Effect of Enriched Ontology Structures on RDF Embedding-Based Entity Linking

Emrah Inan and Oguz Dikenelli

(MTSR 2017)

11th MTSR
Metadata and Semantics Research Conference
Tallinn 2017

Presentation by Yannis Marketakis

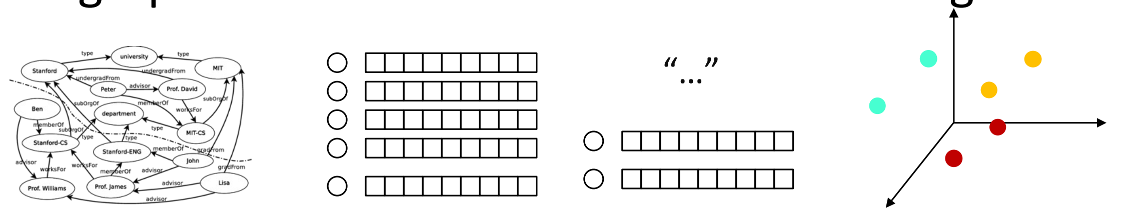
In the context of HY563 (Computer Science Department, UOC)

Outline

In one slide

Background

- RDF KB embeddings resolve ambiguous entities by leveraging semantic representations of entities in RDF graphs
  - Richer graphs → more effective RDF embeddings



