

# Recognizing Film Entities in Podcasts

Ahmet Salih Gundogdu, Arjun Sanghvi, Keith Harrigian

Applied Analytics Solutions - A WarnerMedia Company



# Motivation

- Increasing popularity of podcast usage over previous decade

# Motivation

- Increasing popularity of podcast usage over previous decade
- Limited related research on entertainment properties (i.e. film titles)

# Motivation

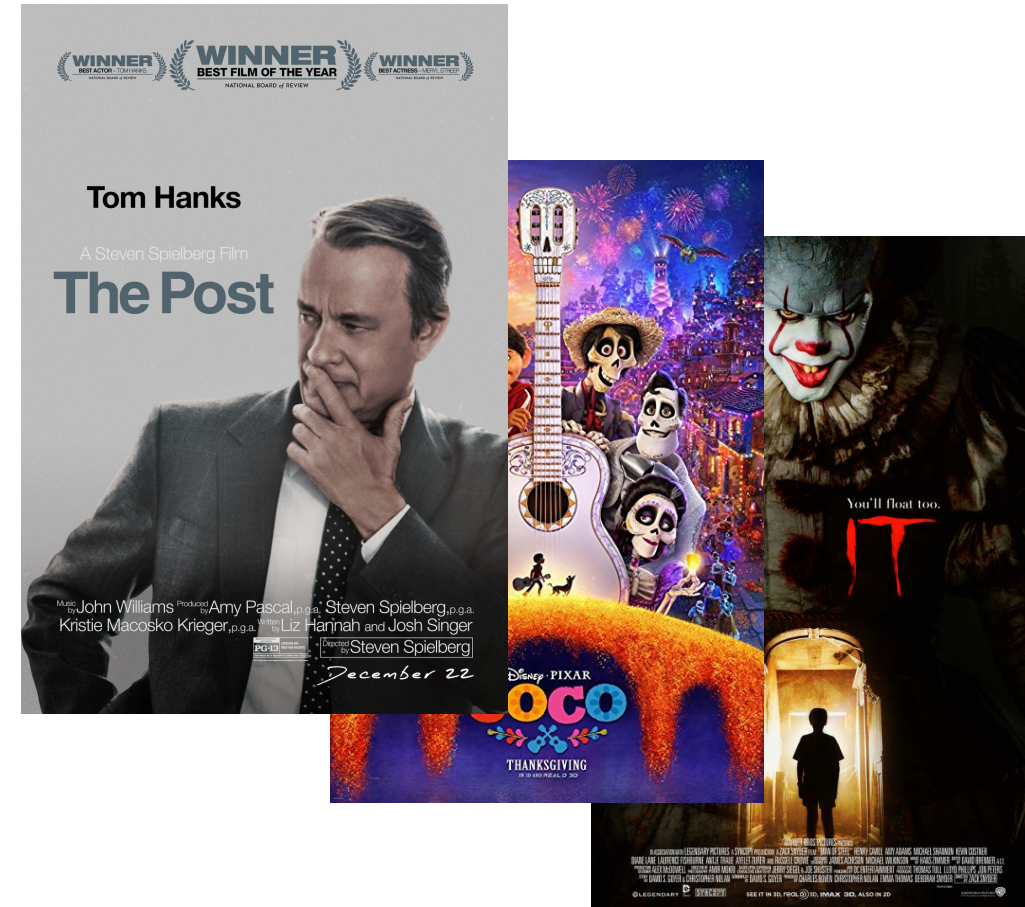
- Increasing popularity of podcast usage over previous decade
- Limited related research on entertainment properties (i.e. film titles)
- Need for Named Entity Recognition (NER) to identify properties within the podcast

# Motivation

- Increasing popularity of podcast usage over previous decade
- Limited related research on entertainment properties (i.e. film titles)
- Need for Named Entity Recognition (NER) to identify properties within the podcast
- Great potential to gauge social opinion
  - Large-scale topic analysis
  - Identifying emotions associated with properties
  - Building predictive models for property-level response variables

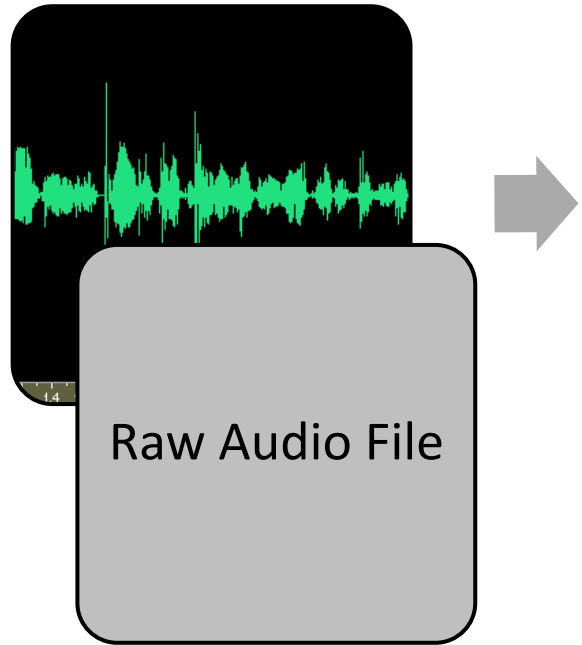
# Main Challenges

- Recognizing creative works
  - Disambiguation from common words
  - New creative works release every day
- Domain of spoken language
  - Need for an intermediary step (transcription)
  - Previous tools are designed for written language

