Semantic Entity Enrichment by Leveraging Multilingual Descriptions for Link Prediction

(Position Paper)

Genet Asefa Gesese, Mehwish Alam, Harald Sack

DL4KG@ESWC
June 2020
What is Link Prediction (LP)?
Why Entity Descriptions for LP?

- **A Brief History of Time** is a 1925 novel written by American author F. Scott Fitzgerald that follows a cast of characters...
- **The Great Gatsby** hasDescription...
- **Stephen Hawking** isA writtenBy **Person** hasDescription...
- **F. Scott Fitzgerald** hasDescription...

OR?

... is a physicist, cosmologist, author and Director of Research at the Centre for Theoretical Cosmology...

... was an American novelist and short story writer, whose works are the paradigmatic writings of the Jazz Age...
Why Multilingual Entity Descriptions for LP?

Samuel Zachary Arkoff (12 June 1918 – 16 September 2001) was an American producer of B movies.[1]

Samuel Z. Arkoff est un producteur américain né le 12 juin 1918 à Fort Dodge, Iowa, et décédé le 16 septembre 2001 à Burbank, Californie.


Il est un des spécialistes de la série B (produisant par exemple Roger Corman), ce qui ne l'a pas empêché de produire des cinéastes comme Martin Scorsese ou Brian De Palma.

Samuel Z. Arkoff is an American producer born on June 12, 1918 in Fort Dodge, Iowa, and died September 16, 2001 in Burbank, California.

He founded in 1956, with James H. Nicholson, the production company American International Pictures (AIP).

He is one of the specialists in the B series (producing for example Roger Corman), which did not prevent him from producing filmmakers like Martin Scorsese or Brian De Palma.

From French to English using Google Translate
Related Work

- DKRL [1] and MKBE [2]
  ○ use CNN to encode entity descriptions.

- Jointly [3]
  ○ uses LSTM to encode entity descriptions.

- SSP [4]
  ○ adopts a topic model to generate a representation for an entity based on its description.

- LiteralE [5]
  ○ represents entity descriptions using a document embedding technique.

Problem:

→ None of these models utilize multilingual entity descriptions!
Problem Statement

How to combine complementary or additional information present in entity descriptions provided in multiple natural languages for the task of LP?
Potential Solutions

1. Applying Language Translators
   - A straightforward way to incorporate multilingual entity descriptions in the existing neural network encoder based KGE models, such as DKRL, MKBE, and Jointly.
   - First, all the descriptions must be converted into one language (English).
   - Then, the pre-trained embeddings of the words present in the descriptions would be passed as inputs to the encoder.

Main challenge:

→ The errors that occur during machine translation will be propagated to the encoder.
Potential Solutions

2. Using Multilingual Word Embeddings (Inspired By KDCoE [6])
   ○ KDCoE leverages a weakly aligned multilingual KG for entity alignment task using descriptions of entities.
   ○ Attentive Gated Recurrent Unit encoder (AGRU) is used, in KDCoE, to encode multilingual descriptions with multilingual word embeddings as inputs.

A CNN encoder adopted from DKRL [1]
Future Work

- Conducting detailed analysis on the nature and quality of the multilingual entity descriptions available in different KGs such as DBpedia and Wikidata.

- Evaluating the proposed potential solutions on different datasets.

- Adopting the solutions for other tasks such as triple classification and entity classification.
References


