



LSDIS

Large Scale Distributed Information Systems



University of Georgia
Computer Science Department

Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection

Boanerges Aleman-Meza¹, Meenakshi Nagarajan¹,
Cartic Ramakrishnan¹, Li Ding², Pranam Kolari²,
Amit P. Sheth¹, I. Budak Arpinar¹, Anupam Joshi², Tim Finin²

¹LSDIS lab

Computer Science
University of Georgia, USA

²Department of Computer Science and
Electrical Engineering

University of Maryland, Baltimore
County, USA



UMBC
AN HONOR UNIVERSITY IN MARYLAND



World Wide Web 2006 Conference
May 23-27, Edinburgh, Scotland, UK

This work is funded by NSF-ITR-IDM Award#0325464 titled "SemDIS: Discovering Complex Relationships in the Semantic Web" and partially by ARDA



Outline

- Application scenario: Conflict of Interest
- Dataset: FOAF Social Networks + DBLP Collaborative Network
- Describe experiences on building this type of Semantic Web Application



Conflict of Interest (COI)

- Situation(s) that may bias a decision
- Why it is important to detect COI?
 - for transparency in circumstances such as contract allocation, IPOs, corporate law, and peer-review of scientific research papers or proposals
- How to detect Conflict of Interest?
 - connecting the dots



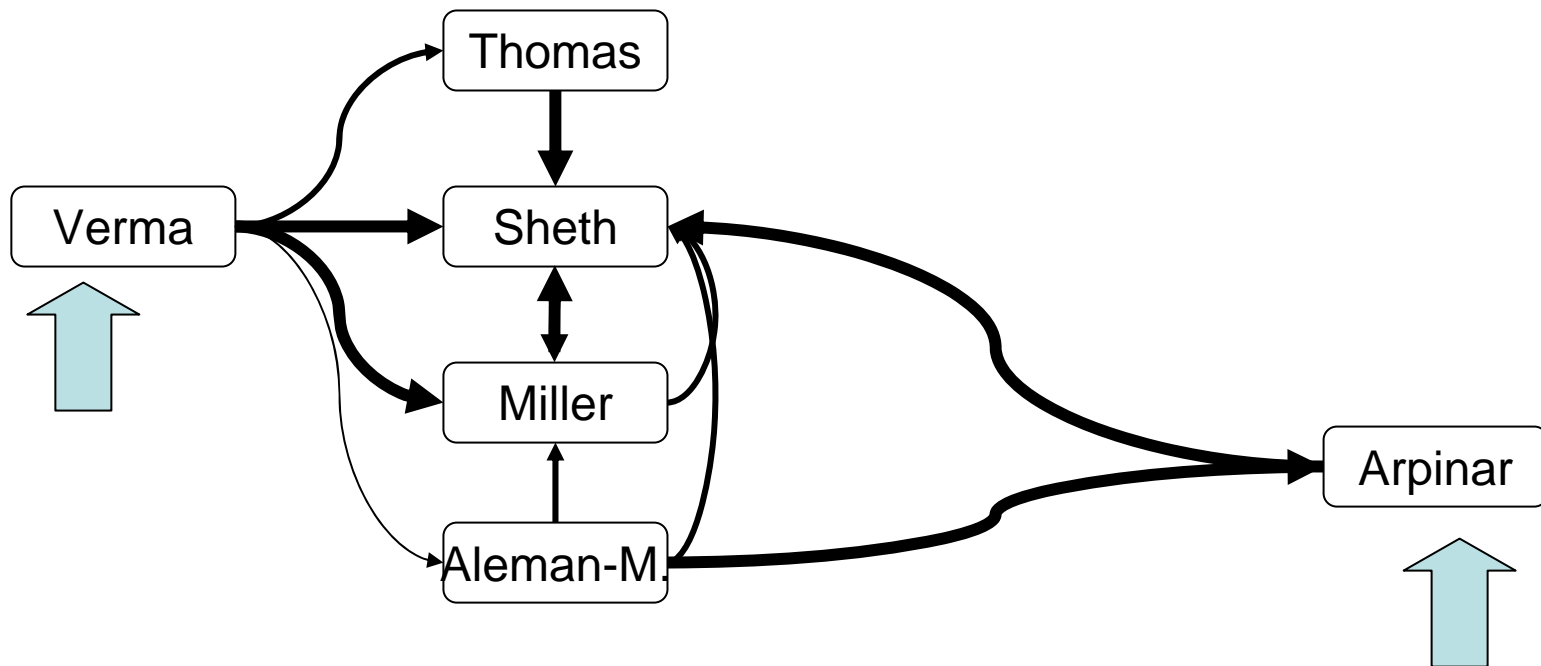
Scenario for COI Detection

- Peer-Review: assignment of papers with the least potential COI
 - Our scenario is restricted to detecting COI only (not paper assignment)
- Current conference management systems:
 - Program Committee declares possible COI
 - Automatic detection by (syntactic) matching of email or names, but it fails in some cases
 - i.e., Halaschek \leftrightarrow Halaschek-Wiener



Conflict of Interest

- Should Arpinar review Verma's paper?





Social Networks

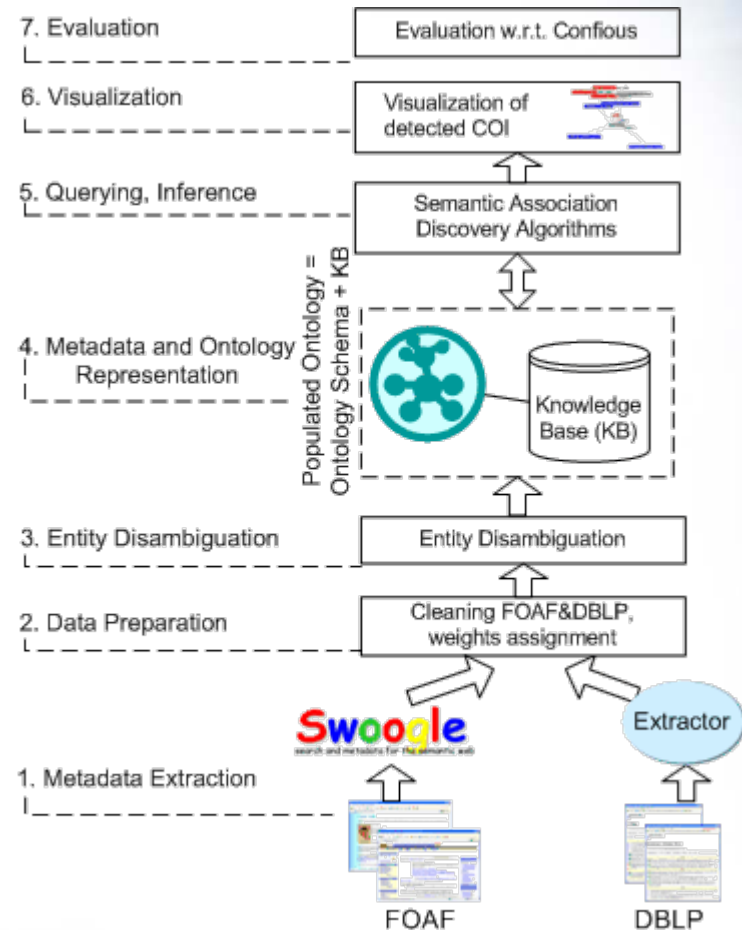
- Facilitate use case for detection of COI
 - But, data is typically not openly available
 - Example: LinkedIn.com for IT professionals
- Our Pick: public, real-world data
 - FOAF, Friend of a Friend
 - DBLP bibliography
 - underlying collaboration network
 - Covering traditional and semantic web data