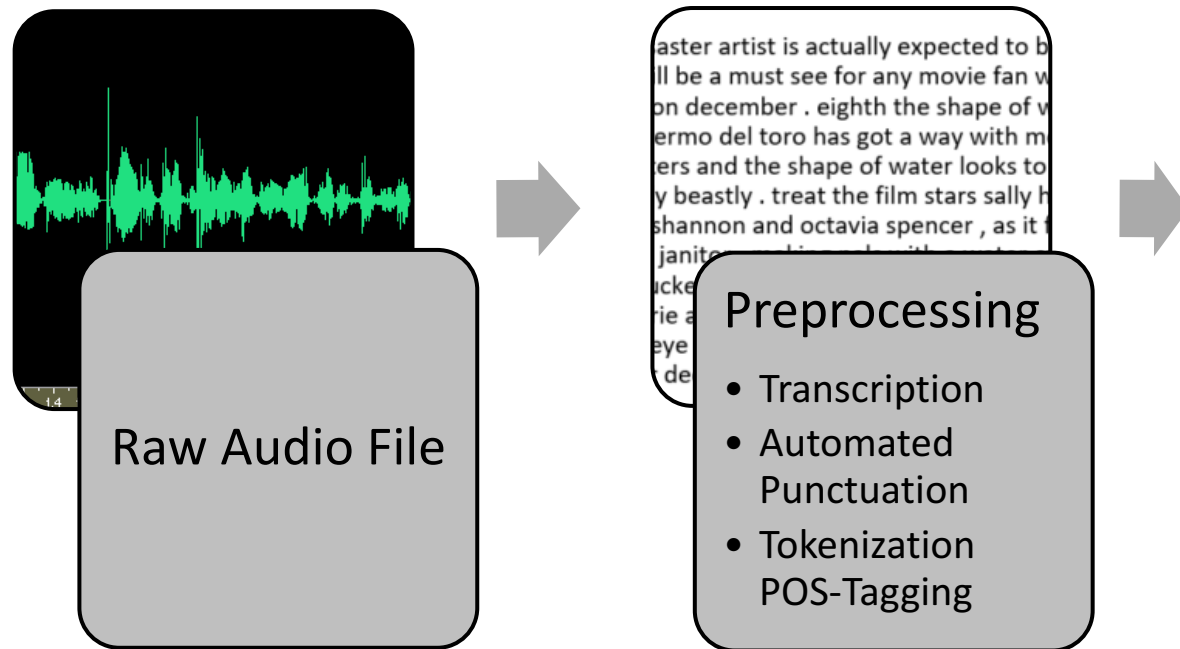


**NER is really challenging task especially for audio applications**

# Proposed System

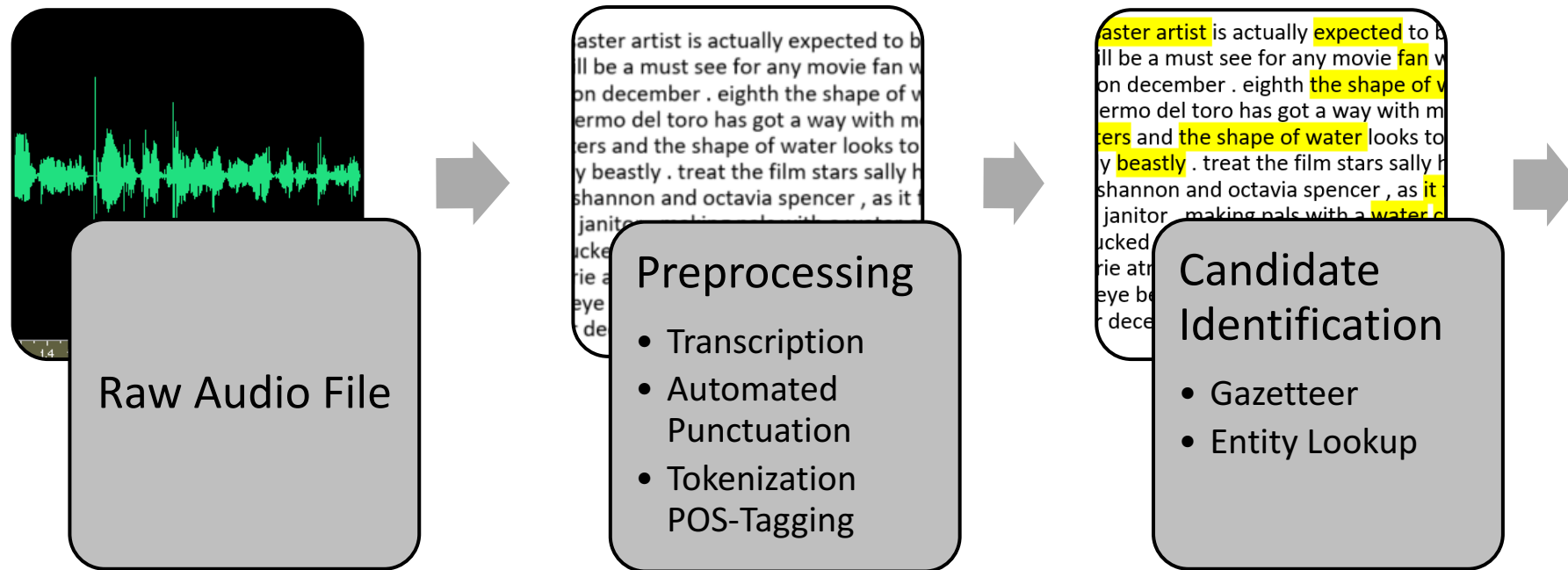


# Proposed System

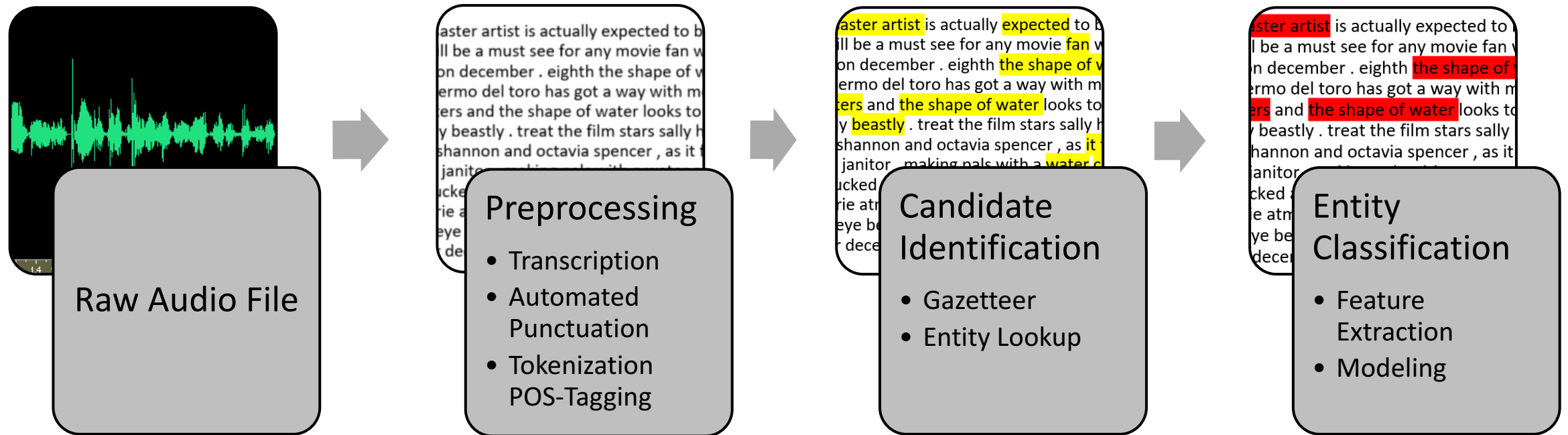




# Proposed System



# Proposed System



# Data and Preprocessing

- Data
  - Manual annotation
  - Podcast lengths (5-15 minutes)

Channels	#1	#2	#3	#4	Total	$\overline{WER}$
Collider	51	2	-	-	53	2%
All Things Consd.	4	2	-	-	6	26%
Looper	4	16	37	-	57	2%
Bob Mondello	20	15	5	-	40	10%
Morning Ed.	3	11	-	-	14	23%
Screen Junkies	20	30	3	11	64	2.5%
Angry Joeshow	18	-	-	-	18	4%
Fresh Air	14	12	-	-	26	25%
SlashFilm	38	-	-	-	38	38%

**Table 1: Label Distributions Across Podcast Channels Episodes with Average WERs**

# Data and Preprocessing

- Data
  - Manual annotation
  - Podcast lengths (5-15 minutes)
- Preprocessing
  - Transcription performed using open source speech recognition tool<sup>3</sup>
  - Punctuation inference<sup>1</sup>
  - POS-Tagging<sup>2</sup>

