BOTTARI: Location based Social Media Analysis with Semantic Web

Emanuele Della Valle
Joint work with:

CEFRIEL: Irene Celino, Daniele Dell’Aglio, Marco Balduini
SALTLUX: Tony Lee, Seonho Kim
SIEMENS: Volker Tresp, Yi Huang
Watch this first :-)  

Bottari - a LarKC application  

http://www.youtube.com/watch?v=c1FmZUz5BOo
What have you seen?

- An augmented reality application for personalized recommendation of restaurants in Seoul
Yet another tripadvisor®?

- Yes and no!
- Same use case, more “democratic”
- We do “reality mining” by listening to the social media
Architecture

SPARQL → Query Rewriter

RDF2Matrix Plug-in → SUNS Plug-in

SLD Plug-in → SOR Invoker

Query Evaluator → out

PULL: Query Initiated

HTTP

Social Media Crawler and Sentiment Miner → Streaming Linked Data Server

Social Media Crawler and Sentiment Miner

Streaming Linked Data Server

SOR geospatial KB

PUSH: Data Initiated

androjena

PUSH: Data Initiated

SOR Invoker
Sentiment Mining

Micropost message

Morphologically Analyzable?

Yes

Rule based Analysis
Auto generated rules

No

SVMs
Syllable Kernel

Learned documents

Sentiment of the tweet

• Precision tests:
  – Auto-generated rules ≈ 70%
  – Manually-coded rules ≈ 90%
  – Syllable kernel ≈ 50~60%

• Our target > 85%
SOR - Geo-Spatial KB

twd:Tweet
  twd:messageID(xsd:string)
  twd:messageTime Stamp(xsd:string)

  twd:talksAbout
    geo:NamedPlace

  twd:following
  twd:reply
  twd:reply
  twd:talksAboutPositively
  twd:talksAboutNeutrally
  twd:talksAboutNegatively

geo:SpatialThing

sioc:creator

sioc:content(xsd:string)

twd:TwitterUser
  twd:screenName(xsd: string)
  twd:discuss
  twd:talksAboutNegatively

sioc:UserAccount
  sioc:id(xsd:string)
  geo:NamedPlace

geo:NamedPlace

geo:SpatialThing

twd:following

twd: follower

geo: NamedPlace

twd: talks About

geo: Spatial Thing

geo: Named Place

twd: talks About Negatively

geo: Named Place

geo: Spatial Thing
C-SPARQL and Streaming Linked Data Server

---

**Re-stream**

- +1 for POI

**Analyze**

- Count +1 for POI
  - [1 DAY]
  - [1 DAY]
  - [1 DAY]
  - [7 GRAPHs]
  - [31 GRAPHs]
  - [7 DAYS]
  - [31 DAYS]

**Publish**

- Most Liked POIs
  - [1 GRAPH]
  - [1 GRAPH]
  - [7 GRAPHs]
  - [31 GRAPHs]

**Visualize**

- Top-10 POIs
  - Weekly Plot
  - Monthly Plot
  - Weekly Map
  - Monthly Map

---

**LEGEND**

- Re-stream from database
- C-SPARQL Query
- Windower
- List View
- Plot Line View
- Heatmap View

---

26.10.2011 - SW Challenge 2011, ISWC 2011, Bonn, Germany
A machine learning framework for inductive materialization

- Detects interesting data patterns
- Predics RDF-triples
  - i.e., which restaurant a user will tweet positively about

Characteristics

- Capability to deal with sparse, high-dimensional and incomplete data
- Multivariate latent space based approach
- Modularized approach for easily integrating contextual information
WHERE {

?poi a ns:NamedPlace ;
    ns:name ?name ;
    geo:lat ?lat ;
    geo:long ?long .
FILTER (f:within_distance(37.5, 126.9, ?lat, ?long, 200))
FILTER (f:dest_point_viewing(37.5, 126.9, ?lat, ?long, 90, 200))

   WITH PROBABILITY ?prob
   ENSURE PROBABILITY [0.5..1] }

}
ORDER BY DESC(?numPos), ?prob,
    f:distance(37.5, 126.9, ?lat, ?long)
LIMIT 10
androjena

**Probabilistic** part of the query to get personalized recommendations (the “for me” button in BOTTARI)

**Streaming** part of the query to get trends in users' sentiment (the “emerging” button in BOTTARI)

**Geo-Spatial** part of the query to get POIs closer to user location

Input user query is split

Results of the different computations are joined

PULL: Query Initiated

PUSH: Data Initiated
Evaluation - Efficacy

- SUNS + C-SPARQL
- for me (SUNS)
- emerging (C-SPARQL)
- knnItem
- random
Evaluation - Efficiency

Hardware: 2.66 GHz Intel Core 2 Duo with 8 GB RAM
Evaluation – Scalability

Query Latency (sec) vs. Number of concurrent users
Conclusions

• End-user application
• Attractive and functional interface
• Real-world dynamic data
• Fully based on Semantic Web technologies
  – RDF as common data format between heterogenous components
  – SPARQL as query language
• Rigorously evaluated
  – Effective
  – High throughput for handling dynamic data
  – Scalable in number of concurrent users
• Commercial Potential
Any question?

Emanuele Della Valle
Joint work with:

CEFRIEL: Irene Celino, Daniele Dell’Aglio, Marco Balduini
SALTLUX: Tony Lee, Seonho Kim
SIEMENS: Volker Tresp, Yi Huang