Our Experiences: Multi-step Process

Building Semantic Web Applications involves a multi-step process consisting of:

1. Obtaining high-quality data
2. Data preparation
3. Metadata and ontology representation
4. Querying / inference techniques
5. Visualization
6. Evaluation





Our Experiences: Multi-step Process

Building Semantic Web Applications
requires:

1. Obtaining high-quality data
 - DBLP, FOAF data



FOAF – Friend of a Friend

- Representative of Semantic Web data
- Our FOAF dataset was collected using Swoogle (swoogle.umbc.edu)
 - Started from 207K Person entities (49K files)
 - After some data cleaning: 66K person entities
 - After additional filtering, total number of Person entities used: 21K
 - i.e., keep all 'edu/ac'





DBLP (dblp.uni-trier.de)

- Bibliography database of CS publications
 - Representative of (semi-)structured data
 - We focused on 38K (out of over 400K authors)
 - authors in Semantic Web area
 - arguably more likely to have a FOAF profile
- DBLP has an underlying collaboration network
 - co-authorship relationships



Combined Dataset of FOAF+DBLP

- 37K people from DBLP
- 21K people from FOAF
- 300K relationships between entities

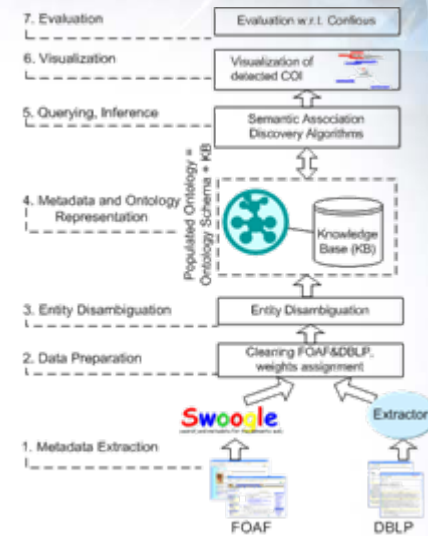


Our Experiences: Multi-step Process

Building Semantic Web Applications requires:

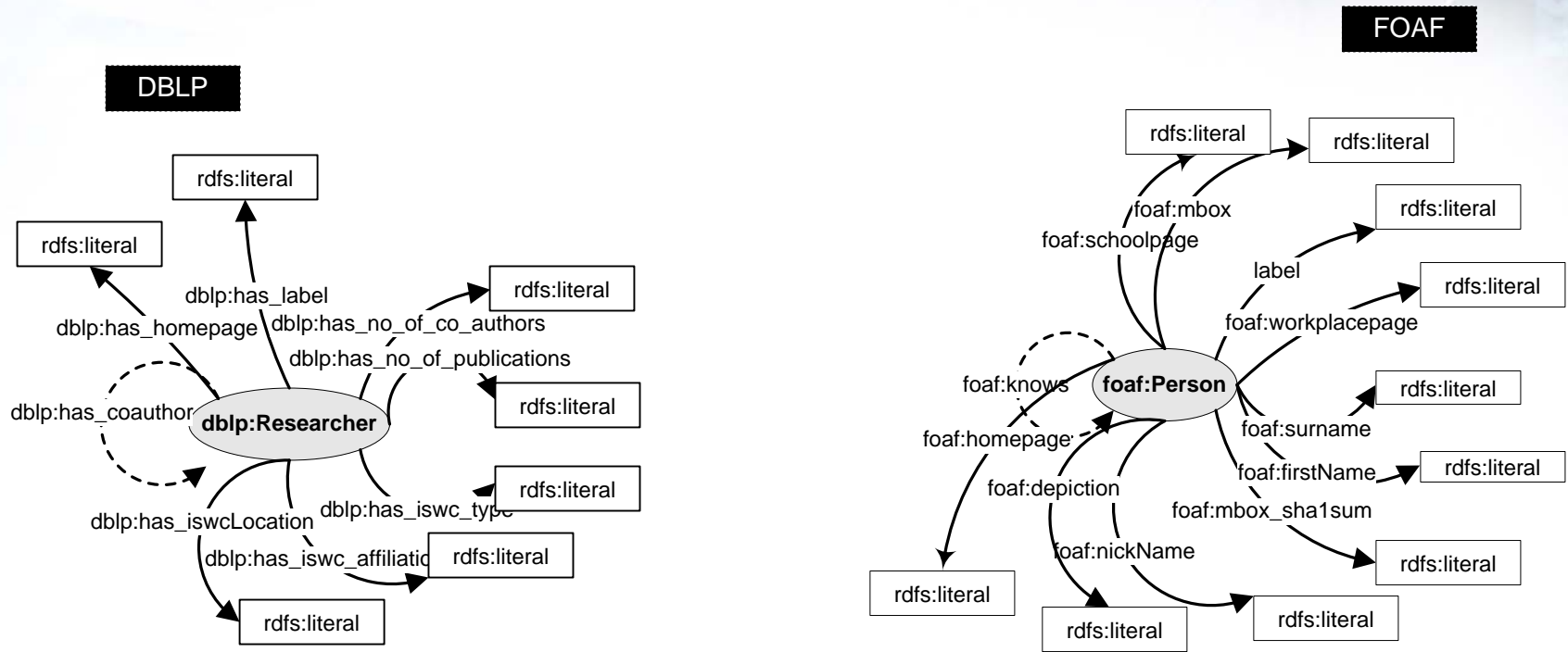
2. Data preparation

- Our goal: Merging person entities that appear both in DBLP and FOAF





Person Entities from two Sources



- Goal: harness the value of relationships across both datasets
 - Requires merging/fusing of entities

