Entity-Level Event Impact Analytics
Govind Govind

To cite this version:
Govind Govind. Entity-Level Event Impact Analytics. WSTNET Web Science Summer School 2018 (WWSSS 2018), Jul 2018, Hannover, Germany. hal-02102815

HAL Id: hal-02102815
https://hal.archives-ouvertes.fr/hal-02102815
Submitted on 17 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Objective: Automatically Predict the Event Diffusion into Foreign Language Communities

Conceptual Approach
“Discovery of semantic connections to languages from named entities in the article”

ELEVATE Framework [1]
- Recursive exploration of *yago* relations
- Country centric:
  - isCitizenOf,
  - diedIn,
  - isLocatedIn,
  - isLeaderOf,
  - isPoliticianOf,
  - wasBornIn,
  - livesIn
- Organization centric:
  - owns,
  - created,
  - worksAt
- Linking of entities to languages

Breadth-first-search (BFS):
Stopping the exploration after the discovery of first language for a named entity

Depth-first-search (DFS):
Revealing all languages associated with a named entity exhaustively

Spread Prediction
Task: Pick the best candidates from all the scored languages

Adjusted Thresholding
- Threshold(θ) = average spread in the ground truth
- k-fold cross-validation
- Risk of picking the irrelevant languages

Multi-label Classification
- Output labels as the languages in event spread
- Candidate language scores as feature vectors
- Classifiers decide the spread

Other Research Works

ELEVATE-Live [3]
- Online news article virality prediction to countries
- Extension of ELEVATE framework
- Available at: https://elevate.greyc.fr

Semantic Fingerprinting [2]
- Fine-grained entity-level Web content classification
- Concise semantic representation of documents based on their entities

Short Biography

3rd year PhD student University of Caen, FranceAdvisor: Prof. Marc Spaniol

Research Interests
- Entity-level analytics
- Data aggregation via LOD
- Deep learning for noisy data analytics

Education
- Master degree in Maths and Computing from Indian Institute of Technology (IIT) Patna (India)
- Bachelor degree in computer science from Guru Jambheshwar University of Science and Technology Hisar (India)

Experimental Results

<table>
<thead>
<tr>
<th>Language</th>
<th>5-days</th>
<th>30-days</th>
<th>60-days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ru &quot;&quot;</td>
<td>0.8794</td>
<td>0.8704</td>
<td>0.7878</td>
</tr>
<tr>
<td>En &quot;&quot;</td>
<td>0.8794</td>
<td>0.8704</td>
<td>0.7878</td>
</tr>
</tbody>
</table>

Macrow average scores for the adjusted threshold based models (+PC: number of predictions)

Education

- Bachelor degree in computer science from Guru Jambheshwar University of Science and Technology Hisar (India)