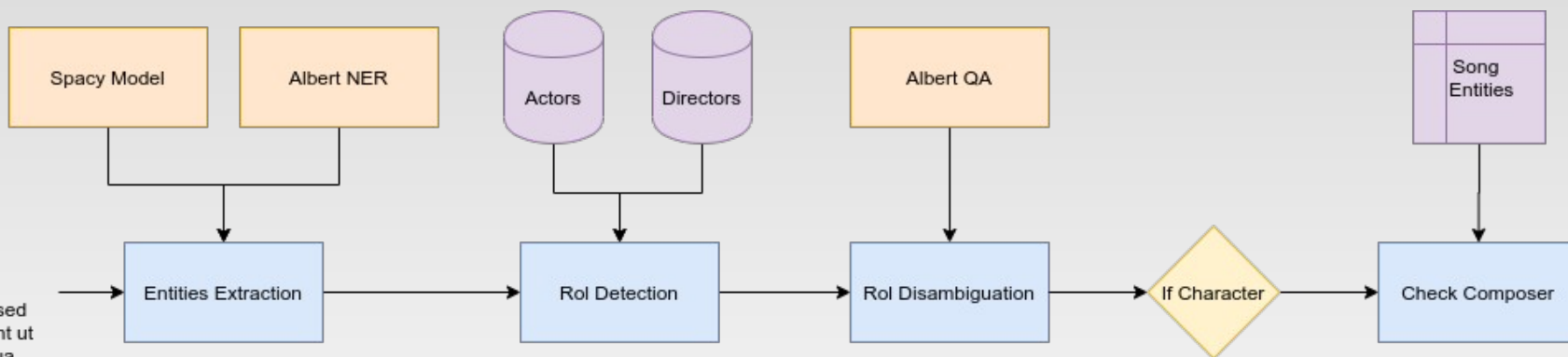


Approach Overview

Machine Learning + Deep Learning + Regular Expressions + Knowledge resources

Review

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

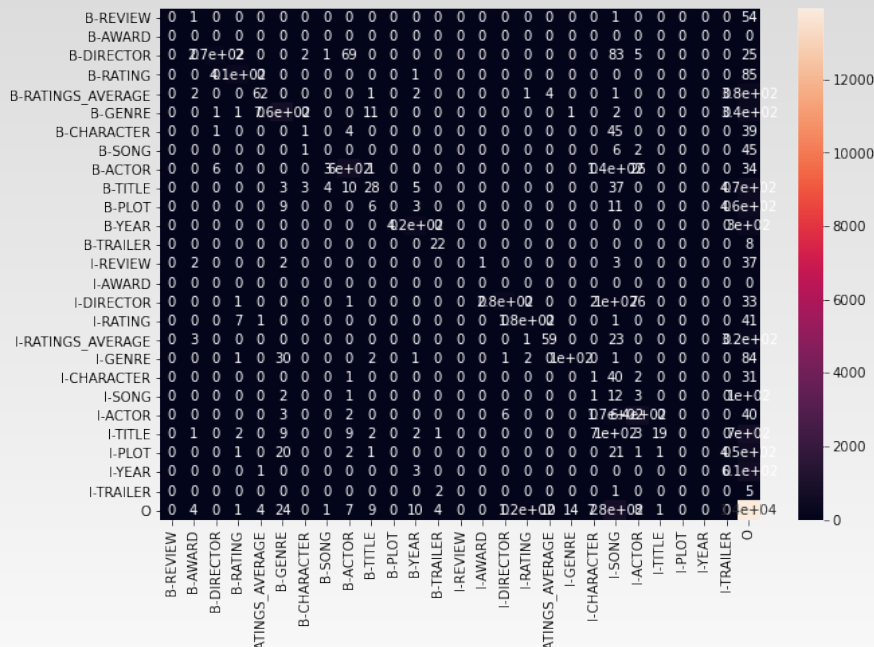




Results & Metrics

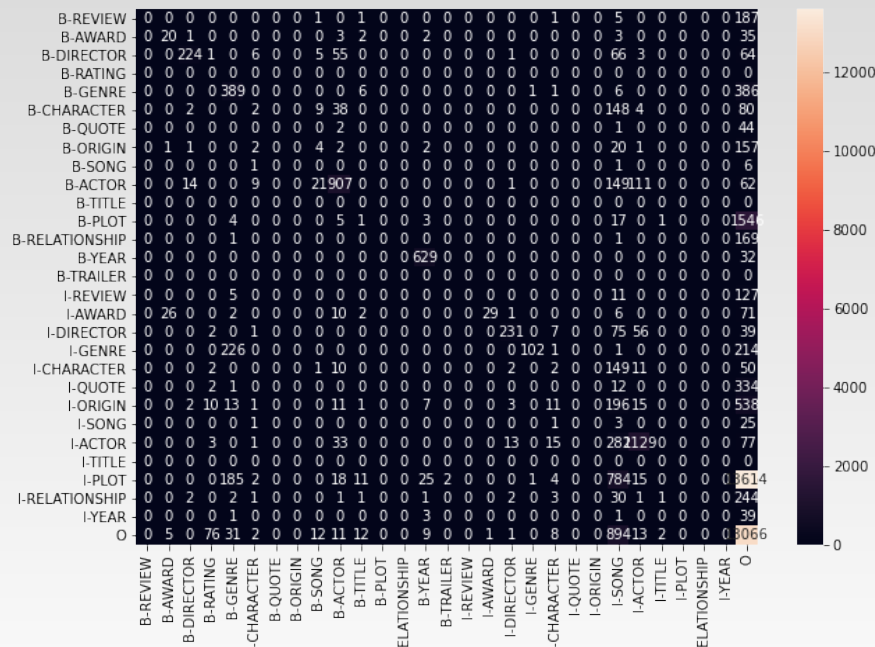
engtest.bio

f1-score: 0,68



trivia10k13test.bio

f1-score: 0,74





Next Steps

- Retrain LM with movie reviews - Domain Adaption
- Truecaser to solve the casing problems
- Train NER names in lowercase
- Use Levenshtein distance
- To improve regular expressions
- To augmentate knowledge resources



DEMO

<http://mmartgal.ddns.net:5000>