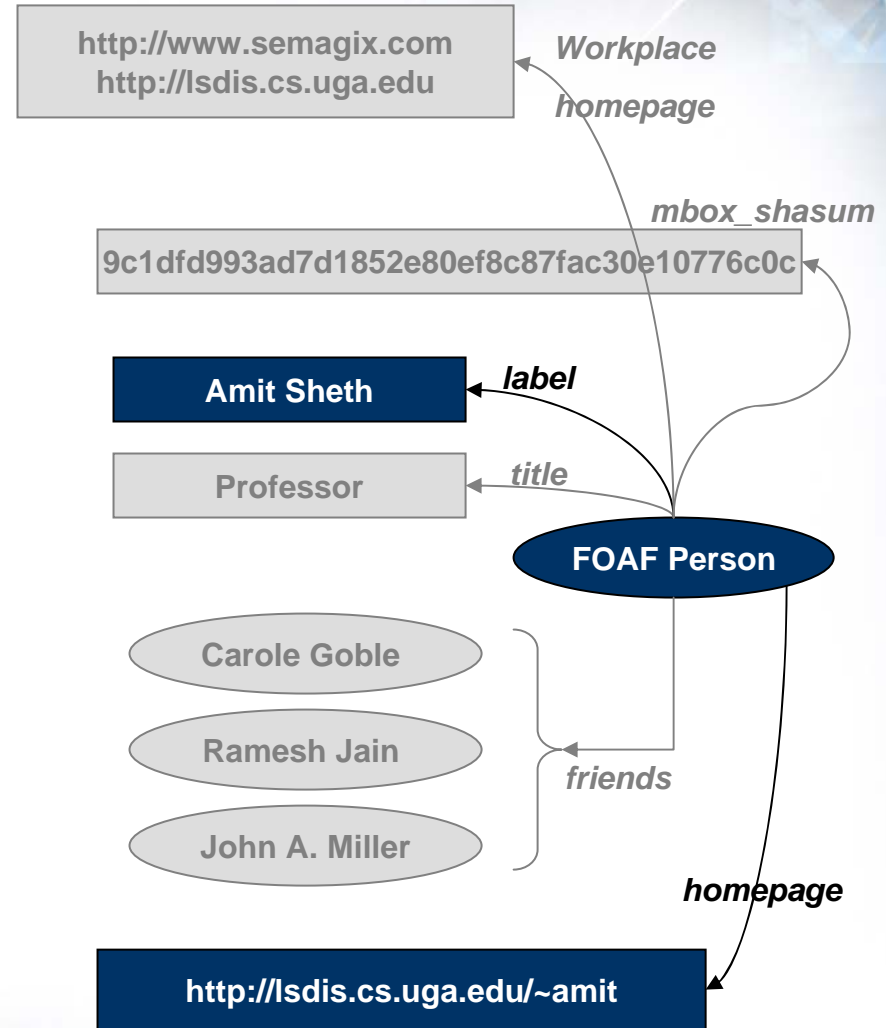
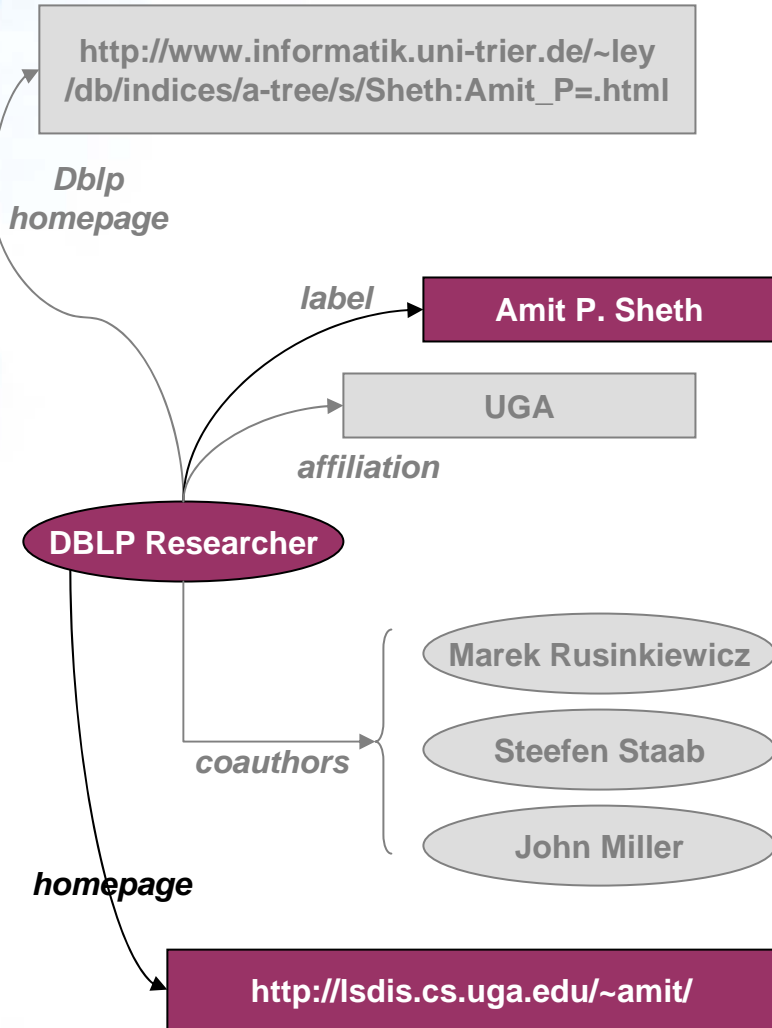


Merging Person Entities

- We adapted a recent method for entity reconciliation
 - *Dong et al. SIGMOD 2005*
- Relationships between entities are used for disambiguation
 - Presupposition: some coauthors also appear listed as (foaf) friends
 - With specific relationship weights
- Propagation of disambiguation results