Channels	#1	#2	#3	#4	Total	$\overline{WER}$
Collider	51	2	-	-	53	2%
All Things Consd.	4	2	-	-	6	26%
Looper	4	16	37	-	57	2%
Bob Mondello	20	15	5	-	40	10%
Morning Ed.	3	11	-	-	14	23%
Screen Junkies	20	30	3	11	64	2.5%
Angry Joeshow	18	-	-	-	18	4%
Fresh Air	14	12	-	-	26	25%
SlashFilm	38	-	-	-	38	38%

**Table 1: Label Distributions Across Podcast Channels Episodes with Average WERs**

# Candidate Identification

- Gazetteer
  - 9,000+ rows of film metadata

Film Name	Synopsis	Keywords	Logline	Production Budget
Inside The Mind Of Leonardo In 3D	The film presents a visually dazzling experience of unparalleled access to Leonardo da Vinci's revolutionary (and controversial) ideas a...	[architect, artist, book, discovery, educational, genius, historical, invention, inventor, Italy, painter, Renaissance, science, scientist]	N/A	N/A
Inglourious Basterds	N/A	[1940s, 1941, 1944, Adolf Hitler, Black List, Black List (2008), epic, family, farmer, film, France, Jewish, military, murder, Nazi, Par...	A group of Jewish American soldiers are enlisted to take down the leaders of the Third Reich.	\$70.0MM

- Entity Lookup using Fuzzy matching based on Levenshtein ratio
  - Word Error Rate ("A Fantastic Woman" transcribed as "The Fantastic Woman")
  - Homonyms ("Coco" transcribed as "Cocoa")

