



# **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>

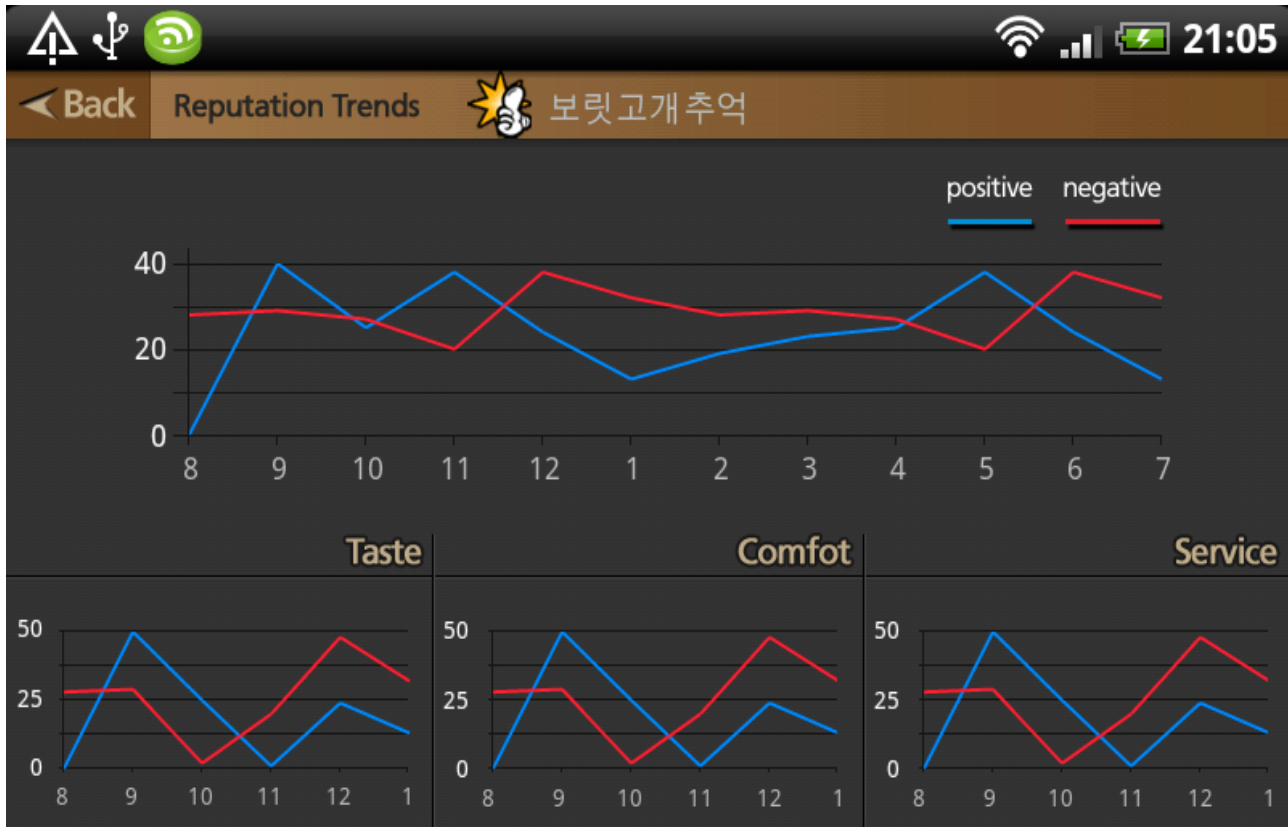
# What have you seen?

- An augmented reality application for personalized recommendation of restaurants in Seoul