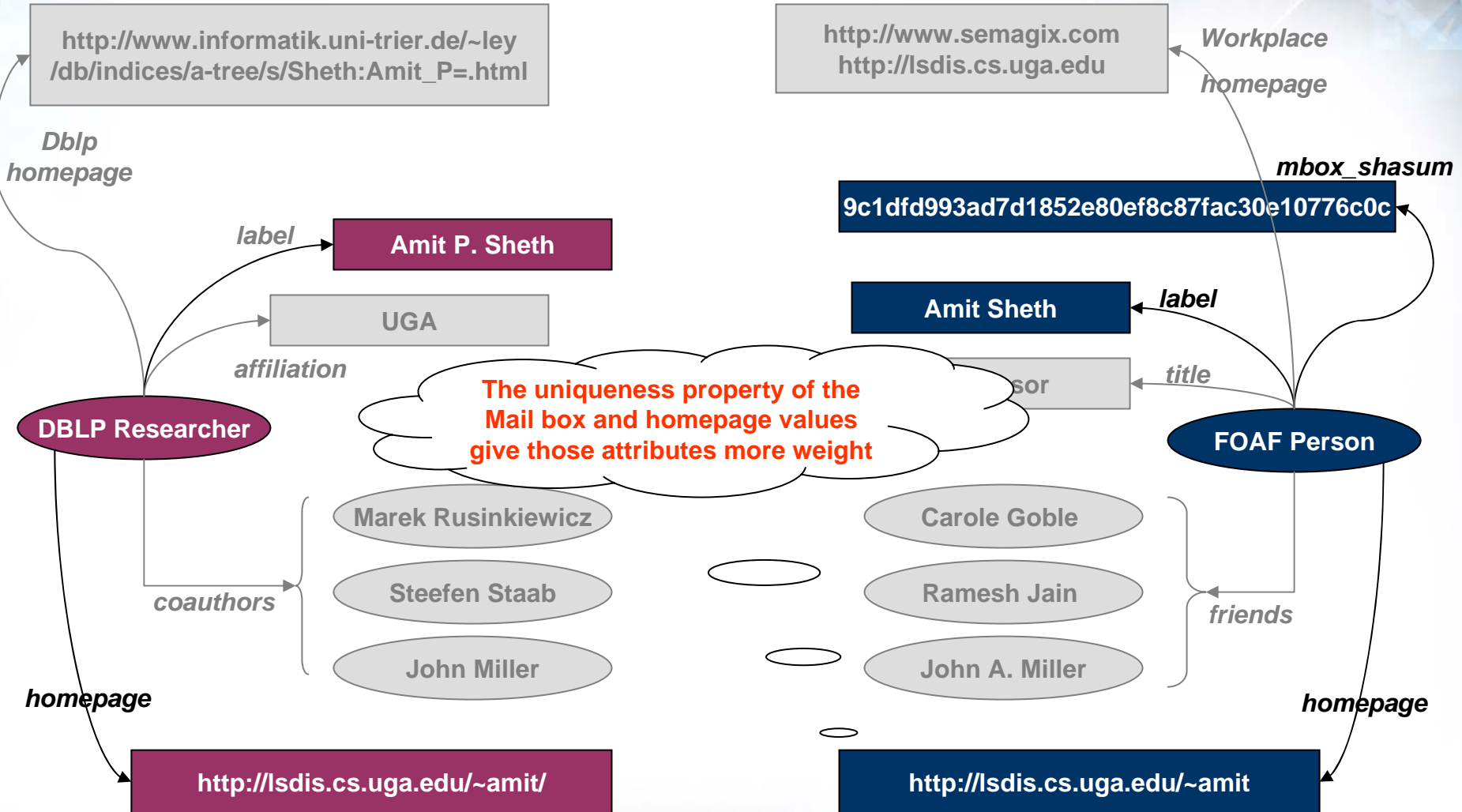


Syntactic matches



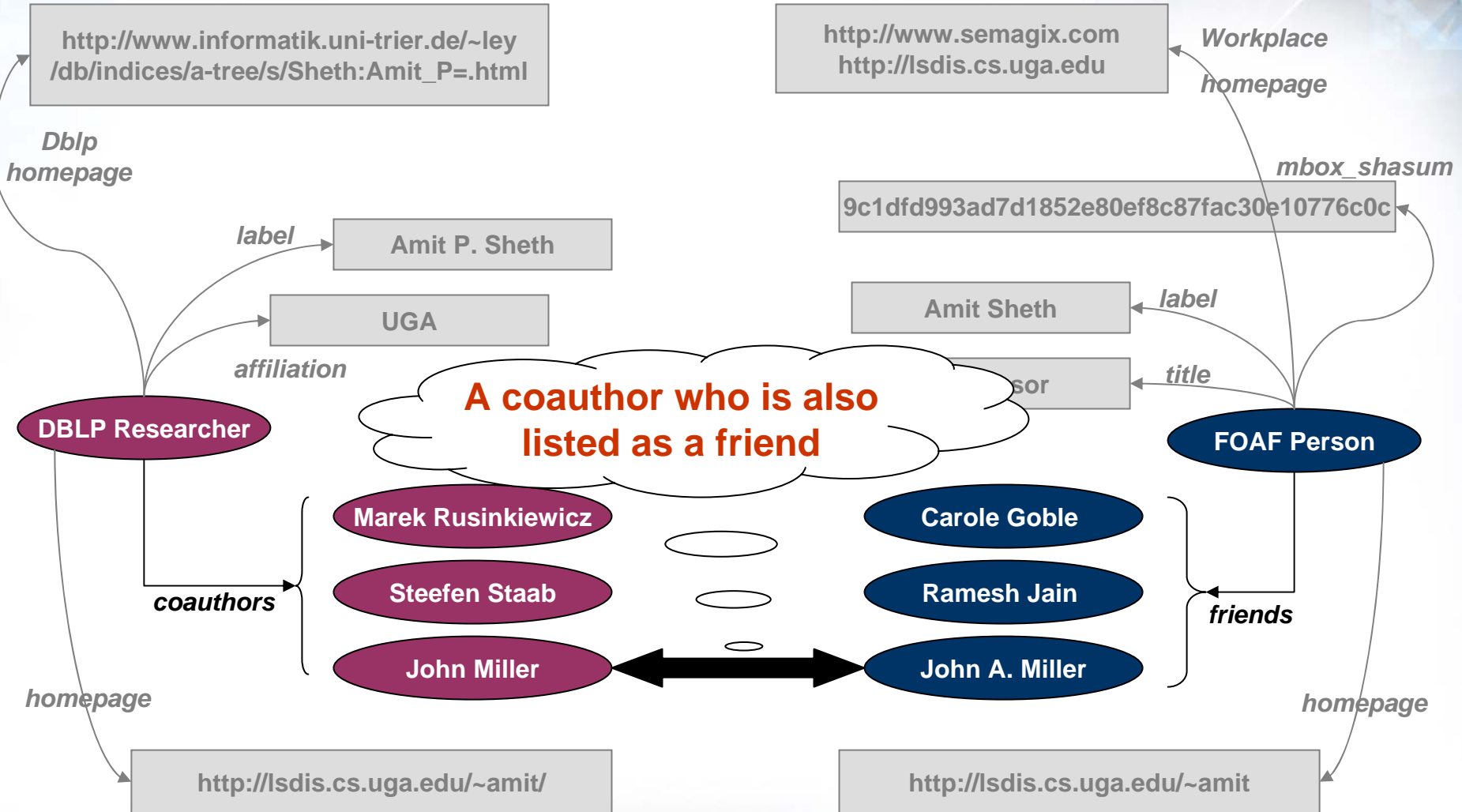


... with Attribute Weights





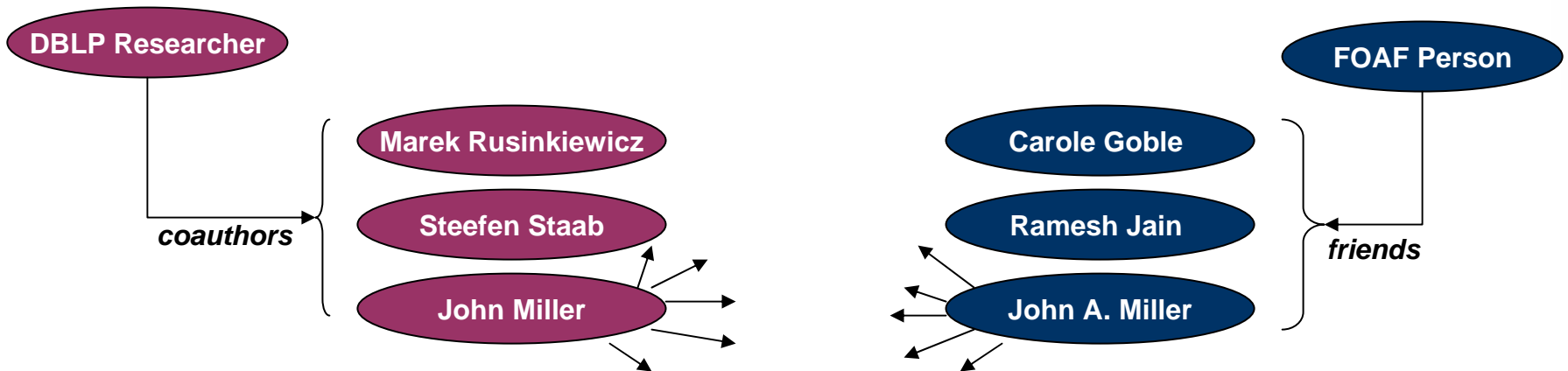
Relationships with other Entities





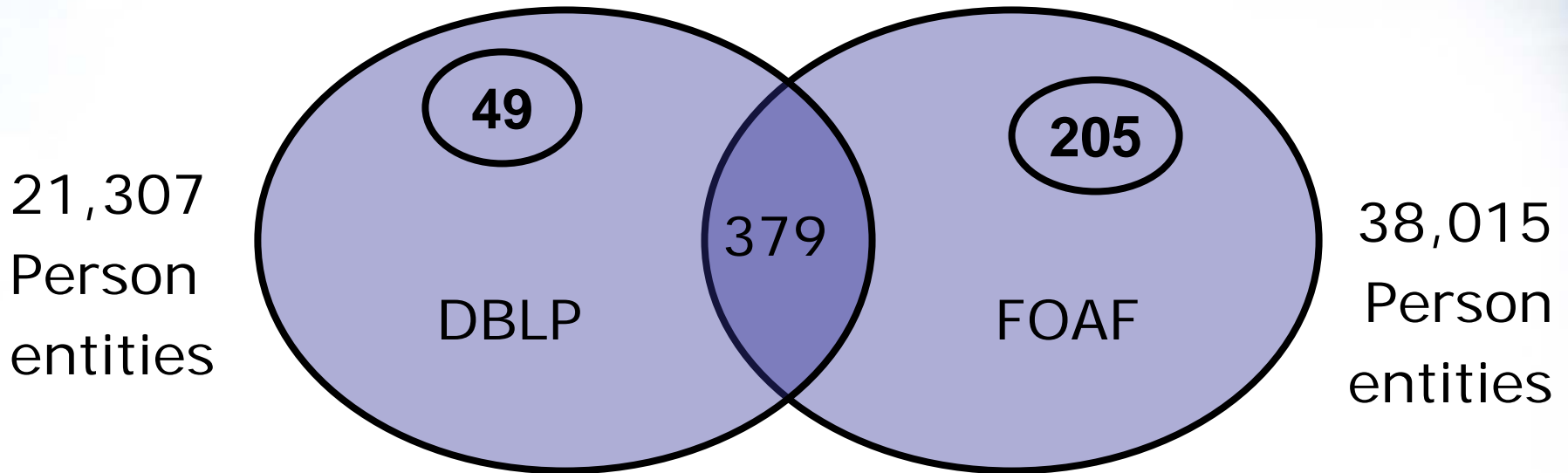
Propagating Disambiguation Decisions

- If **John Miller** and **John A. Miller** are found to be the same entity, there is more support for reconciliation of the entities Amit P. Sheth and Amit Sheth
 - based on the presupposition that some coauthors can also be listed as (foaf) friends