RDF Graph *Graph Embeddings* *Text Embeddings* *Similarity Calculation*

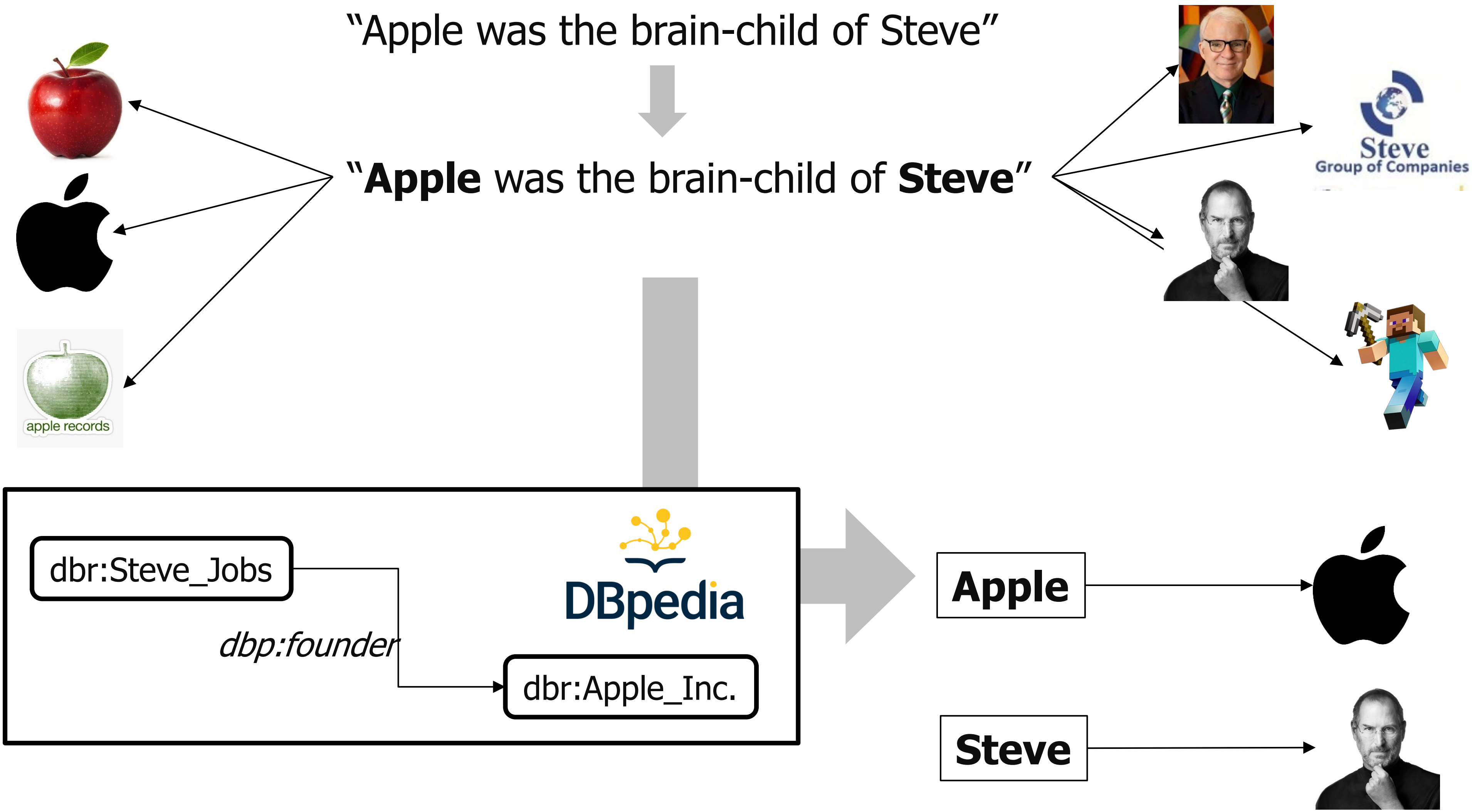
Goal

- Study the impact of different enriched structures of ontology & instances
  - There are more entities to disambiguate

Results

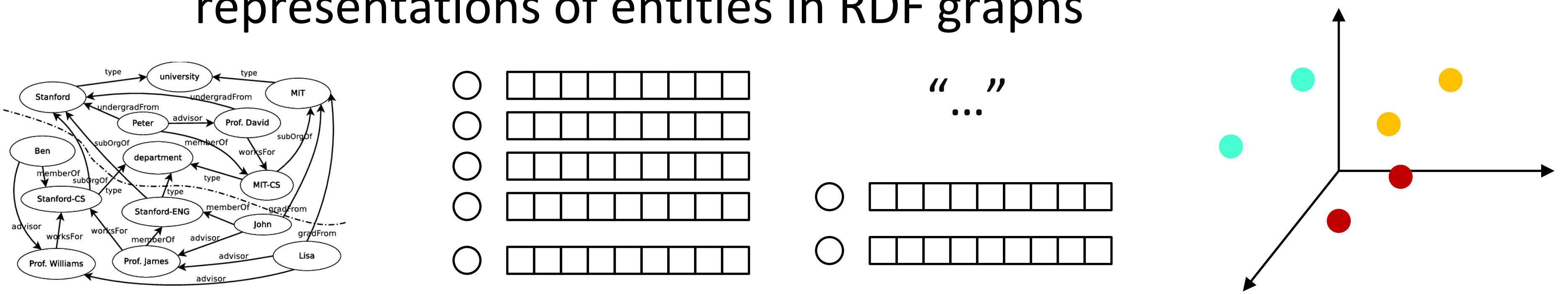
- Evaluation
  - In a knowledge-agnostic manner
  - Using state of the art systems: AGDISTIS, DoSeR

Introduction – Entity Dissambiguation



Introduction – ED using RDF Embeddings

- Resolve ambiguous entities by leveraging semantic representations of entities in RDF graphs



RDF Graph *Graph Embeddings* *Text Embeddings* *Similarity Calculation*

- Based on the similarities between text mentions and RDF entities, the entity with the highest similarity score is selected
- Additionally: refine and improve disambiguation through
  - Domain-specific knowledge
  - Contextual information
  - Entity popularity

Introduction – ED Ranking entities

- Similiary Calculation relies on ranking of candidate entities
  - Independent (or local) ranking
  - Collective (or global) ranking
- Semantic relatedness is widely used but…
  - Most approaches rely on Wikipedia/Dbpedia
- Knowledge-agnostic ED approaches overcome such problems
  - AGDISTIS [5]
    AGnostic DISambiguation of named entiTIeS using linked open data

ο/η ʼΑγδιστης

  - DoSeR [6]
    Dissambiguation of Semantic Resources

[5]: Usbeck et al. (2014) AGDISTIS-graph-based disambiguation of named entities using linked data
[6]: Zwicklbauer et al. (2016). Doser-a knowledge-base-agnostic framework for entity disambiguation using semantic embeddings