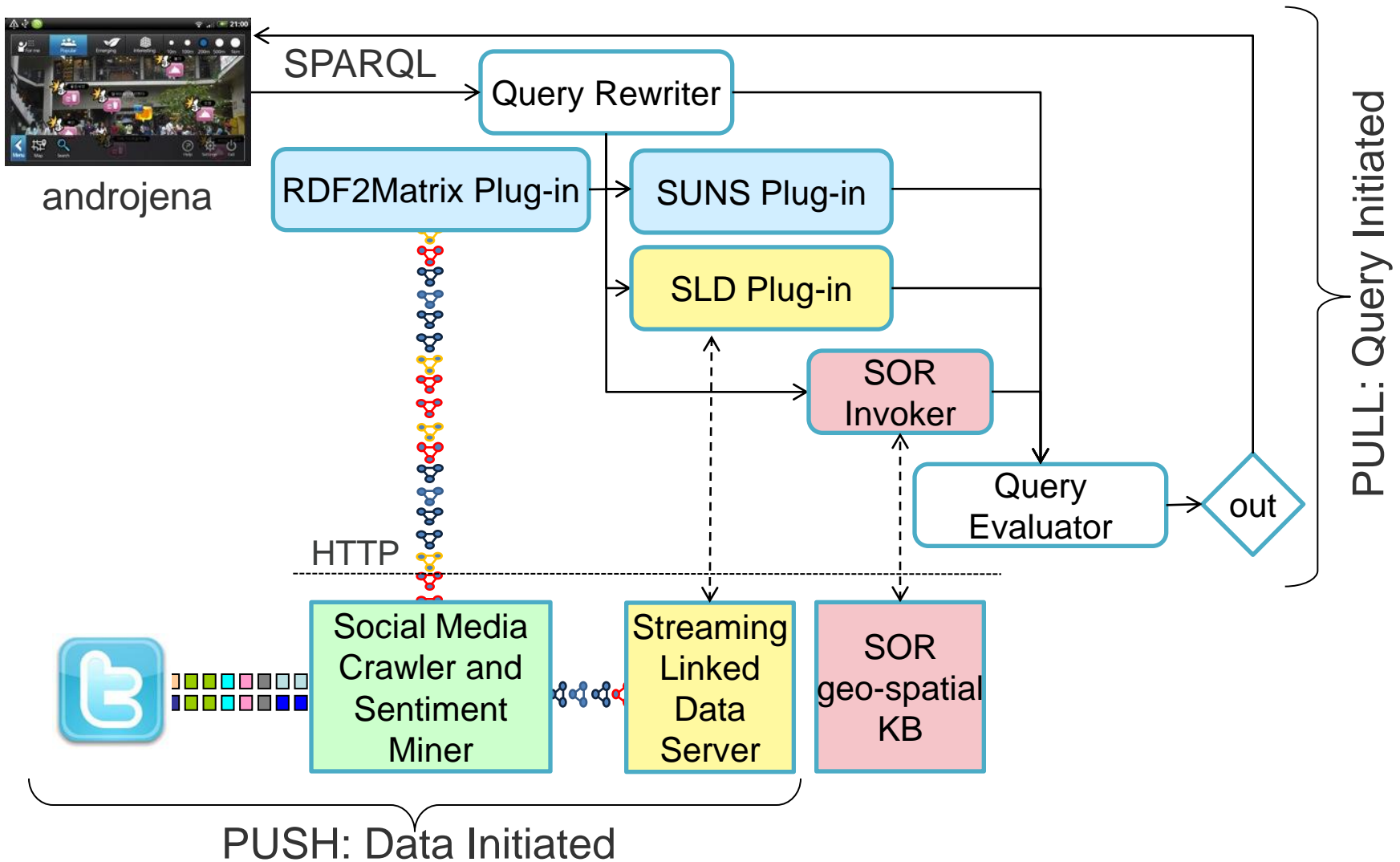


# Yet another tripadvisor® ?

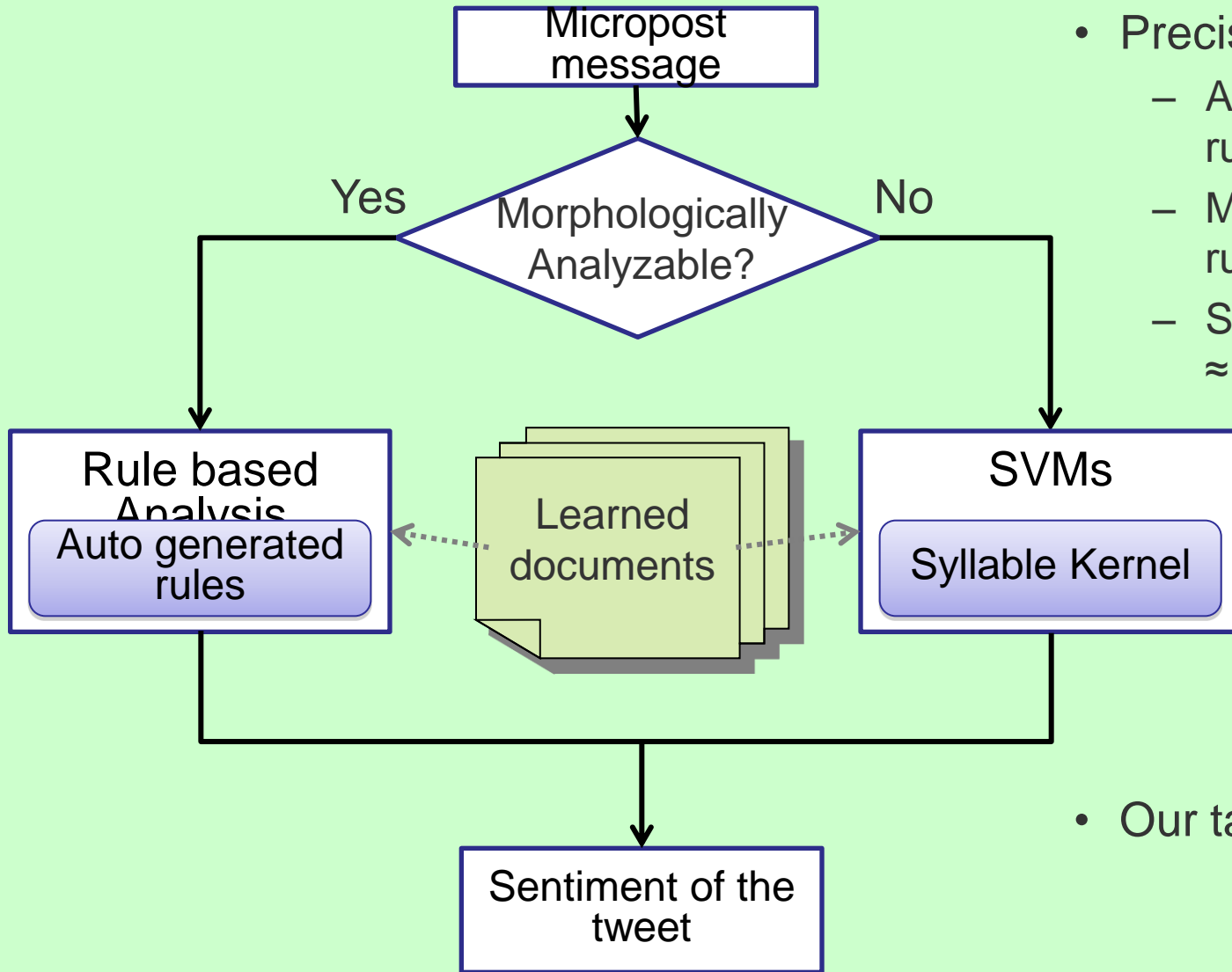
- Yes and no!
- Same use case, more “democratic”
- We do “reality mining” by listening to the social media