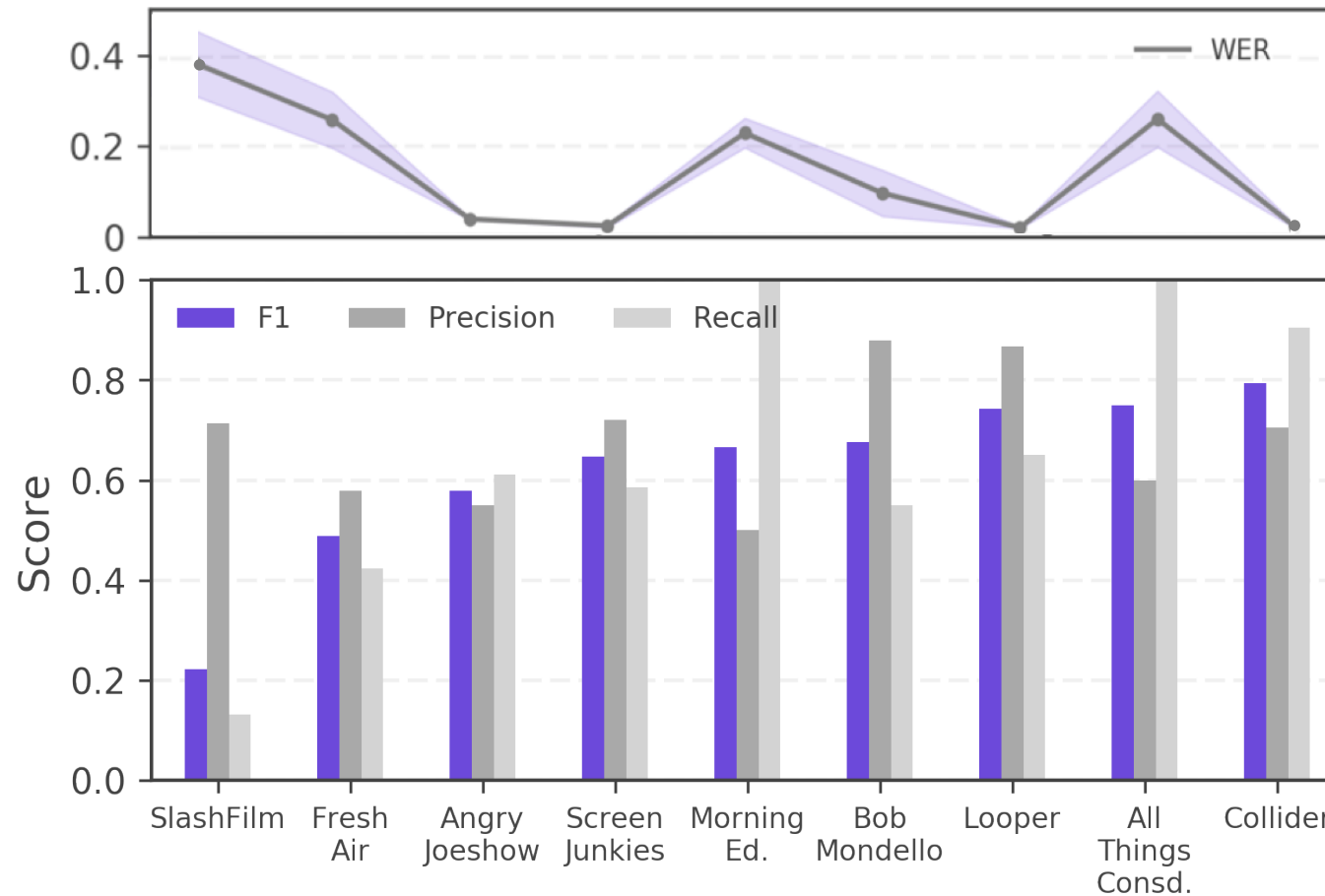
# Entity Classification

- Feature Extraction
  - Extract contextual and syntactical features from entity candidates
- Modeling
  - Logistic Regression used to identify true film mentions and filter out false positives
  - Model does not require retraining when the new film releases

Features	Definition
Closeness	Closeness value between movie $m$ and keyword $k$
Levenshtein Ratio	$lev(a, b)$ measures the minimum number of deletions, insertions and substitutions to transform $a$ into $b$
Production Budget	Normalized budget for film $m$ within the transcript
Title POS-tags	Bag of POS-tags in the title of movie $m$
Pre/Post POS-tags	POS-tags of preceeding and succeeding word for candidate
N-gram Levels	Number of tokens in entity candidate

# Results: Generalization Error

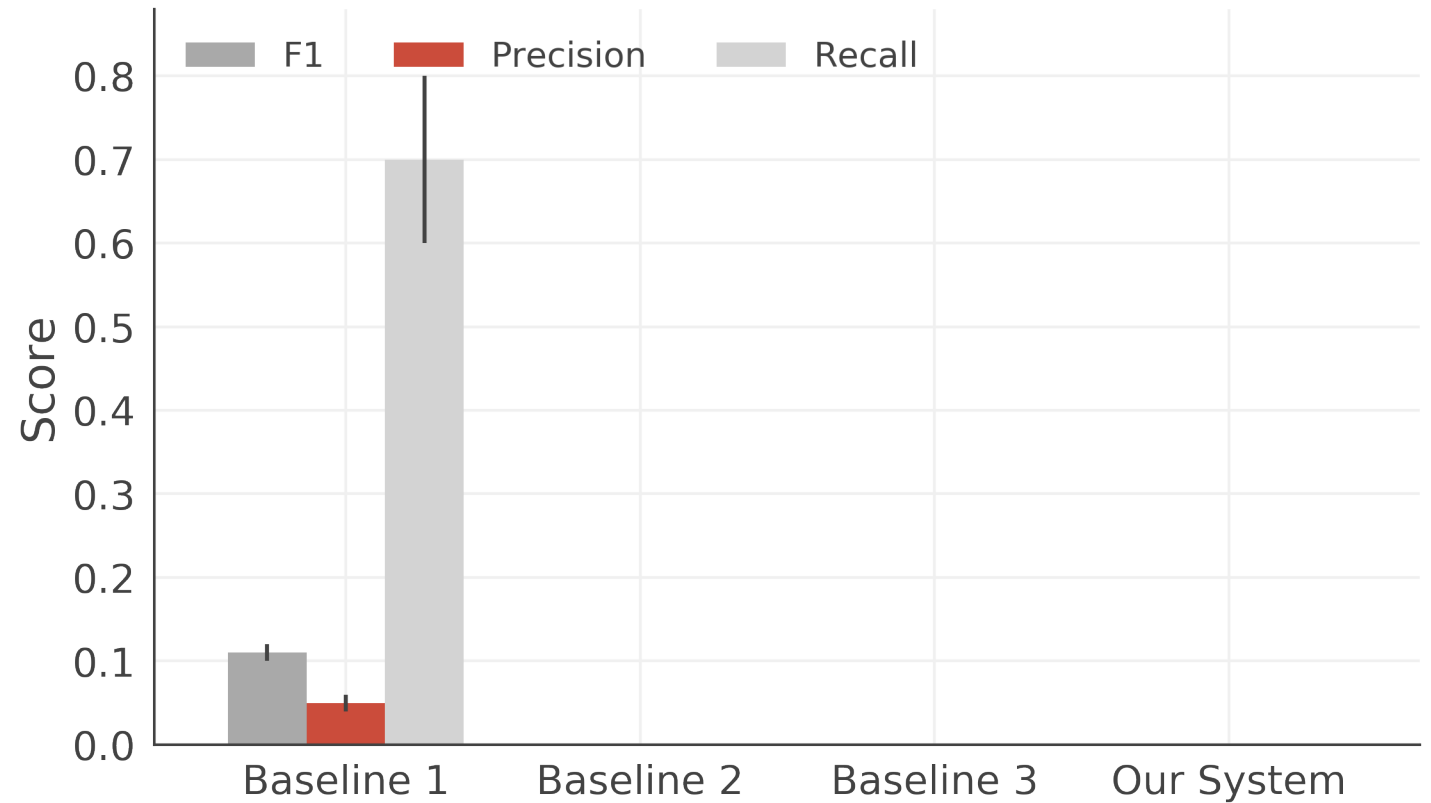
- Evaluate generalization error using 'Leave One Channel Out' cross-validation (Training on data from all the channels except one)



