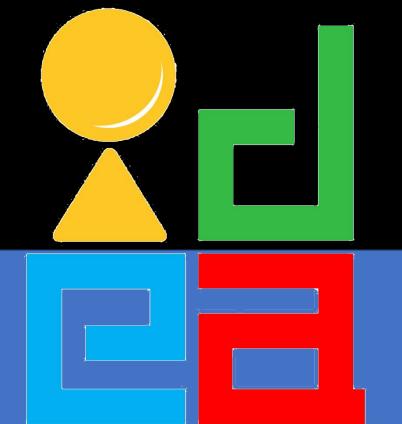


Learning Non-taxonomic Ontologies from Social Media Events



Isaac Martinez (imtz90b@gmail.com)

Intelligent Data Engineering and Applications Laboratory
National Tsing Hua University, Taiwan

Introduction and Motivation

A domain ontology represents a formal specification of a shared explicit knowledge related to a specific domain. Ontologies describe the relationships between objects, these relationships can be hierarchical(taxonomic) or not. Building and maintaining an ontology is a resource intensive task that requires human expertise.

Social Media streams contain not only temporal and geolocated data, but also information written by humans about a certain topic written in natural language.

**Contain not only temporal and geolocated data, but also information written by humans about a certain topic written in natural language.

The main motivation behind this work is to make this human knowledge available in a structured way, so that it enables further reutilization and exploration.

Objective

- To extract the group of Entities and non-taxonomic Relationships from a set of tweets related to an event that provide the most information gain possible about
- To merge related concepts into the same Entity, and merge related verbs into the same Relationship.
- To build an ontology from the set of semantic relationships extracted.

Framework Linguistic Analysis Preprocessing Twitter Data Lemmatization Remove Language Filtered by Spam Filtering Extraction Subject-Verb-Object triads Build Wall @realDonald | rump Want Graph building Assign weight based on temporal and frequency features Trump realDonaldTrum Wxg1vg1 Reduce Graph Establish weight centrality and degree thresholds Generate Ontology Merge equivalent entities based on similarity OWL format files Evaluation Merge related verb pairs based on node similarity

Tania Taekema @dhbmom1 · 17h

Humphrey Gervais @lifecaredotbz · Jun 13

Changed my mind #Trump SHOULD build wall BUT around whole country

lark all maps "Here be America. Stay away. Very dangerous." #Orlando

The #LGBTQ should be alarmed, #Muslims hate them and #Hillary wants to

open our borders to them! @donlemon @sallykohn #p2 #Orlando #CNN

Frequent Entities

Trump

@realDonaldTrump

-#Hillary

-wall

-...

Frecuent Verbs

-build

-wants

-...

Expected Outcome and Applications

- The expected output is an ontology containing the key entities related to an event and the relationships between them in a machine readable format.
- This output can be later utilized to classify news and links related to an event, to link similar events or any other activity that relies on this kind of knowledge.