# Architecture



# Sentiment Mining

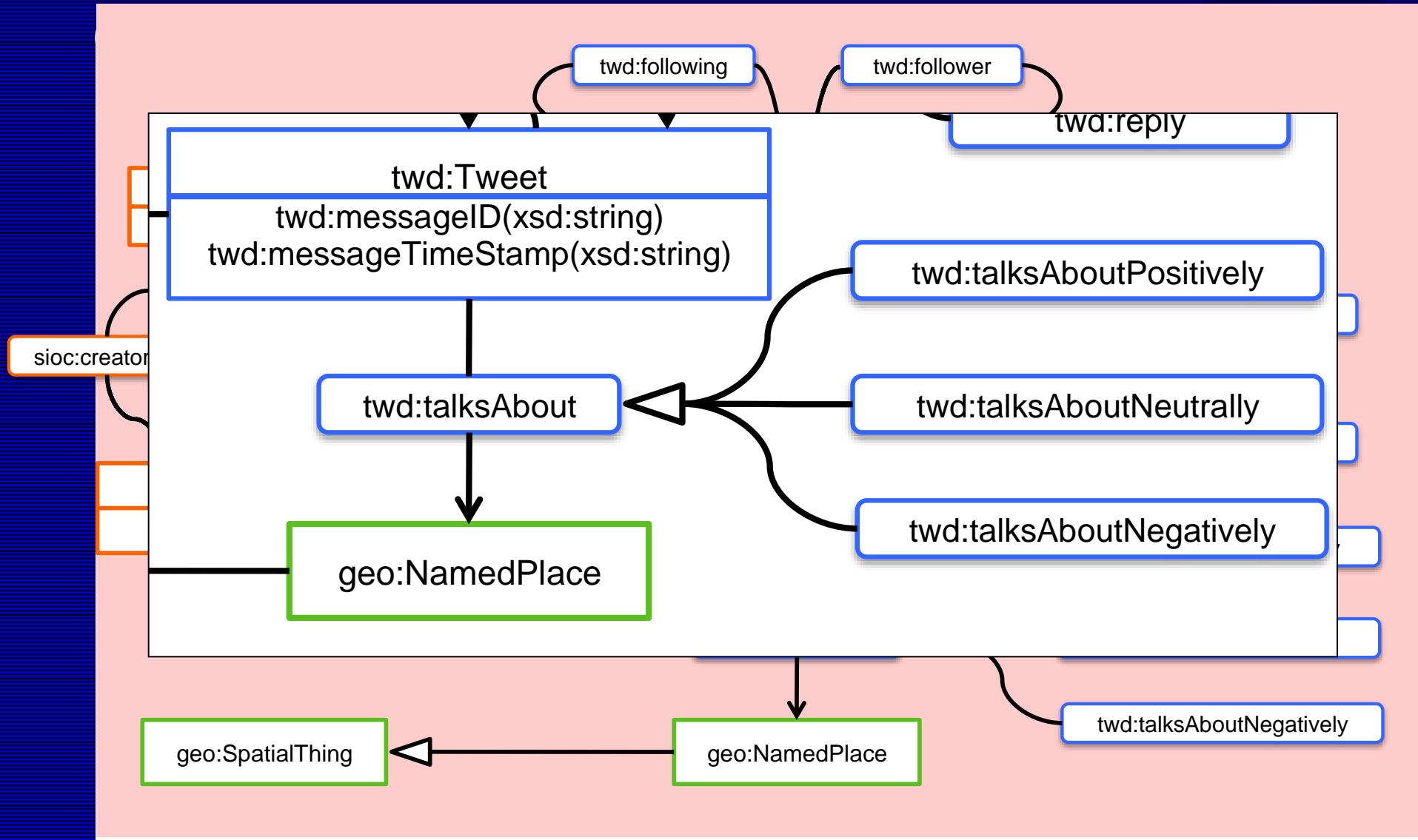


- Precision tests:
  - Auto-generated rules  $\approx 70\%$
  - Manually-coded rules  $\approx 90\%$
  - Syllable kernel  $\approx 50\sim 60\%$

