



MMGPS Workshop, London

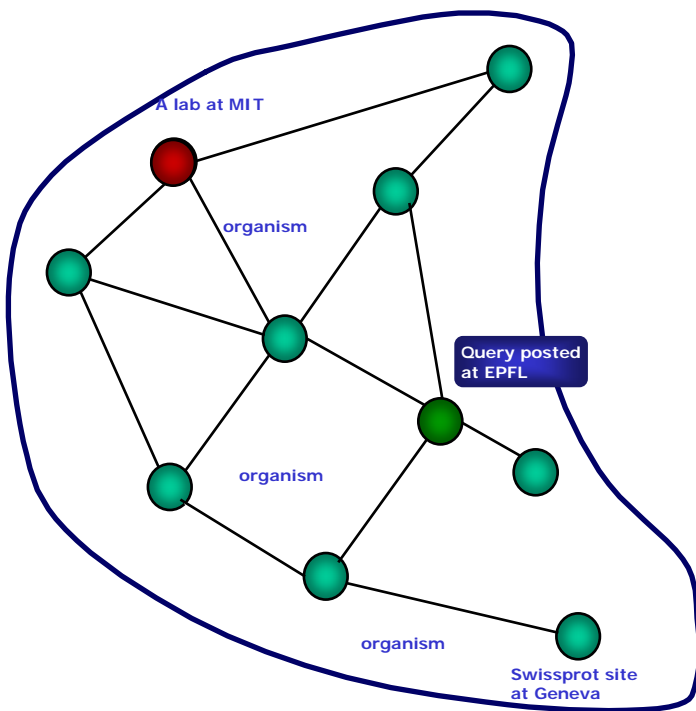
The Chatty Web approach for global semantic agreements

Philippe Cudré-Mauroux, Karl Aberer

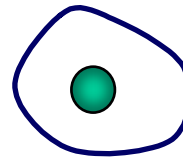
Distributed Information Systems Laboratory (LSIR)
Swiss Federal Institute of Technology, Lausanne (EPFL)

The Problem (1)

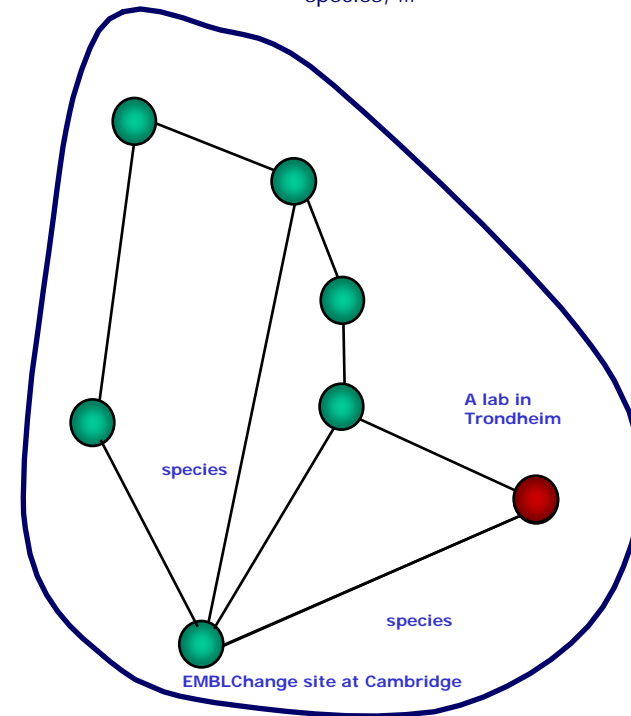
SwissProt peers
authors, titles, organism, ...



other peers
authors, ...



EMBLChange peers
species, ...