Results of Disambiguation Process



Number of entity pairs compared: 42,433

Number of reconciled entity pairs: 633

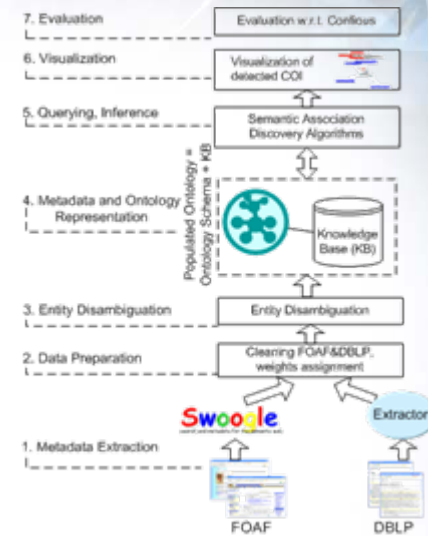
(a *sameAs* relationship was established)



Our Experiences: Multi-step Process

Building Semantic Web Applications requires:

3. Metadata and ontology representation
(How to represent the data)





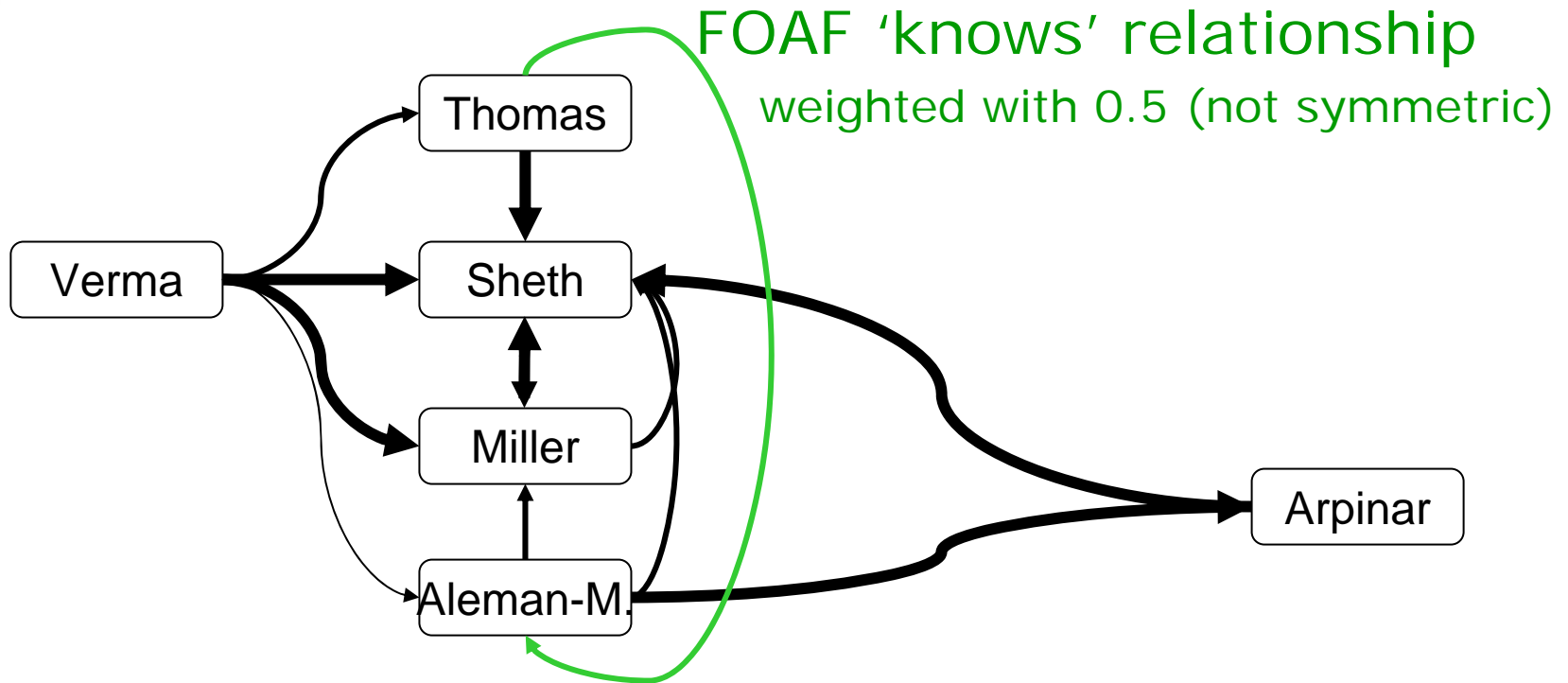
Assigning weights to relationships

- Weights represent collaboration strength
- Two types of relationships (in our dataset)
 - ‘knows’ in FOAF (directed)
 - ‘co-author’ in DBLP (bidirectional)
 - Anna → co-author → Bob
 - Bob → co-author → Anna



Assigning weights to relationships

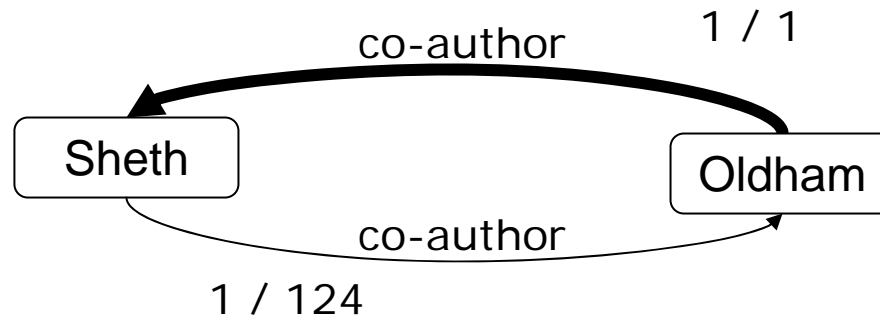
- Weight assignment for FOAF *knows*





Assigning weights to relationships

- Weight assignment for co-author (DBLP)
 $\# \text{co-authored-publications} / \# \text{publications}$



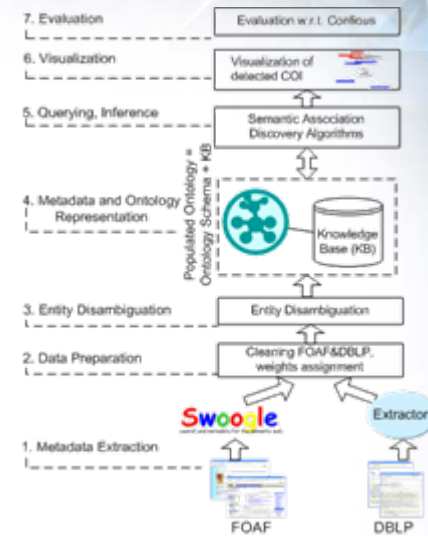
- The weights of relationships were represented using *Reification*



