

# Discovering and Ranking Semantic Associations over a Large RDF Metabase



30<sup>th</sup> International Conference on Very Large Data Bases Toronto, Canada Aug 31 - Sep 3, 2004

Chris Halaschek, Boanerges Aleman-Meza,  
I. Budak Arpinar, Amit P. Sheth  
LSDIS Lab, Computer Science, University of Georgia



## Motivation

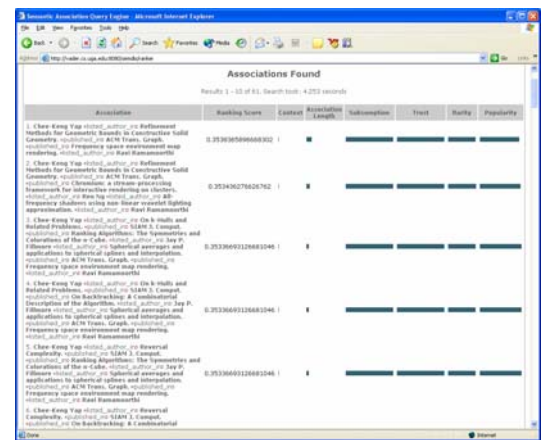
1. Query between "Hubwoo" (company) and "SONERI" (bank) results in 1,160 semantic associations (SWETO testbed)
2. Cannot expect users to sift through resulting associations
3. Results must be presented to users in a relevant fashion...need ranking

## Ranking Demo



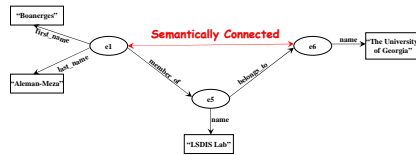
## Ranked Semantic Associations

(Entities: Chee-Keng Yap ←→ Ravi Ramamoorthi)



## Background

### Semantic Associations



### Selecting entities of interest



## Approach

### Association Rank Function

$$\text{Score} = k_1 \times \text{Context} + k_2 \times \text{Subsumption} + k_3 \times \text{Trust} + k_4 \times \text{Rarity} + k_5 \times \text{Popularity} + k_6 \times \text{Association Length}$$

(for flexibility,  $k_i$  values are assigned by user)

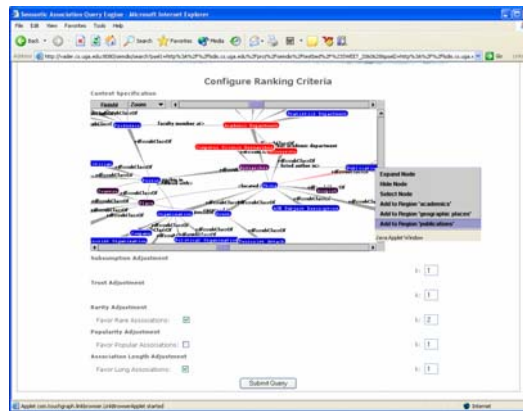
### Semantic criteria

- Context (classes/relations of interest)
- Subsumption
- Trust

### Statistical criteria

- Rarity
- Popularity
- Association Length

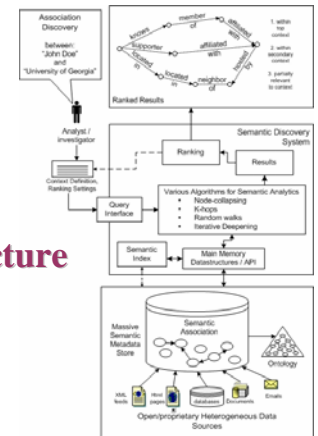
### Configuration of Ranking Criteria



### Acknowledgements

- SemDis Team – LSDIS Lab
- NSF-ITR-IDM Award #0325464, 'SemDis: Discovering Complex Relationships in the Semantic Web'
- NSF-ITR-IDM Award #0219649, 'Semantic Association Identification and Knowledge Discovery for National Security Applications'

## SemDis System Architecture



### SWETO Test-bed (RDF Metabase)