- Our target  $> 85\%$

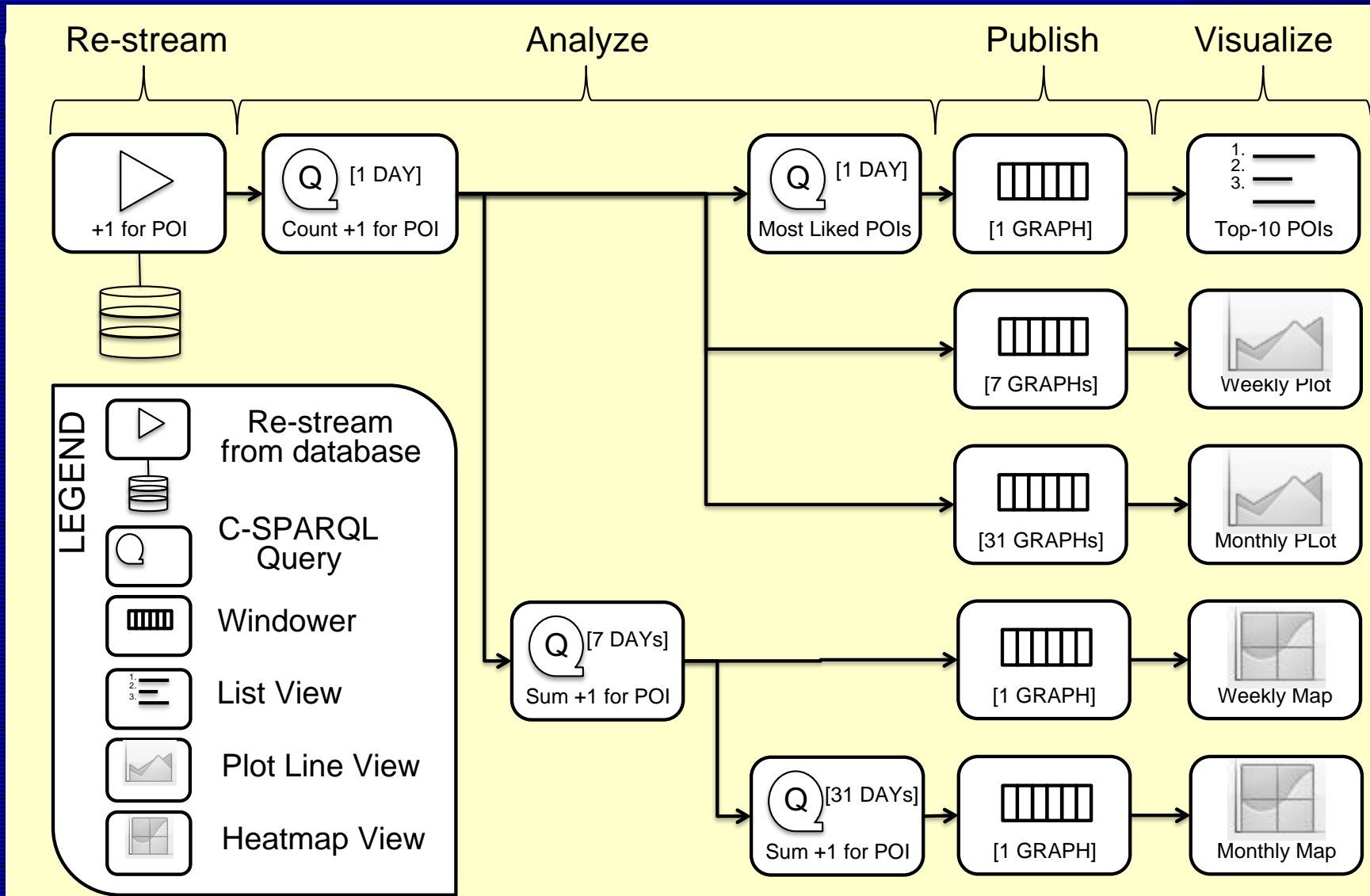


# SOR - Geo-Spatial KB





# C-SPARQL and Streaming Linked Data Server



# SUNS (Statistical Unit Node Sets)

- A machine learning framework for **inductive materialization**
  - Detects interesting data patterns
  - Predicts 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

# Query Processing

```
SELECT DISTINCT ?poi ?name ?lat ?long ?numPos ?prob
WHERE {
```

```
  ?poi a ns:NamedPlace ;
        ns:name ?name ;