Our Experiences: Multi-step Process

Building Semantic Web Applications requires:

4. Querying and inference techniques



Top of the most active researchers (which builds upon research in Semantic Web Services)



Semantic Analytics for COI Detection

- Semantic Analytics:
 - Go beyond text analytics
 - Exploiting semantics of data (“A. Joshi” is a Person)
 - Allow higher-level abstraction/processing
 - Beyond lexical and structural analysis
 - Explicit semantics allow analytical processing
 - such as *semantic-association* discovery/querying



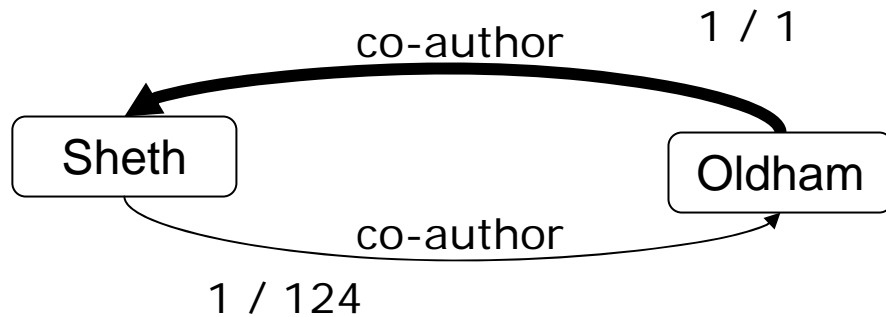
COI - Connecting the dots

- Query all paths between Persons A, B
 - using ρ operator: semantic associations query
 - Anyanwu & Sheth, WWW'2003
 - Only paths of up to length 3 are considered
- Analytics on paths discovered between A, B
 - Goal: Measure Level of Conflict of Interest
 - Trivial Case: 'Definite' Conflict of Interest
 - Otherwise: High, Medium, Low 'potential' COI
 - Depending on direct or indirect relationships



Case 1: A and B are Directly Related

- Path length 1
 - COI Level depends on weight of relationships

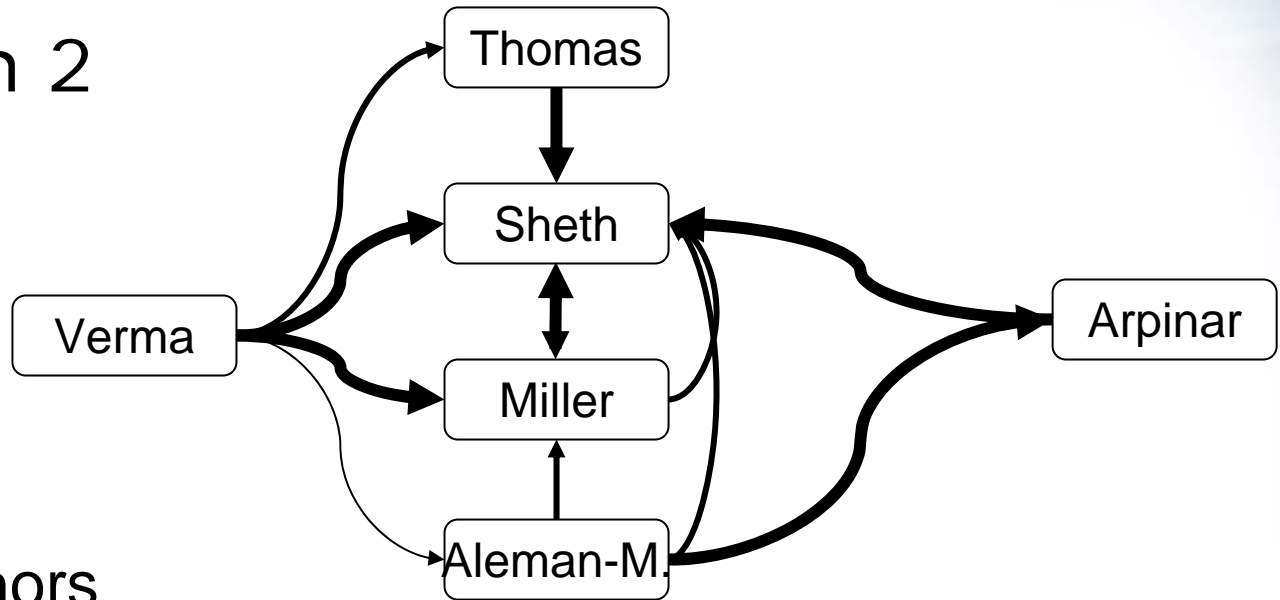


--	--	--



Case 2: A and B are Indirectly Related

- Path length 2



Number of co-authors
in common > 10 ?

If so,
then COI is: Medium

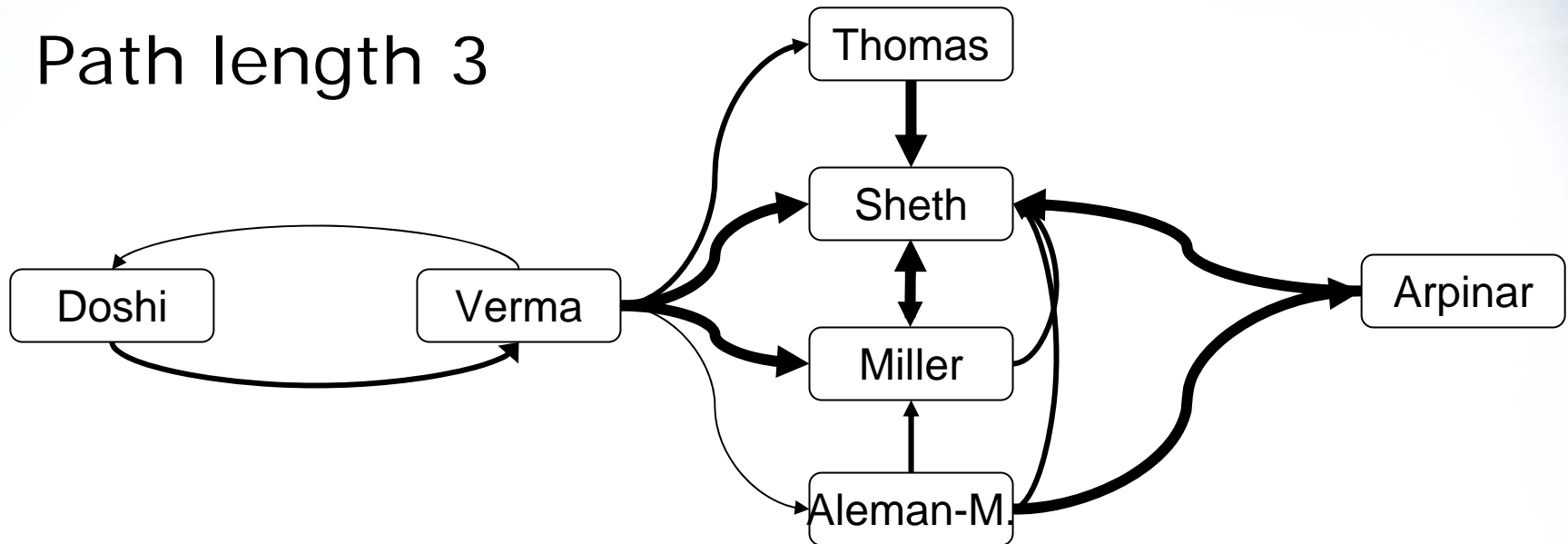
Otherwise, depends on weight

--	--



Case 3: A and B are Indirectly Related

- Path length 3



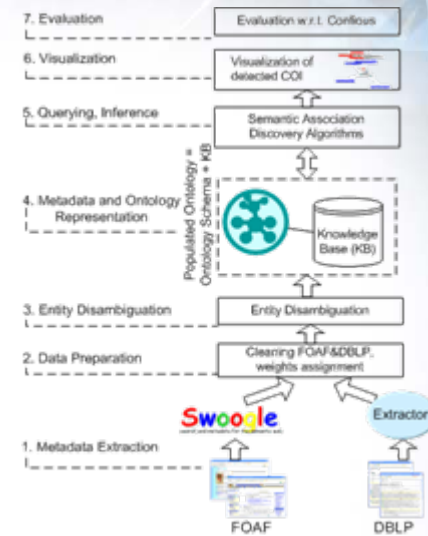
COI Level is set to: Low
(in most cases, it can be ignored)



