EvenPers: Event-based Person Exploration and Correlation

Christian Kapp, Jannik Strötgen, and Michael Gertz
Database Systems Research Group, Heidelberg University

Motivation

Information extraction
• a lot of information only published in unstructured format → textual documents

Events
• happen at specific place and time
event = (space,time)
• events can be of different granularities
• can be associated with persons
Place, time & person information
• widely spread in text documents
• can be extracted and normalized

Huge document collections
• same events in different documents
• different persons can be associated with same events
→ persons can share events

Objectives

Extraction
• normalized events and persons
• person event profiles

Similarity Calculation
• similarities between events
• similarities between persons

Exploration
• explore a person’s events
• explore participants of an event
• identify similar persons
→ document-independent exploration

Person Event Profile Extraction

Selecting Documents (a)

Linguistic Preprocessing (b)

Person NER (f)

Geo Tagging (d)

Coreference Res. (g)

Temporal Tagging (c)

Person Normalization (h)

Event Extractor (e)

Person Event Profile Extractor (i)

DB (j)

Temporal and geo tagging:
• HeidelbergTime [1]
• Yahoo! Placemaker

Person NER and coreference
• StanfordNER, OpenNLP NER
• Arkref, Cherrypicker, Illinois package

Person normalization
• NER and coreference output results in person chains
• Wikipedia and JRCNames are checked for name variations for each chain
• normalized persons across documents

Person event profiles
• co-occurrences of events and persons
• \( \text{pep}(p) = \{e_1, p_1\}, \ldots, \{e_n, p_n\} \)

Existing similarities

event-centric document similarities [2]
– each document as set of events
– document similarity based on similarities of their events

Adaptation:
• event-centric person similarities
– each person as person event profile
– person similarity based on similarities of their events
• granularity-dependent event similarities
  \( e_1 = \text{(Magdeburg, 2013-03-13)} \)
  \( e_2 = \text{(Germany, 2013-03-13)} \)
  \( e_3 = \text{(Germany, 2013)} \)
  \( \sim(e_1, e_1) > \sim(e_2, e_2) > \sim(e_3, e_3) \)
  \( \sim(e_1, e_2) > \sim(e_1, e_3) \)
• the more similar events in \( \text{pep}(p_1) \) and \( \text{pep}(p_2) \), the higher \( \sim(p_1, p_2) \)

Event-centric exploration

• starting with a person
• explore her (most important) events
• explore (most) similar persons
• explore their (shared) events

References


Contact information

Jannik Strötgen
Heidelberg University
stroetgen@uni-hd.de
http://dbs.ifi.uni-heidelberg.de/