        geo:lat ?lat ;
        geo:long ?long .
```

**GEO-SPATIAL**

```
  FILTER (f:within_distance(37.5, 126.9, ?lat, ?long, 200))
```

```
  FILTER (f:dest_point_viewing(37.5, 126.9, ?lat, ?long, 90, 200))
```

```
  { :someUser sioc:creator_of ?tweet .
```

**PROBABILISTIC**

```
    ?tweet twd:talksAboutPositively ?poi .
```

```
    WITH PROBABILITY ?prob
```

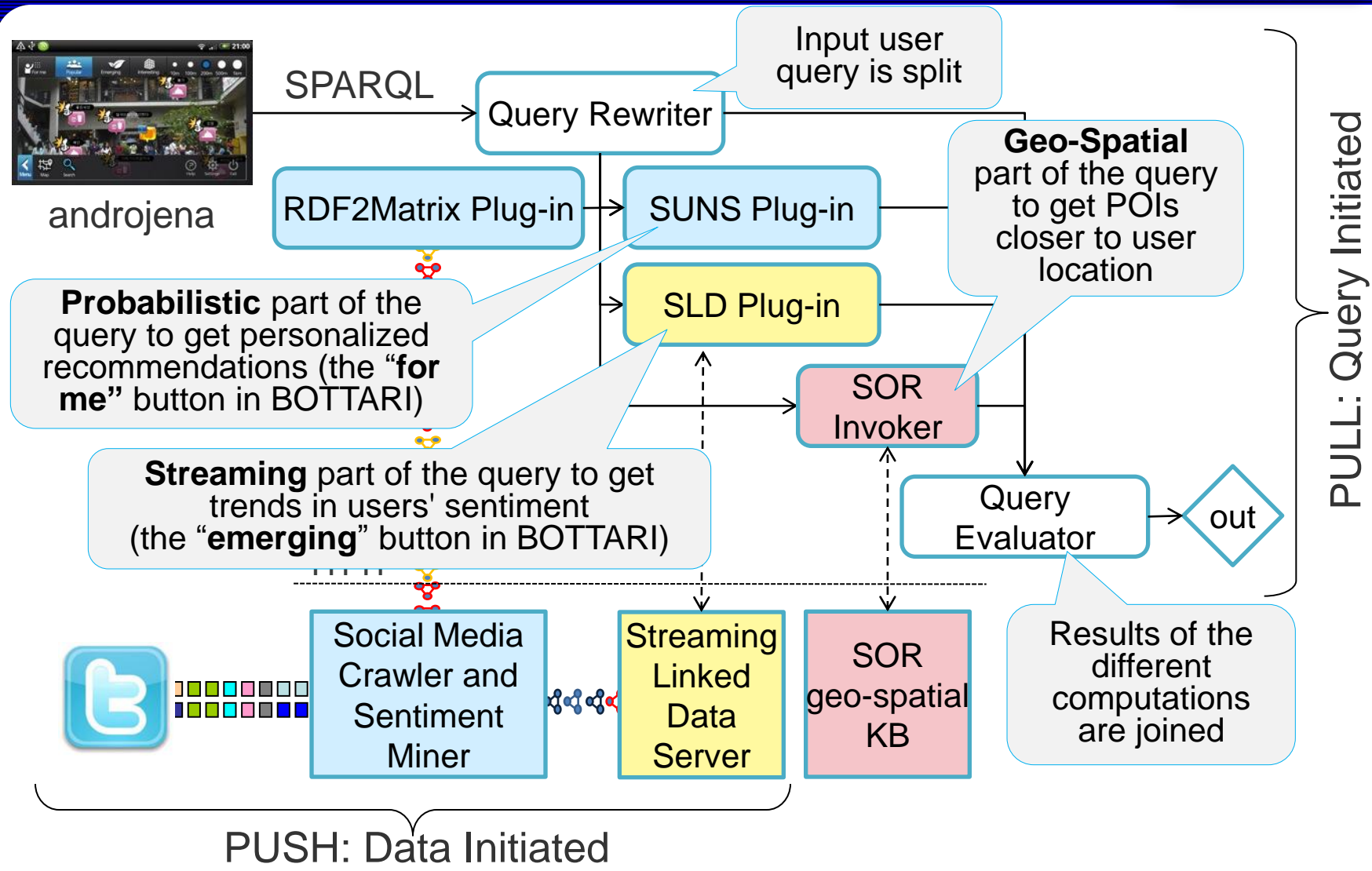
```
    ENSURE PROBABILITY [0.5..1) }
  ?poi twd:numberOfPositiveTweets ?numPos .
```