Our Experiences: Multi-step Process

Building Semantic Web Applications requires:

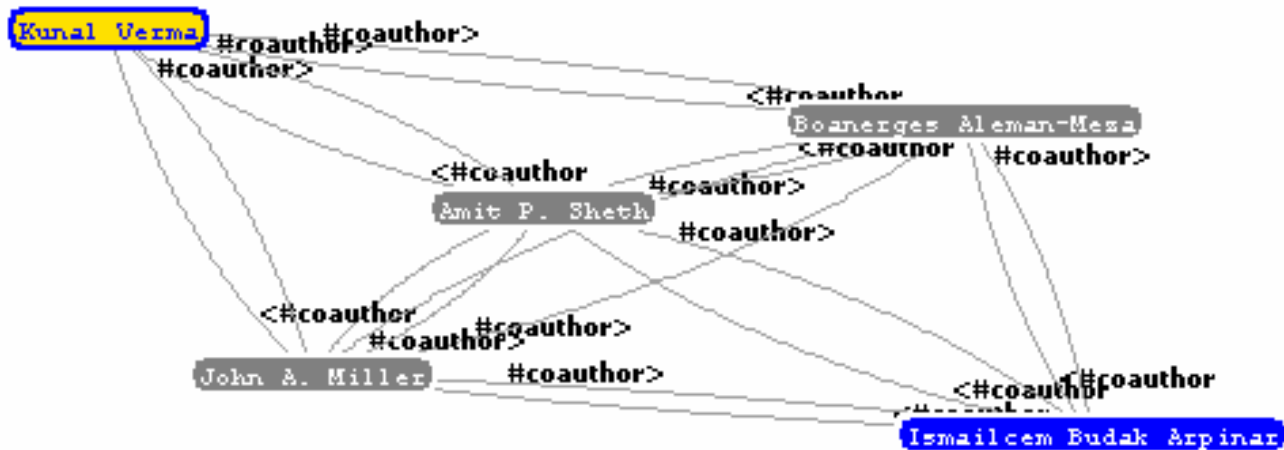
5. Visualization





Visualization

- Ontology-based approach enables providing 'explanation' of COI assessment



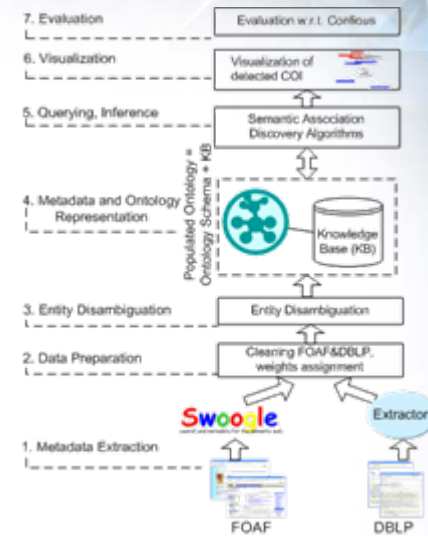
- Understanding of results is facilitated by named-relationships



Our Experiences: Multi-step Process

Building Semantic Web Applications requires:

6. Evaluation





Evaluating COI Detection Results

- Used a subset of papers and reviewers
 - from a previous WWW conference
- Human verified COI cases
 - Validated well for cases where syntactic match would otherwise fail
- We missed on very few cases where a COI level was not detected
 - Due to lack of information or outdated data



Examples of COI Detection

Wolfgang Nejdl, Less Carr

Low level of potential COI

1 collaborator in common

(Paul De Bra co-authored once with Nejdl and once with Carr)

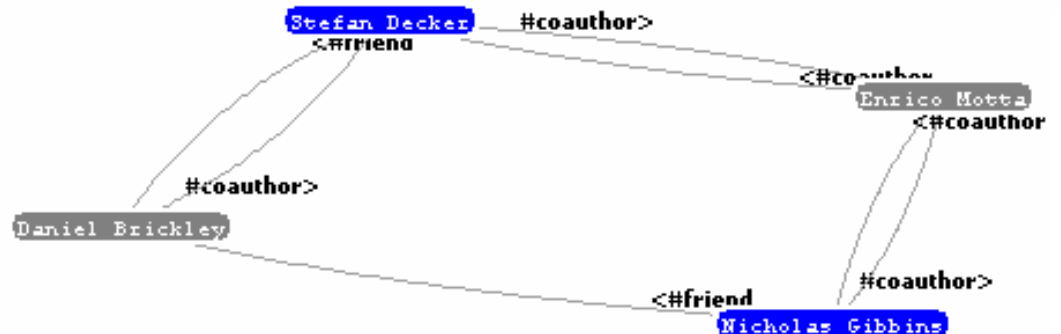


Stefan Decker, Nicholas Gibbins

Medium level of potential COI

2 collaborators in common

(Decker and Motta co-authored in two occasions, Decker and Brickley co-authored once, Motta and Gibbins co-authored once, Brickley and Motta never co-authored, but Gibbins (foaf)-knows Brickley)



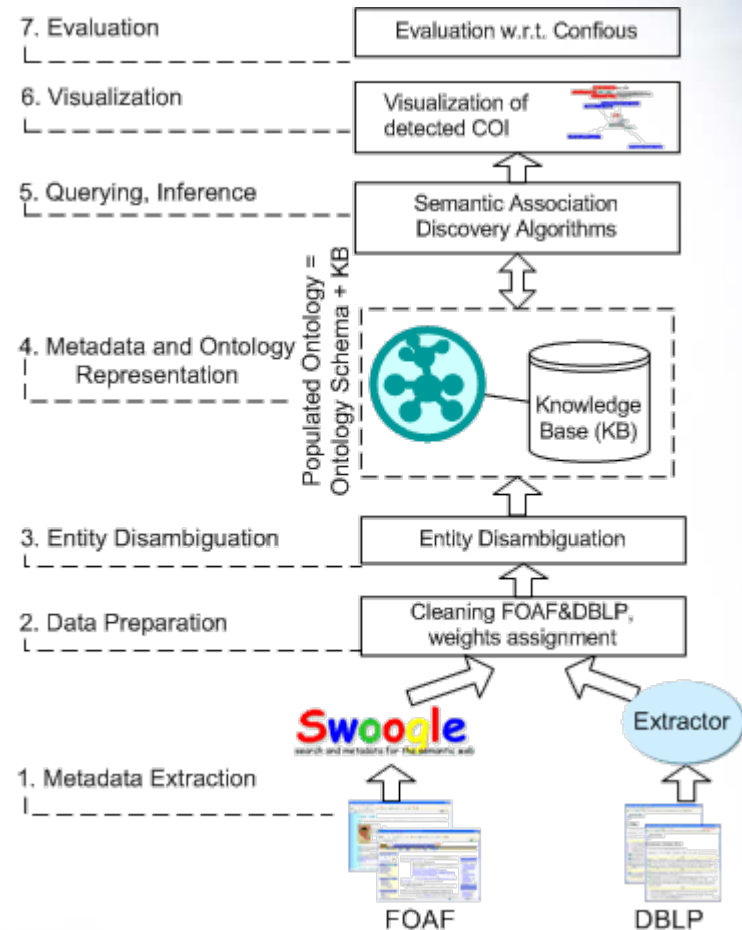
Demo at <http://lsdis.cs.uga.edu/projects/semdis/coi/> or, search for: coi semdis