# Results: Model Comparisons

- Baseline comparison
  - B1: Identified Candidates

**Figure 1: Average Results of CV for Baseline 1**

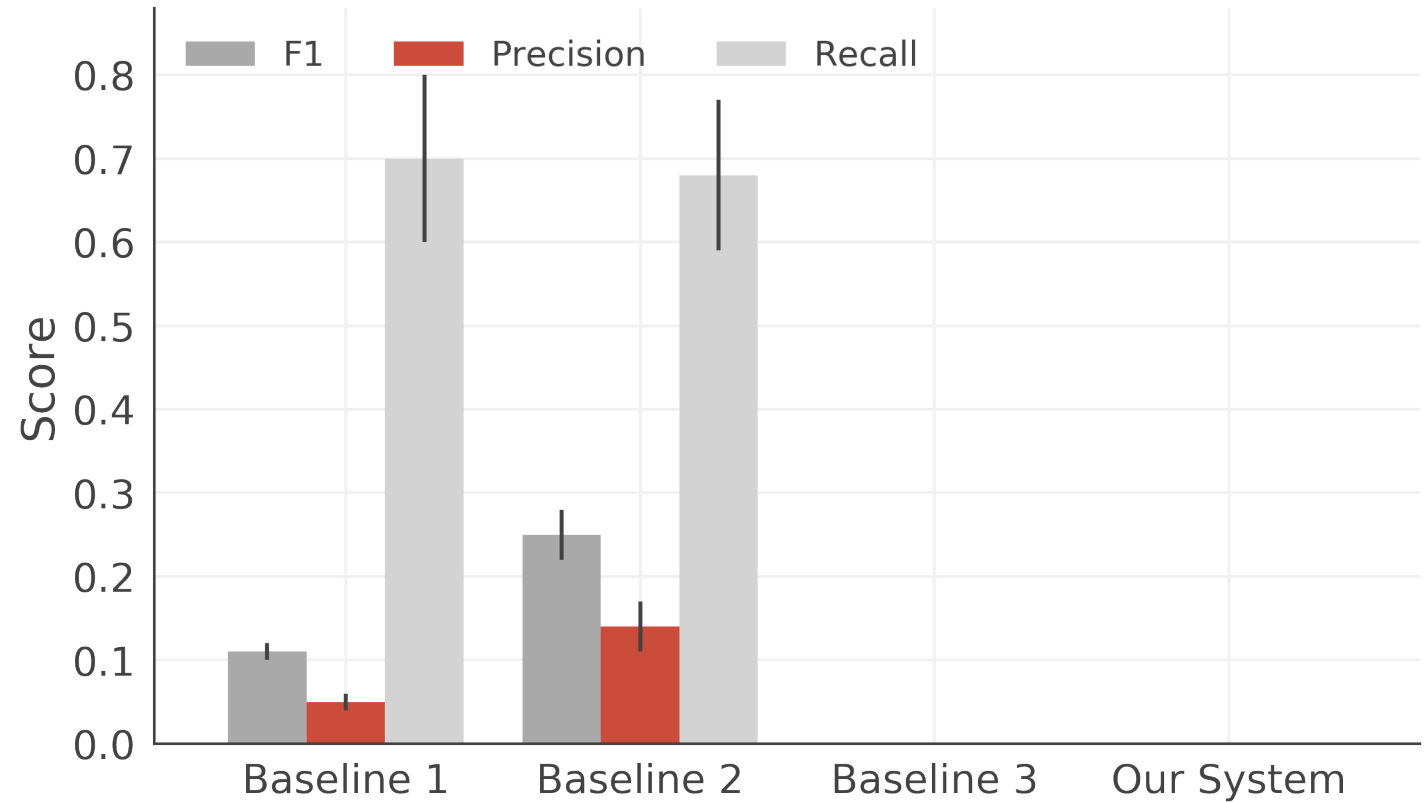




# Results: Model Comparisons

- Baseline comparison
  - B1: Identified Candidates
  - B2: Name Phrase Filter applied on B1

