Market and Technology Monitoring driven by Knowledge Graphs

Andreas Belger, Renato Budinich, Ralph Blum, Martin Zablocki, Roland Zimmermann

Motivation

The idea of this research project is to work on tools, which reveal lines of timely developments by analyzing a "stream" of publicly available information, usually issued on a daily, weekly or monthly basis in public domains. More specifically the focus is on timely monitoring of technologies readiness (or maturity). The project considers continuously information starting from 2018, which report on the **e-mobility domain** and retrieve information from those sources to answer the following sample questions:

A. *Which companies may constitute potential acquisition targets or sales leads in the e-mobility market?*
B. *In what stage of development are the existing technologies and which are emerging in the e-mobility market?*

Methods

We use an easy-to-understand model, which is expressive enough to capture the described aspects and reduce complexity to being able to interact with the Knowledge Graph. This approach ensures that on the one hand we are able to disambiguate relationships from different sources, which actually represent the same thing, and merge them by means of a domain-specific ontology. On the other hand, information is made unambiguous without losing the provenance of the information.

Technically, we continuously retrieve information from a list of online sources and annotate the texts with Named Entity Linking and Named Entity Recognition tools. Once the metadata, mentions and entities are stored in the RDF triple store, we use WordNet to define 3 classes of verbs (develop, test, order) and use them to detect *Facts* between concepts (entities or superclasses of these). For interacting and monitoring, we define questions as a set of Sparql queries, which are available through a REST API to a standard BI frontend.

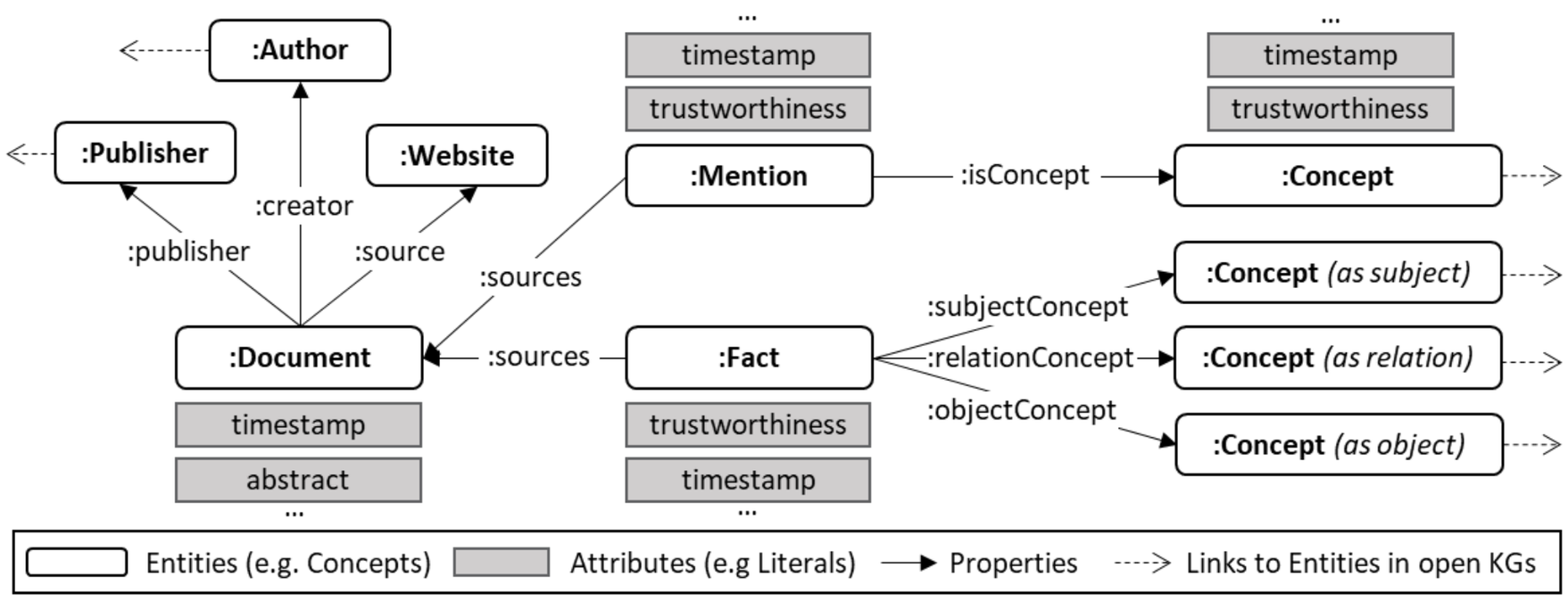

Fig. 1 Our simplified RDF model representing three main components: (1) input documents, (2) identified mentions, (3) retrieved facts.

Insights and Future Research

To answer question *A*, we analyzed data such as type and size of a company, which was provided by the disambiguated DBpedia concepts. The Semantic Web structure allowed us to analyze along multiple meaningful dimensions, e.g. find all companies in the same sector as any given (already recognized) company.

Regarding question *B*, we manually checked a small number of the produced facts and found false-positive rates of 15%, 39% and 14% for the three classes respectively. In order to do a deeper analysis where we could compute also false negatives, and due to the lack (to our knowledge) of a domain relevant dataset, we are in the process of manually annotating a random sample of our text corpus.

Open questions:

- Which methods from social networks analysis could be adapted in order to detect different types of node neighborhoods that could signal relevant features?
- How can we further integrate and exploit temporal aspects and dynamical changes?
- How can we define a semantic model which captures something like a "trend" as part of the graph and enables us to detect new and emerging ones?

Affiliations

Fraunhofer IIS
TECHNISCHE HOCHSCHULE NÜRNBERG GEORG SIMON OHM
Nuremberg Campus of Technology
trivadis makes IT easier.