Learning Class Definitions by Link Prediction (preliminary)[*]

First Author[1][0000−1111−2222−3333], Second Author[2,3][1111−2222−3333−4444], and Third Author[3][2222−−3333−4444−5555]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Abstract. The abstract should briefly summarize the contents of the paper in 15–250 words. Test citation [**Berners-Lee2001**]

Keywords: First keyword · Second keyword · Another keyword.

1 Research Questions

Are knowledge graph embeddings helpful in predicting entity classes?

2 Empirical Semantics

With the help of your tutor, provide a definition for Empirical Semantics that applies to the perspective that your team is taking for this problem. Probably, this will be one of the last things you will do...

3 Introduction

[1 Page]
Explain your perspective on the problem of Empirical Semantics. Give both the intuition and motivate, by relying on use cases and examples, why this perspective is important. Briefly describe what is the state of the art and how you're pushing it with your contribution. Also mention what data and methods you use in your work. Conclude by clearly stating what is your contribution.

4 Related Work

[1 Page]
List the main relevant work (a bullet list is ok) and for each of them write a paragraph describing (i) the key contribution of the related work, (ii) how your contribution relates/differentiate from it.

---

[*] Technical report of the task force 42 from ISWS 2022 led by Heiko Paulheim.

5   Resources

Our use-case focusses on DBpedia categories (why?). A category in DBpedia has a type of *skos:Concept*. There are more than 2 million categories in DBpedia. We manually selected 13 categories (motivation?) for our experiment. For every category, we manually selected predicate-object pairs that represent and distinguish a category. For example, the pair "author"-"Stephen_King" was chosen as an indicative of the category "Novels_by_Stephen_King". For six categories, more than one predicate-object pair was selected.

Four datasets were retrieved from DBpedia with SPARQL-queries:

1. Getting the number of subjects with the same predicate and object for every category. Criteria: (1) predicates are in the DBpedia ontology, (2) predicates not referring to Wikipedia (not containing the word "wiki" in their names). For example, there are 44 objects with predicate "author" and object "Stephen_King" in the category "Novels_by_Stephen_King". This predicate-object pair is the most frequent in the category and it is indicative for the category. The purpose of this dataset is to calculate the frequency of the predicate-object pairs in every category (ranking) and calculate reciprocal rank. The overview of the selected categories, their indicative object-predicate pairs and ranks is presented in Table 1.
2. Getting all triples for every category. Criteria: (1), (2), and (3) objects with URI containing "http://dbpedia.org/resource/" to filter out objects with literal values. This dataset is used for evaluation.
3. Getting all combinations of subjects and predicate-object-pairs in a category. Criteria: (1), (2), (3).
4. Counting the occurrences of the predicate-object pairs for every category in the whole DBpedia. This dataset is used in calculating TF-IDF.

The resulting datasets, SPARQL-queries, and code for pre-processing are documented and openly available on GitHub[4].

6   Proposed approach

[2 pages]
Describe your proposed method.

7   Evaluation and Results: Use case/Proof of concept - Experiments

[2 pages]
Show here that your proposed approach addresses your research questions or how you intend to show it. This can be done by either or both:

---

[4] https://github.com/AndreiNesterov/42

Category URI	Predicate	Object	N subjects	Rank	Reciprocal rank
Novels_by_Stephen_King	author	Stephen_King	44	1	1
Swedish_death_metal_musical_groups	genre	Death_metal	61	1	1
	hometown	Sweden	22	2	0.5
Red_Hot_Chili_Peppers_songs	artist	Red_Hot_Chili_Peppers	49	1	1
English_pop_pianists	genre	Pop_music	7	1	1
	birthPlace	England	2	6	0.166
	instrument	Piano	2	16	0.062
1990s_American_sitcoms	genre	Sitcom	273	1	1
Films_produced_by_Denzel_Washington	producer	Denzel_Washington	6	2	0.5
Hilltowns_in_Emilia-Romagna	region	Emilia-Romagna	2	1	1
	country	Italy	2	1	1
	province	Province_of_Forlì-Cesena	1	3	0.333
Languages_of_Namibia	spokenIn	Namibia	13	3	0.333
Italian_Renaissance_painters	movement	Italian_Renaissance	15	3	0.333
Philippine_television_talk_shows	genre	Talk_show	78	1	1
	country	Philippines	23	6	0.166
Scottish_clans	country	Scotland	1	N/A	0
Argentine_Primera_División_players	team	Argentina_national_football_team	656	5	0.2
Birds_of_Europe	class	Bird	1	N/A	0

Table 1. DBpedia categories selected for experiment

- Describing an experimental setting design, including research hypotheses, methods and metrics of measurements
- Describing a proof of concept/use case, based on real data, that support your claim

8   Discussion and Conclusions

[1 page]
Identify strengths and weaknesses of your proposal, discuss lessons learned: what are the key issues you have encountered or that you think should be taken into account to develop your proposal/experiments, and what are possible ways to address them.