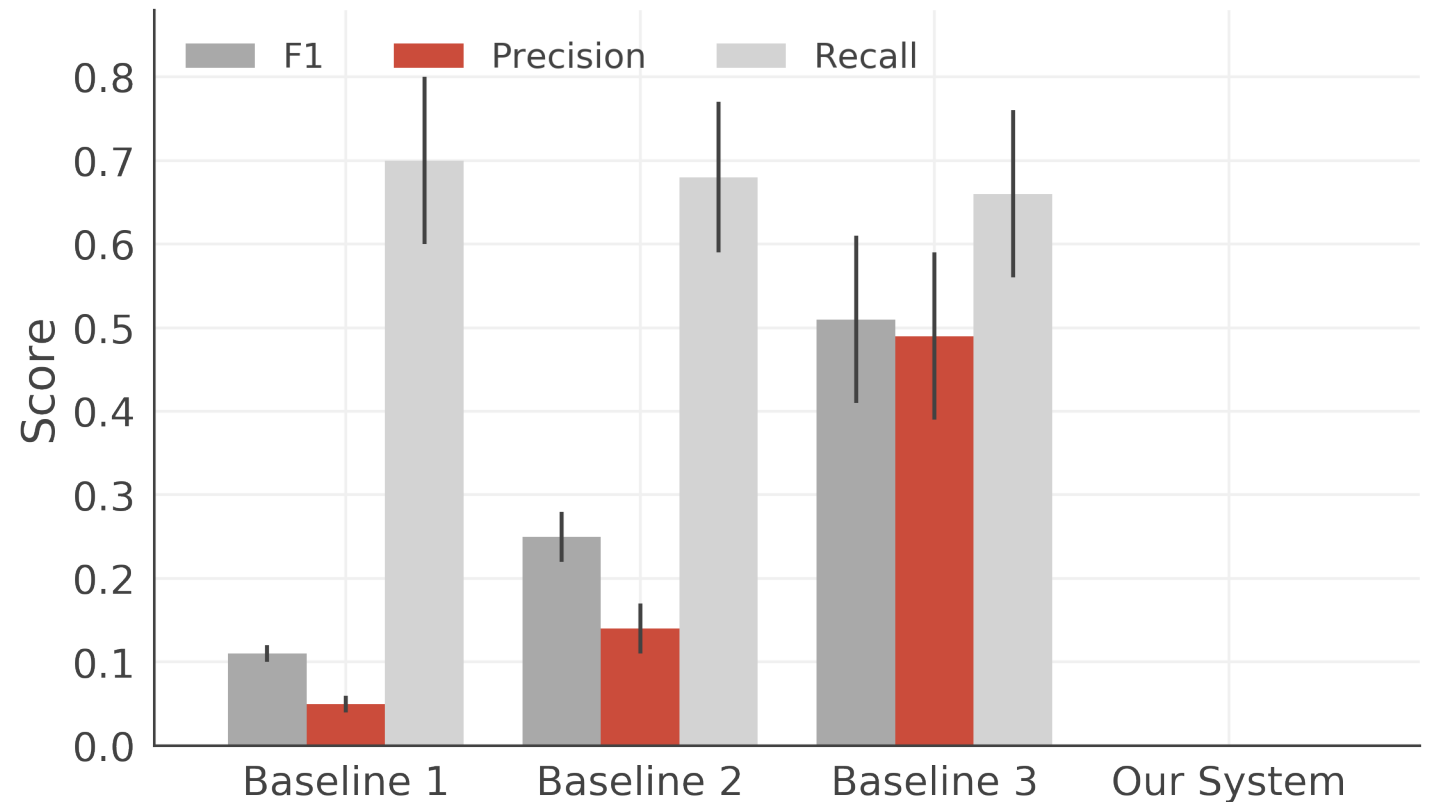
**Figure 2: Average Results of CV for Baseline 1 and 2**



# Results: Model Comparisons

- Baseline comparison
  - B1: Identified Candidates
  - B2: Name Phrase Filter applied on B1
  - B3: Rule-based

