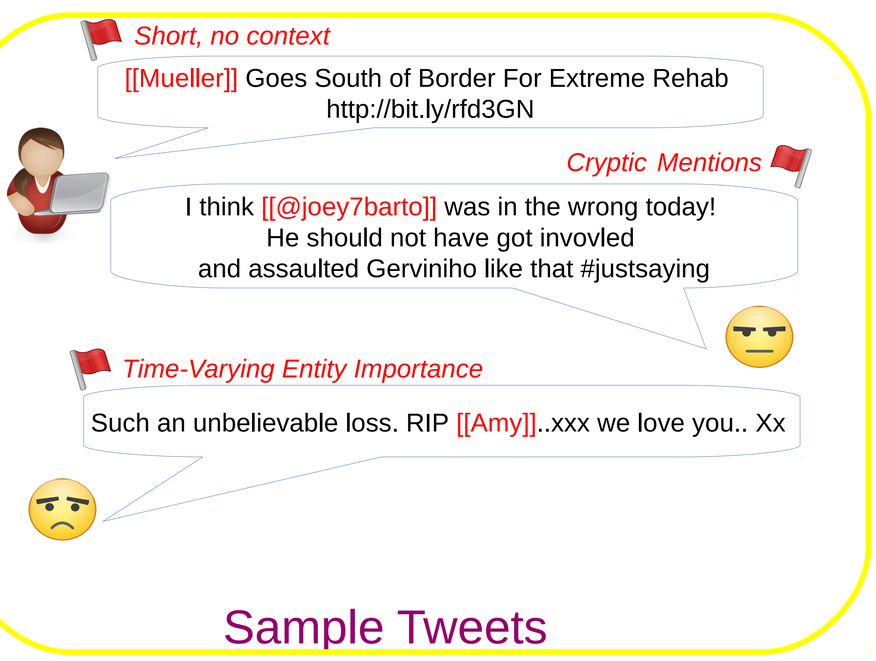
AIDA-Social: Entity Linking on the Social Stream

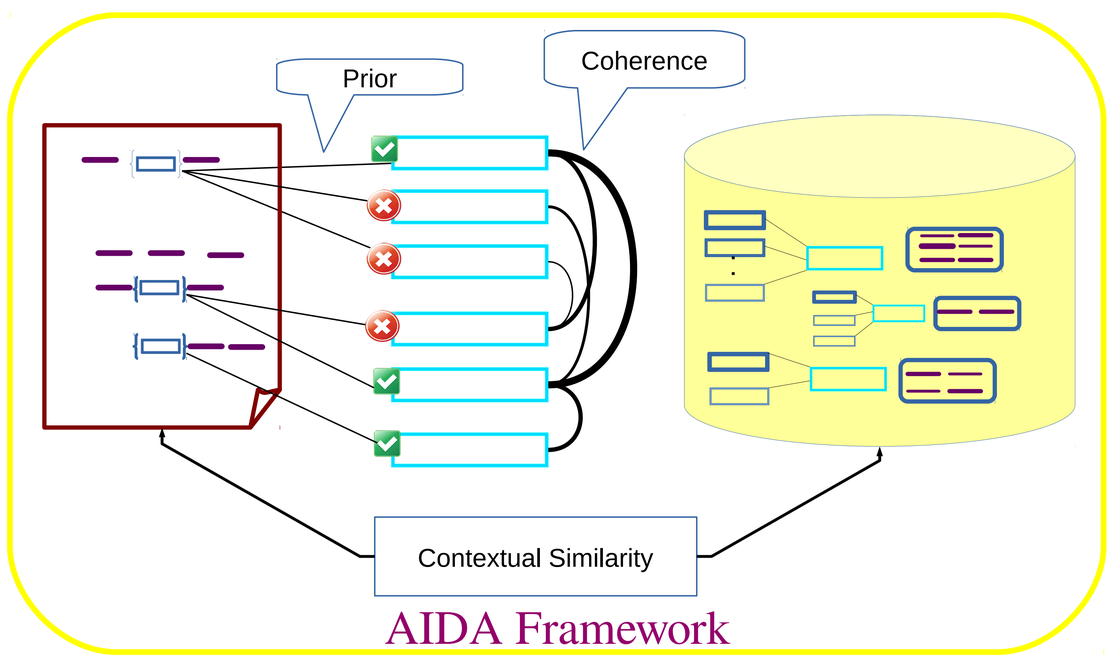
Yusra Ibrahim, Mohamed Amir Yosef, Gerhard Weikum
{yibrahim|mamir|weikum}@mpi-inf.mpg.de

Motivation

- The tremendous increase in social media popularity drove an abundance of user-generated content.
- Adding semantics to this content assists in subsequent Information Retrieval tasks such as relation extraction and semantic search.
- Named Entity Linking (NEL) disambiguate names to their corresponding canonical entities in Knowledge Bases such as YAGO.
- NEL in social media, more specifically in microblogs, is a challenging task due to the brevity, lack of contextual information, and time-varying importance of entities.
- Twitter is the most popular microblog with 500 million Tweets per day.