Introduction - Motivation

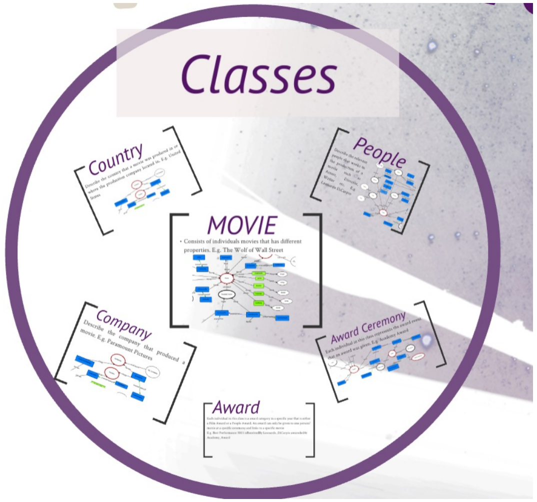
- The successfull results of DoSeR sparked the idea of investigating which joined component of ontology structures has the most impact to RDF embeddings-based semantic relatedness
- Create different enhancements of a knowlegde base
  - Ontological updates: classes, properties
  - Instances updates: ontology instantiations

- Adapted the evaluation systems accordingly
  - HITS algorithm in AGDISTIS
  - PageRank algorithm in DoSeR

Related Work

- Open-domain Entity Linking & Disambiguation
  - DBpedia Spotlight
  - Babelfy
  - WAT Entity Linker
  - TAGME (exploits Wikipedia links)
  - AIDA-light (exploits YAGO2 and Wikipedia)
- Domain-specific approaches
  - AGDISTIS
    - It first identifies entities, and then creates a disambiguation graph that uses HITS algorithm to match the best mention-entity pairs
  - DoSeR
    - Exploits RDF embedings of a knowledge base to compute semantic similarities between entities using a personalized PageRank algorithm

Enriched Ontology Structures



- Used as basis a ontology about movies
  - The Movies Ontology
  - Independent of Dbpedia links
- Variations/enhancements
  - MO-Per (Personal)
    - Enriched director and cast information
  - MO-Fin (Financial)
    - Enriched budget, distributor and producer properties
  - MO-Loc (Locational)
    - Extended location and language properties
  - Plus their combinations
- Enhanced in terms of adding: (a) classes, (b) properties and (c) instances

Enriched Ontology Structures – Quality

- Attribute Richness: average number of attributes for each class → indication of the density of classes
- Relational Richness: diversity of relations between classes
- Axiom/Class ratio: average number of axioms per class
- Class/Relation ratio: average number of classes per relationship

Schema metrics	MO-Per	MO-PerFin	MO-PerLoc	MO-PerFinLoc
Attribute Richness	0.054	0.056	0.12	0.248
Relational Richness	0.414	0.53	0.642	0.759
Axiom/Class ratio	10.865	10.882	11.487	11.98
Class/Relation ratio	0.702	0.785	0.836	0.878

Enriched Ontology Structures – Quality

- Average population: ratio of individuals per class
- Class richness: distribution of instances across classes. It is the ratio of classes having instances with total number of classes → high values means that ontology knowledge is represented by comprehensive data (instances)

KB metrics	MO-Per	MO-PerFin	MO-PerLoc	MO-PerFinLoc
Axioms	891	1062	1287	1521
Class count	82	95	112	127
Properties count	48	65	84	96
Individual count	298	372	468	565
Average population	3.634	3.92	4.185	4.456
Class richness	0.67	0.762	0.834	0.912

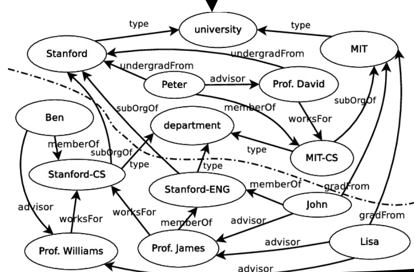
Entity Dissambiguation

TAG·ME

Text snippets

using

mentions (up to 6 words)