Our Experiences: Multi-step Process

Building Semantic Web Applications involves a multi-step process consisting of:

1. Obtaining high-quality data
2. Data preparation
3. Metadata and ontology representation
4. Querying / inference techniques
5. Visualization
6. Evaluation





Evaluation

Two of the most active researchers in the field of Very Large Scale Distributed Information Systems (VLSIS) and Web Search

Authors:	Reviewers:	Karl Aberer	Sean Bechhofer	Mark Burstein	Isabel Cruz	Stefan Decker	Aldo Gangemi	R. V. Guha	Jeff Heflin	Ian Horrocks	Jane Hunter	M. Koubarakis	J. Mylopoulos	Wolfgang Nejdl	Guus Schreiber	Nigel R. Shadbolt
Dennis Quan			L3							<u>L2</u>						<u>L1</u>
Sean Bechhofer	<u>L1</u>	D	<u>L1</u>	<u>L1</u>	<u>M1</u>					H						<u>L4</u>
Alon Y. Halevy		<u>L1</u>			<u>L2</u>	<u>L5</u>		<u>L1</u>		LR, <u>L6</u>		<u>L3</u>	<u>L3</u>	<u>L1</u>		
Wendy Hall	<u>L1</u>	MR, <u>L1</u>				<u>L1</u>				<u>L2</u>	<u>L1</u>		<u>L1</u>	<u>L1</u>	<u>L2</u>	MR, <u>L2</u>
Leslie Carr	<u>L1</u>	M, <u>L1</u>				<u>L1</u>				<u>M1</u>				<u>L1</u>	<u>L2</u>	MR, <u>L7</u>
Timothy Miles-Board		M, <u>L2</u>				<u>L1</u>				<u>M1</u>						<u>L1</u>
Christopher Bailey																<u>L2</u>
Daniel Schwabe					<u>L1</u>	LR, <u>L2</u>				<u>L1</u>			<u>L1</u>	<u>L2</u> *	<u>L1</u>	
m. c. Schraefel	<u>L1</u>	<u>M1</u>		<u>M1</u>			<u>L1</u>							<u>L1</u>		M, <u>L2</u>
Nigel Shadbolt	<u>L1</u>	<u>L4</u>				<u>L8</u>				<u>L1</u>					MR, <u>L1</u>	D
Nick Gibbins	<u>L1</u>	<u>M1</u>				<u>M2</u> *									<u>L2</u>	H
Hugh Glaser	<u>L1</u>	<u>L2</u>													<u>L1</u>	M, <u>L5</u>
Steve Harris		<u>M1</u>													<u>L1</u>	H

Underlined: Confious would have failed to detect COI

Demo at <http://lsdis.cs.uga.edu/projects/semdis/coi/> or, search for: coi semdis



Our Experiences: Discussion

What does the Semantic Web offer today?
(in terms of standards, techniques and tools)

- Maturity of standards - RDF, OWL
- Query languages: SPARQL
 - Other discovery techniques (for analytics)
 - such as path discovery and subgraph discovery
- Commercial products gaining wider use



... Our Experiences: Discussion

What does it take to build Semantic Web applications today?

- Significant work is required on certain tasks
 - such as entity disambiguation
 - We're still on an early phase as far as realizing its value in a cost effective manner
- But, there is increasing availability of:
 - data (i.e., life sciences), tools (i.e., Oracle's RDF support), applications, etc



... Our Experiences: Discussion

How are things likely to improve in future?

- Standardization of vocabularies is invaluable
 - such as in MeSH and FOAF; but also: microformats
- We expect future availability/increase of
 - Analytical techniques used in applications
 - Larger variety of tools
 - Benchmarks
 - Improvements on data extraction, availability, etc



What do we demonstrate wrt SW

We demonstrated what it takes to build a broad class of SW applications: “connecting the dots” involving heterogeneous data from multiple sources- examples of such apps:

- Drug Discovery
- Biological Pathways
- Regulatory Compliance
 - Know your customer, anti-money laundering, Sarbanes-Oxley
- Homeland/National Security
-



Our Contributions

- Bring together semantic + structured social networks
- Semantic Analytics for Conflict of Interest Detection
- Describe our experiences in the context of a class of Semantic Web Applications
 - » Our app. for COI Detection is representative of such class



LSDIS

Large Scale Distributed Information Systems



University of Georgia
Computer Science Department

Data, demos, more publications at
SemDis project web site,
<http://lsdis.cs.uga.edu/projects/semdis/>

Thanks!

Questions



References

Related SemDis Publications (LSDIS Lab - UGA)

- B. Aleman-Meza, C. Halaschek-Wiener, I.B. Arpinar, C. Ramakrishnan, and A.P. Sheth: [Ranking Complex Relationships on the Semantic Web](#), IEEE Internet Computing, 9(3):37-44
- K. Anyanwu, A.P. Sheth, [p-Queries: Enabling Querying for Semantic Associations on the Semantic Web](#), WWW'2003
- C. Ramakrishnan, W.H. Milnor, M. Perry, A.P. Sheth, [Discovering Informative Connection Subgraphs in Multi-relational Graphs](#), SIGKDD Explorations, 7(2):56-63

Related SemDis Publications (eBiquity Lab – UMBC)

- L. Ding, T. Finin, A. Joshi, R. Pan, R.S. Cost, Y. Peng, P., Reddivari, V., Doshi, J. and Sachs, [Swoogle: A Search and Metadata Engine for the Semantic Web](#), CIKM'2004
- T. Finin, L. Ding, L., Zou, A. Joshi, [Social Networking on the Semantic Web](#), The Learning Organization, 5(12):418-435

Other Related Publications

- X. Dong, A. Halevy, J. Madahvan, Reference Reconciliation in Complex Information Spaces, SIGMOD'2005
- B. Hammond, A.P. Sheth, K. Kochut, [Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content](#), In Kashyap, V. and Shklar, L. eds. Real, World Semantic Web Applications, Ios Press Inc, 2002, 29-49
- A.P. Sheth, I.B. Arpinar, and V. Kashyap, [Relationships at the Heart of Semantic Web: Modeling, Discovering and Exploiting Complex Semantic Relationships](#), Enhancing the Power of the Internet Studies in Fuzziness and Soft Computing, (Nikravesh, Azvin, Yager, Zadeh, eds.)
- A.P. Sheth, [Enterprise Applications of Semantic Web: The Sweet Spot of Risk and Compliance](#), In IFIP International Conference on Industrial Applications of Semantic Web, Jyväskylä, Finland, 2005
- A.P. Sheth, [From Semantic Search & Integration to Analytics](#), In Dagstuhl Seminar: Semantic Interoperability and Integration, IBFI, Schloss Dagstuhl, Germany, 2005
- A.P. Sheth, C. Ramakrishnan, C. Thomas, [Semantics for the Semantic Web: The Implicit, the Formal and the Powerful](#), International Journal on Semantic Web Information Systems 1(1):1-18, 2005