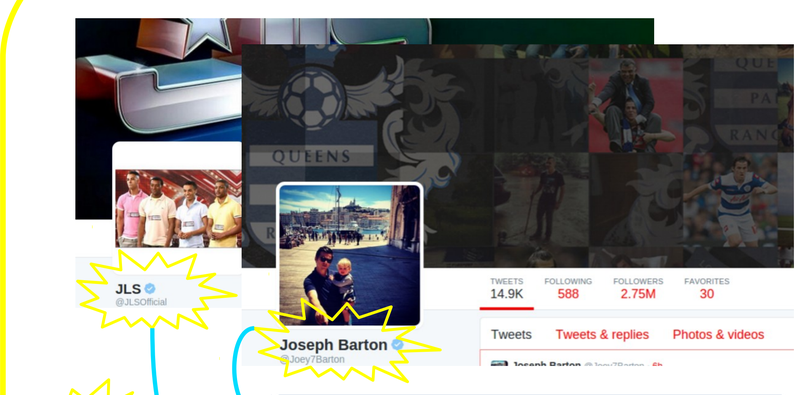
Challenges





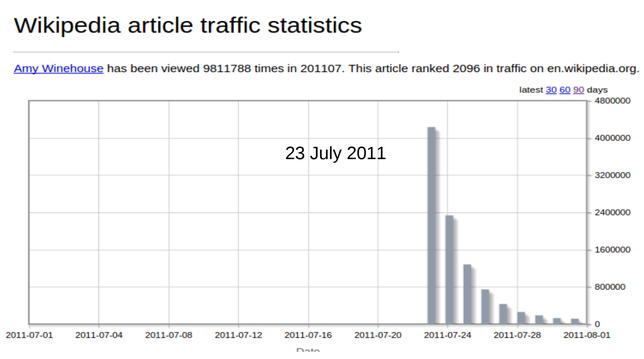
Proposed Techniques

Mention Normalization



Build a normalized set of mentions for every Cryptic mention

Temporal Importance

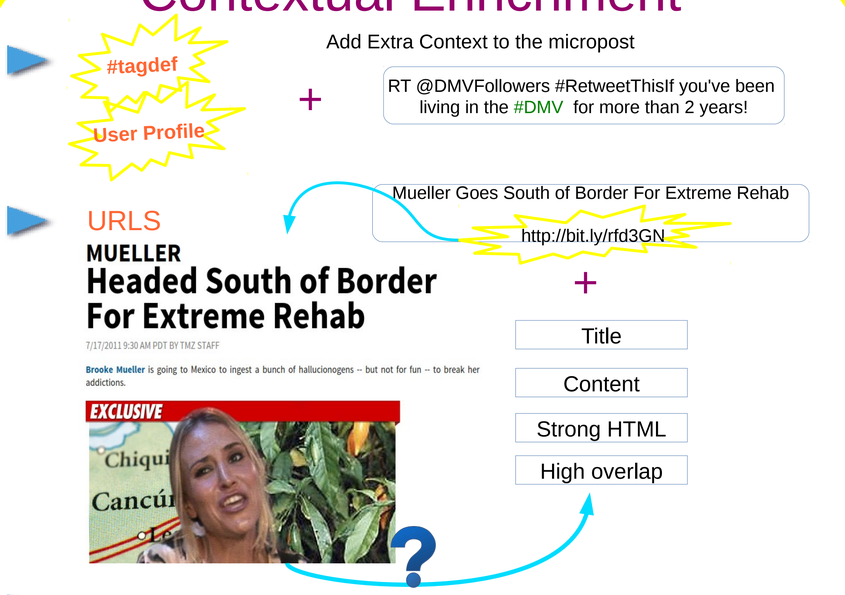


Such an unbelievable loss. RIP Amy..xxx we love you.. Xx

Amy Wright vs. Amy Winehouse
116 vs. 4231460

Calculate Entity importance based on the micropost's publication date and time .

Contextual Enrichment



Experimental Results

#Microposts2014 annotated **corpus**
Filter out: numbers, General Concepts, OOKB

Tweets	2809
Mentions	3392
Unique Mentions	2019
@	506
#	1085
Entities	1683

Baseline: AIDA - Prior + Contextual Similarity

Mention Normalization	Temporal Importance	Contextual Enrichment	P@1
✗	✗	✗	58.41%
✓	✗	✗	61.07%
✗	✓	✗	67.35%
✗	✗	✓	62.98%
✓	✓	✗	69.69%
✓	✗	✓	66.00%
✗	✓	✓	69.07%
✓	✓	✓	72.19%

Conclusion and Future Work

- ✓ Temporal Entity Importance, Contextual Enrichment, and Mention Normalization yields +13% gain over P@1.
- ✓ Adding extra context assists in mention-entity similarity measures
- ✓ Temporal Importance improves the accuracy of NEL, specially on plain tweets.
- ✓ Adding context from URLS yields better gain than adding context from #Tags and @userMentions
- ✓ Clustering assists in coherence measures.
- → TODO: Other Datasets
- → TODO: Other ways to estimate Temporal Importance
- → TODO: More on Clustering

MAX-PLANCK-GESELLSCHAFT

max planck institut informatik