**STREAMING**

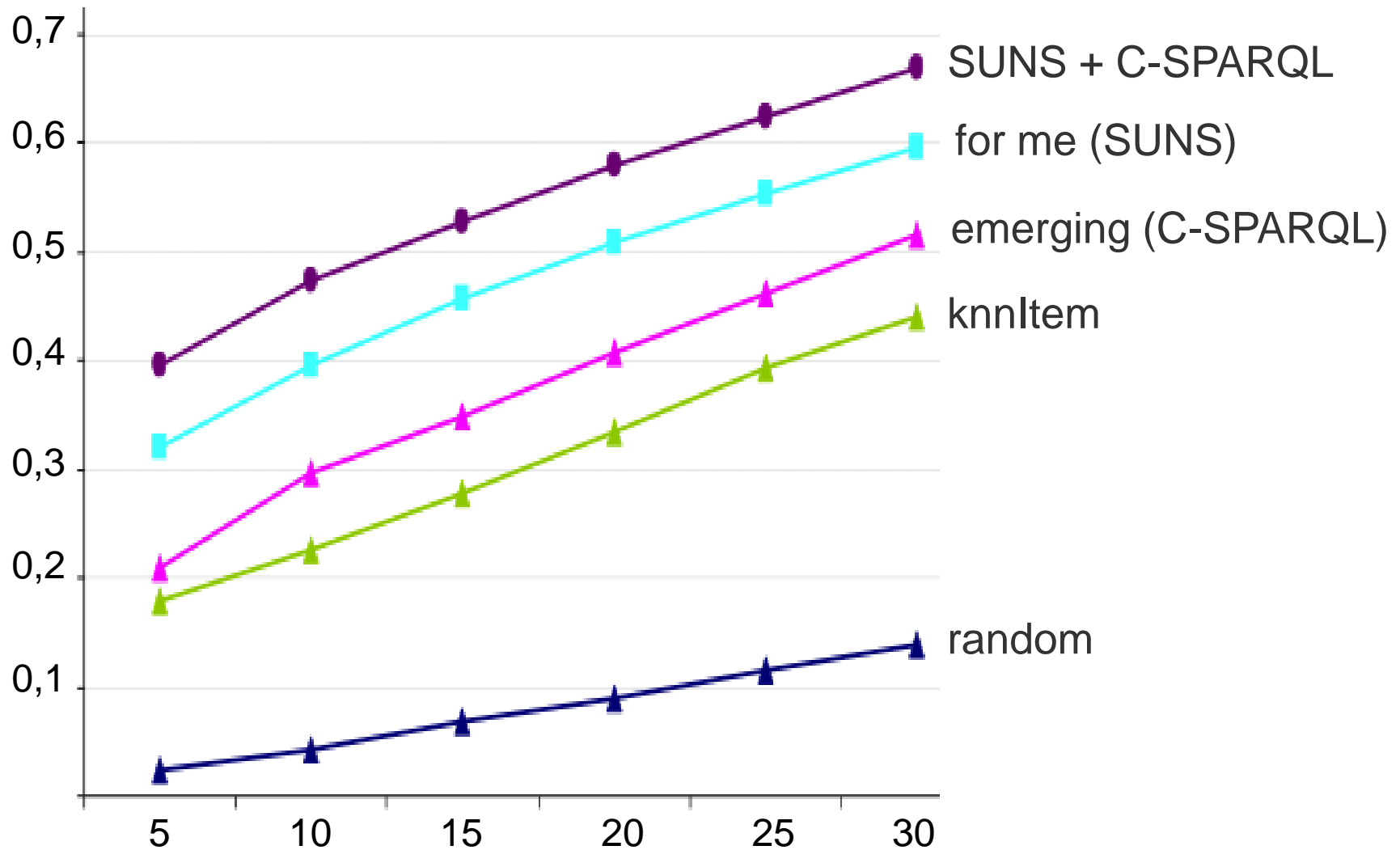
```
}
ORDER BY DESC(?numPos), ?prob,
           f:distance(37.5, 126.9, ?lat, ?long)
```

```
LIMIT 10
```