Figure 3: Average Results of CV for Baseline 1, 2, and 3

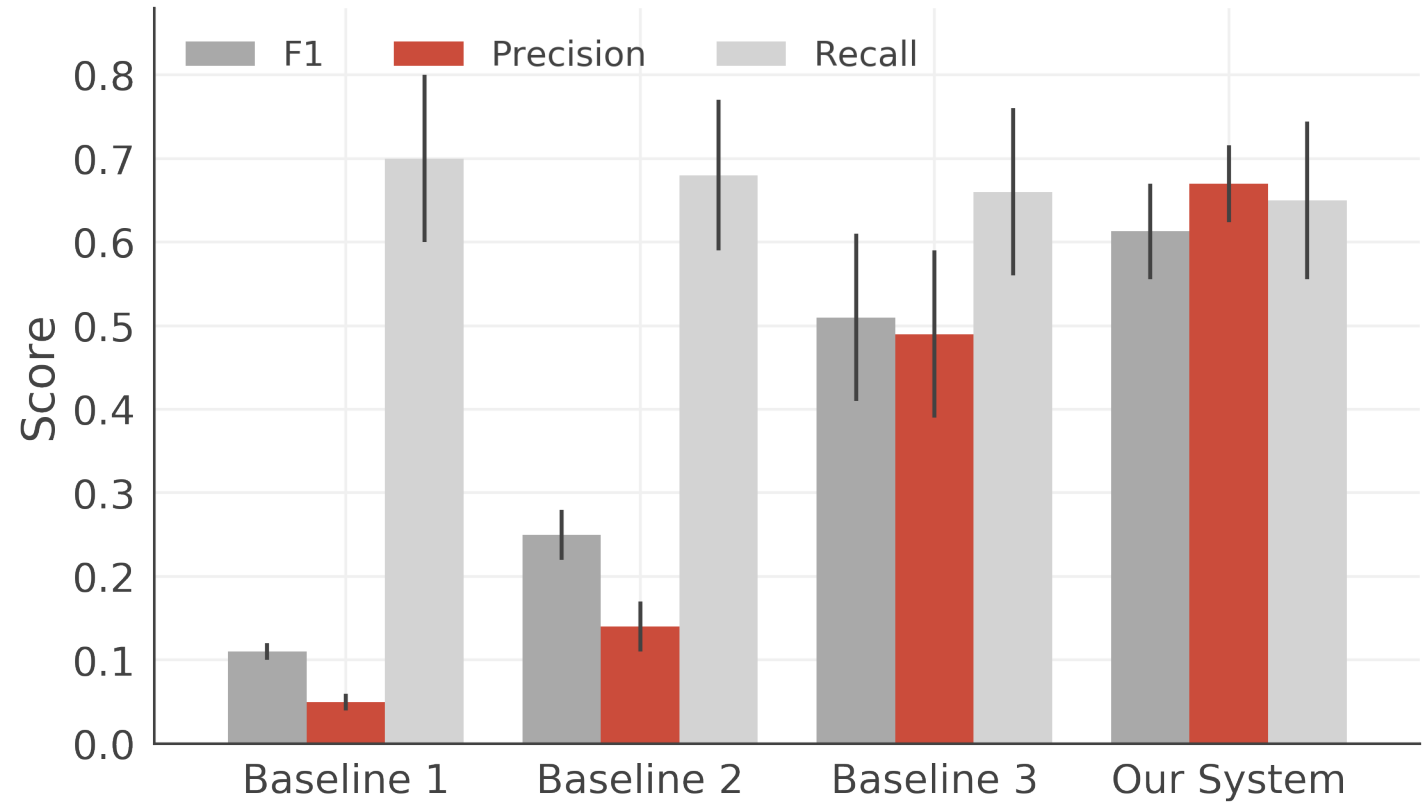


**The rule-based approach demonstrates that metadata adds most of the predictive power.**

# Results: Model Comparisons

- Baseline comparison
  - B1: Identified Candidates
  - B2: Name Phrase Filter applied on B1
  - B3: Rule-based
- Our System

Figure 4: Average Results of CV for Baseline 1, 2, 3, and our system

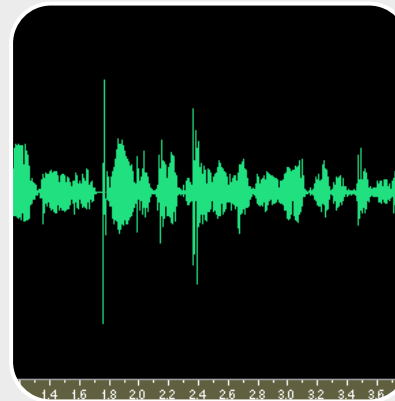


**The linear model is better at taking into account multiple features.**

# Future Directions

- Incorporate additional metadata (release date, production studio, cast members)
- Create larger and more granular set of podcast transcripts to use data-greedy sequence learning models.

- Thank you for listening...
- Questions?



ster artist is actually expected to  
be a must see for any movie fan  
on december . eighth the shape of  
ermo del toro has got a way with m  
ers and the shape of water looks to  
beastly . treat the film stars sally  
hannon and octavia spencer , as it  
janitor , making pals with a water c  
cked away in a secret research tan  
ie atmosphere and authority of de  
ye behind this film . consider your  
december eighth already booked

Link to presentation: <https://github.com/asgundogdu/kdd18-RecognizingFilmEntitiesinPodcasts>

# References

- [1] Tilk, O., & Alumäe, T. (2016). Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration. In *Interspeech* (pp. 3047-3051).
- [2] Toutanova, K., & Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger.
- [3] Zhang, A. (2017). Speech Recognition (Version 3.8) [Software]. Available from [https://github.com/Uberi/speech\\_recognition#readme](https://github.com/Uberi/speech_recognition#readme).



# Results: Robustness to Transcription error

- Estimating the effect of WER
  - Correctly identifies 27/33 true film mentions in the human-curated transcript
  - As opposed to 22/33 true mentions in the computer-generated transcript.

**The system has room to improve given more stable speech transcription models.**