Movie ontology & instances

AGDISTIS
Adapted HITS

DoSeR
adapted PageRank

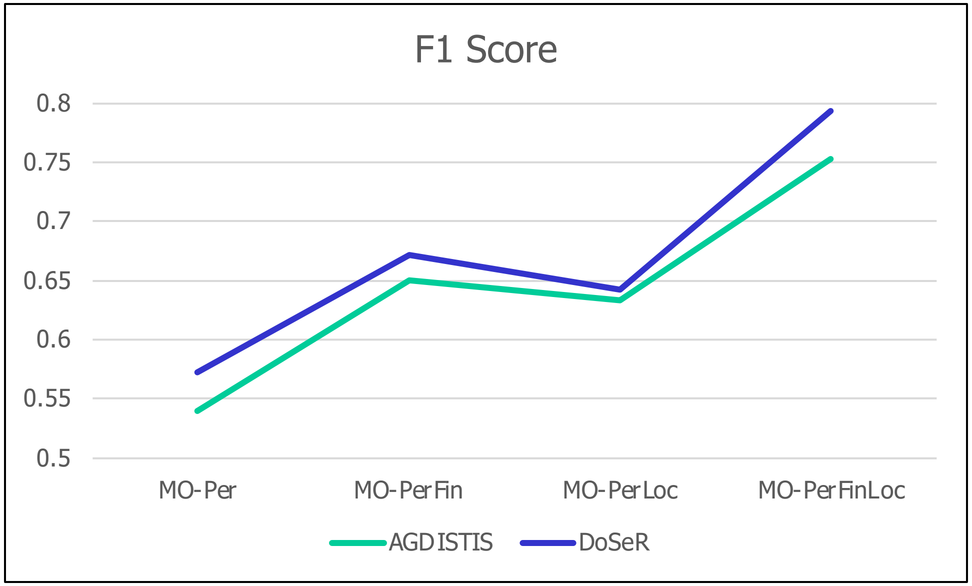
Evaluation – The evaluation set

- WeDGem*: an evaluation set generator for specific domains using Wikipedia articles
  - Annotated documents including entities for movies and their features (i.e. director, cast, genre, etc.)
- In numbers:
  - 945 annotated texts in English
  - 3,648 entities
- Used also Wikipedia disambiguation pages to increase the number of potential entities
  - https://en.wikipedia.org/wiki/Wicker_Park
    - https://en.wikipedia.org/wiki/Wicker_Park_(film)
    - https://en.wikipedia.org/wiki/Wicker_Park_(soundtrack)
    - https://en.wikipedia.org/wiki/Wicker_Park_(Chicago_park)
  - ~28.51 ambiguity of entities

*: https://github.com/einan/WeDGeM

Evaluation

Ontology & KB Variation	AGDISTIS			DoSeR		
	Prec	Rec	F1	Prec	Rec	F1
MO-Per	0.65	0.463	0.54	0.704	0.47	0.573
MO-PerFin	0.724	0.59	0.65	0.738	0.616	0.672
MO-PerLoc	0.709	0.572	0.633	0.729	0.574	0.642
MO-PerFinLoc	0.781	0.728	0.753	0.819	0.77	0.793



Conclusions

- Studied the efficiency and side-effects of different enriched ontology structures using state of the art knowledge-agnostic Entity Linking systems
  - AGDISTIS – using HITS algorithm
  - DoSeR – using PageRank algorithm
- Study contacted on a small sample set

- In future
  - Evaluate approached on bigger collections
  - Evaluation on different domains (other than movies)
  - Hybrid approach that will correlate the results of domain-specific and knowledge-agnostic Entity Linking systems

My Review

- My **positive** comments
  - Easy to follow, not very technical
- My **negative** comments
  - A comparison of structural vs instances enrichment would be interesting
  - Not clear why Movie Ontology (MO) was not used as a baseline for the evaluation
  - MO and enrichment are not available
  - Mixed the notion of ontology and knowledge base in the text
  - Language issues (grammar and syntax issues mainly)

Motivation (0-5)	4
Soundness (0-5)	2
Novelty (0-5)	2
Technical Depth	Easy
Overal	**Borderline (with revision)**

References and Links

- Online:
  - https://link.springer.com/chapter/10.1007/978-3-319-70863-8_2
- Citation
  - Inan, E. and Dikenelli, O., 2017. Effect of enriched ontology structures on RDF embedding-based entity linking. In *Metadata and Semantic Research: 11th International Conference, MTSR 2017, Tallinn, Estonia, November 28–December 1, 2017, Proceedings 11* (pp. 15-24). Springer International Publishing.
- Cited by **5** publications

Thank You