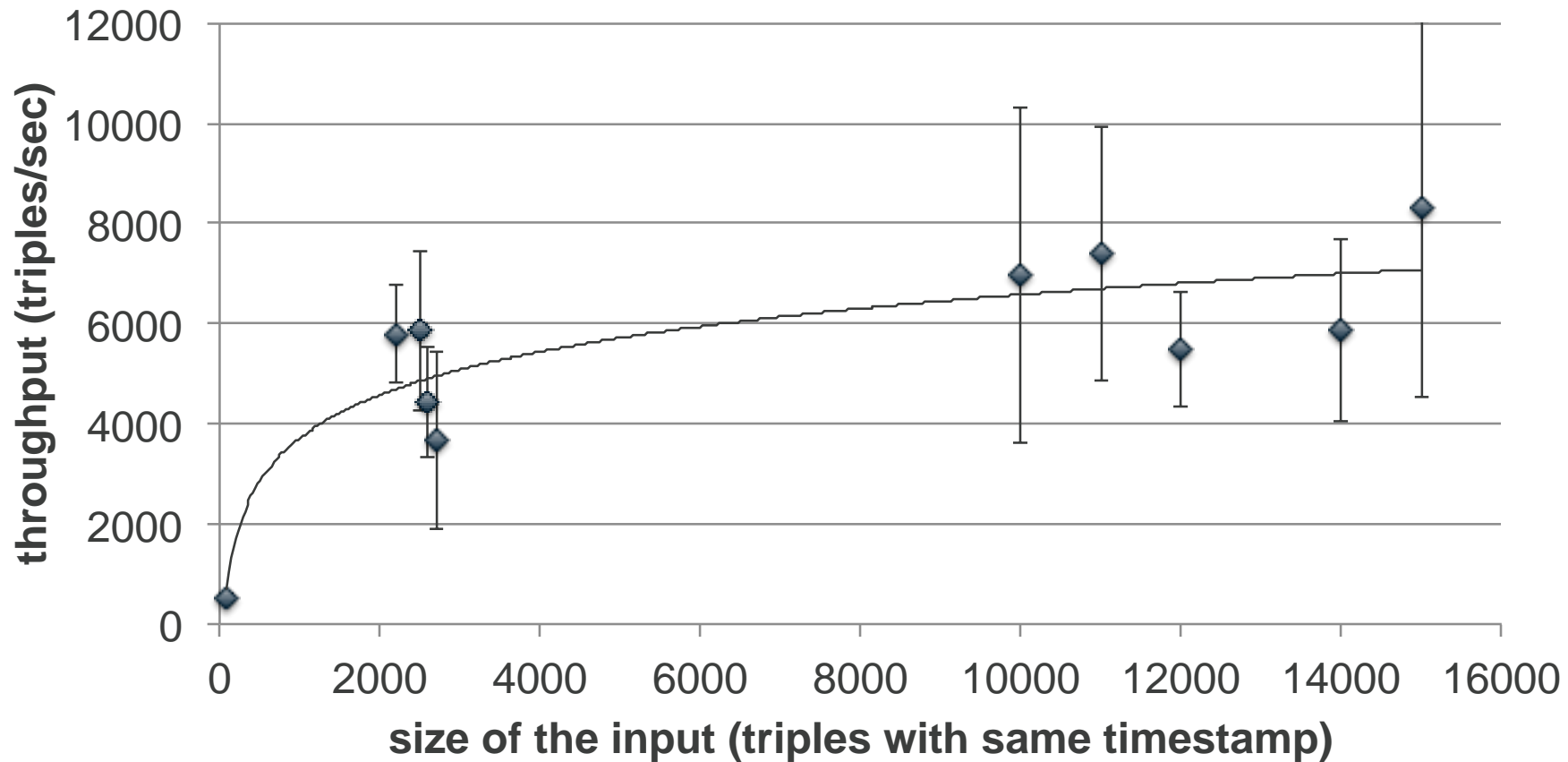
# LarKC At Work



# Evaluation - Efficacy

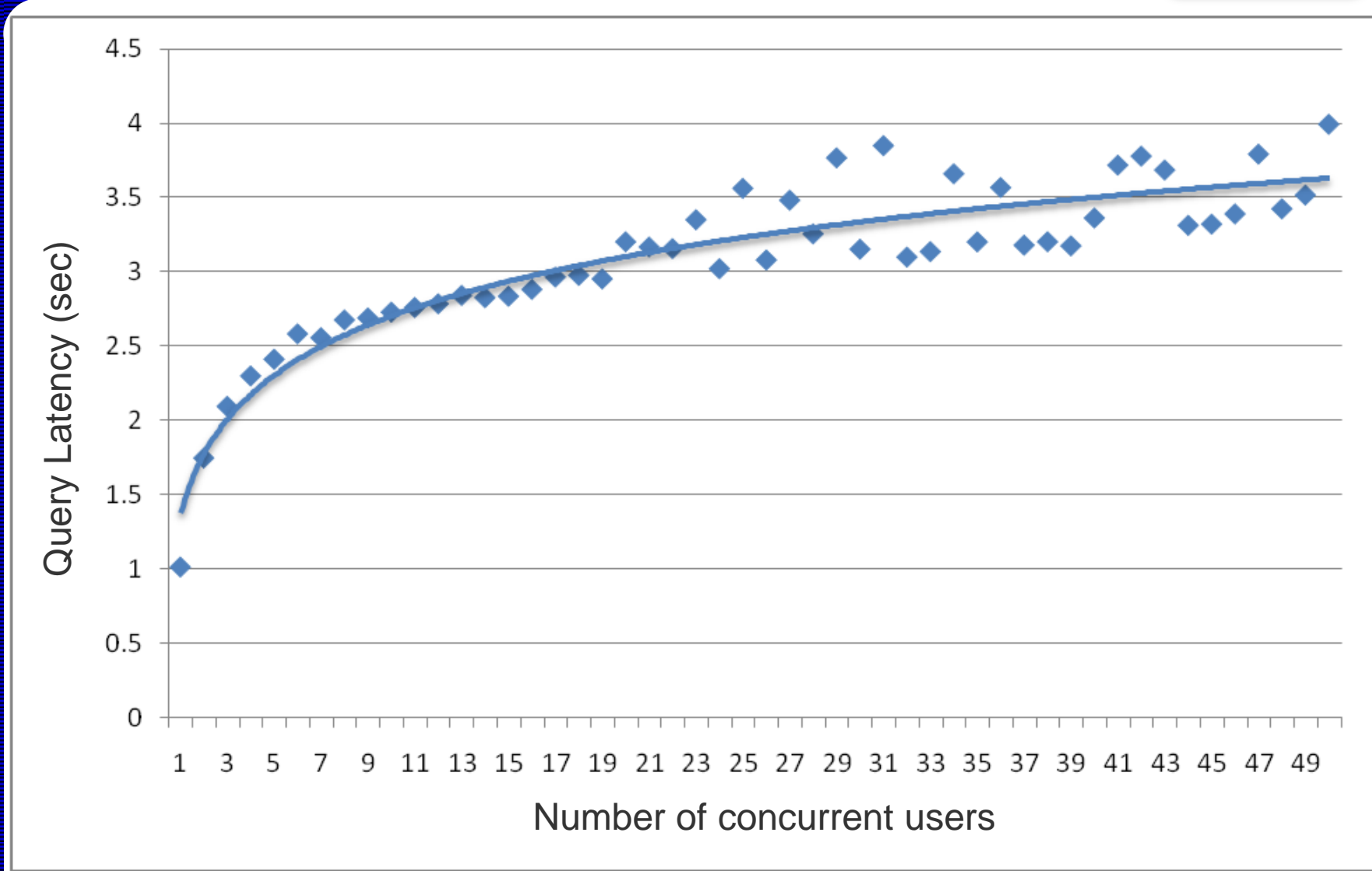


# Evaluation - Efficiency



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

# Evaluation – Scalability



# 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





# ENJOY KOREA WITH LARKC

**Any question?**

Emanuele Della Valle

Joint work with:

**CEFRIEL:** Irene Celino, Daniele Dell'Aglio, Marco Balduini

**SALTUX:** Tony Lee, Seonho Kim

**SIEMENS:** Volker Tresp, Yi Huang