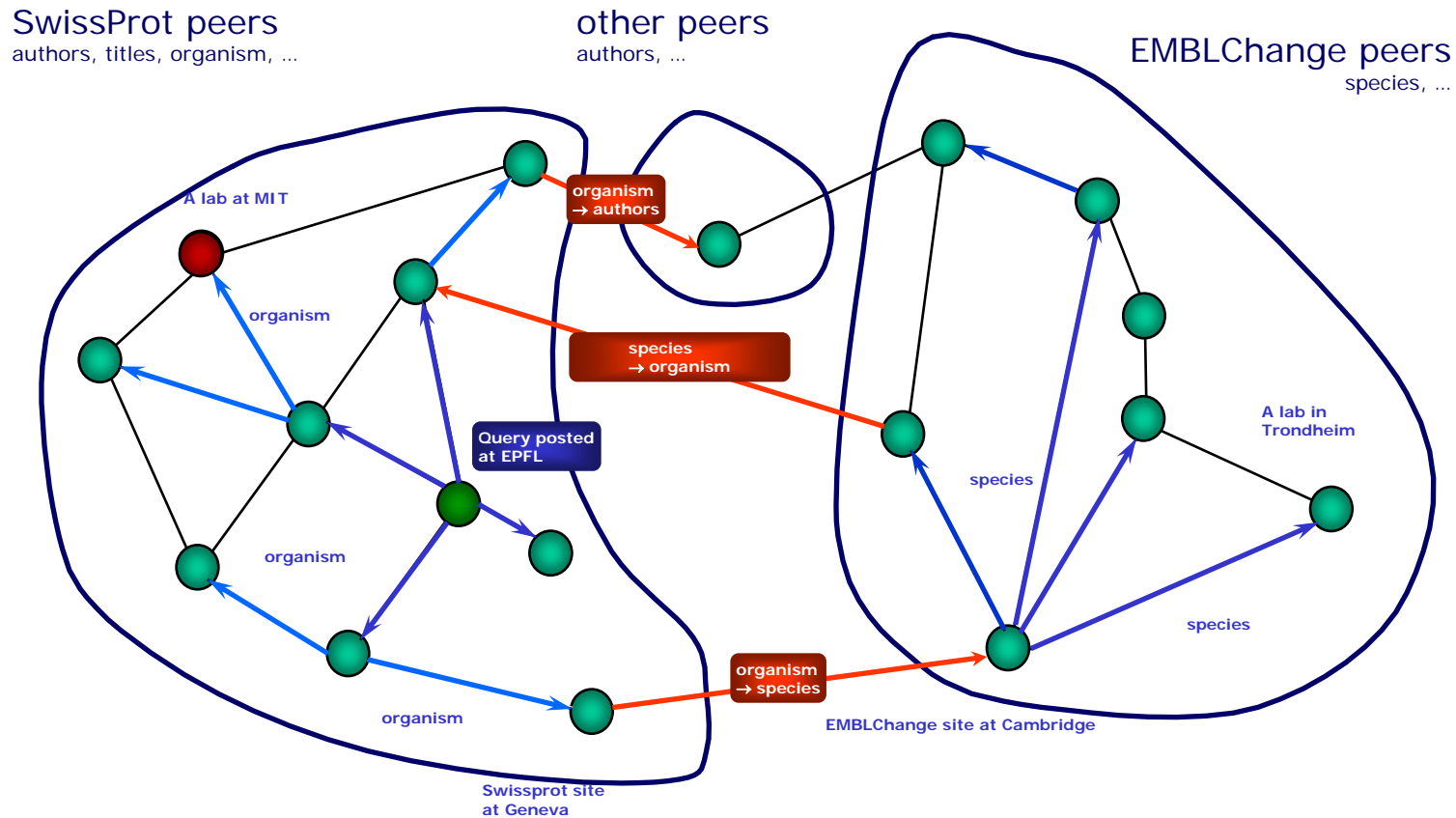


The Problem (2)

- How to obtain semantic interoperability among heterogeneous data sources without relying on pre-existing, global semantic models?

Outline of the solution

- Local translations enabling global agreements





Query Forwarding

- To whom shall we send the queries?
 - To peers susceptible of sending us a response...
- Simplistic solutions
 - Local Neighboring (same schema)
 - Low recall
 - Query Flooding (entire network)
 - Low precision, high network load
- Semantic Gossiping
 - Query forwarding by selecting the right peers
 - Query dependant PHBs (Per-Hop Behaviors)
 - Query / transformed queries analysis
 - Intrinsic measures (syntactic distances)
 - Extrinsic measures (semantic distances)

On Translations

Q1 =
 FOR \$p IN "zoran_project.xml"/*
 WHERE "Jie Project" IN p/title
 RETURN
 <start> \$p/duration/start </start>

```
<zoran_project>
  <title> My Project </title>
  <acronym> MP </acronym>
  <duration>
    <start>10/11/01</start>
    <end>13/10/05</end>
  </duration>
  <team>
    <member>1</member>
    <member>2</member>
  </team>
</zoran_project>
```

Q2 =
 FOR \$pr IN T12
 WHERE "Jie Project" IN p/title
 RETURN
 <start> \$p/duration/start </start>

```
<jie_project>
  <Name> Jie Project </Name>
  <Begin> 02/05/02 </Begin>
  <Level>Diploma</Level>
  <Location>EPFL</Location>
  <Lab>LSIR</Lab>
  <Institute>IIF</Institute>
  <Faculty>I&C</Faculty>
  <Length>6 months</Length>
  <Benefits>...</Benefits>
  <Report>Yes</Report>
</jie_project>
```

T12 =
 FOR \$p IN "jie_project.xml"/*
 RETURN
 <zoran_project>
 <title> \$p/Name </title>
 <acronym> </acronym>
 <duration>
 <start>\$p/Begin</start>
 ...

$$(T_{p_1 \rightarrow p_2}(q_{p_1}))(DB) = q_{p_1}(q_T(DB))$$



Similarity Measures

- **Syntactic Similarity**

- Similarity measure between an original and a transformed query.
- Iterative computation of information loss in selections / projections.

- **Semantic Similarities**

- Probabilistic analysis (max. likelihood) upon the correctness of translations based on feedback received



Semantic Similarity

- **Cycles Detection**

- Detection of query cycles:

- - $(T1 \rightarrow n) (A_i) = (A_i)$ ✓
 - - $(T1 \rightarrow n) (A_i) = (A_j)$ ✗
 - - $(T1 \rightarrow n) (A_i) = \emptyset$

- **Results Analysis**

- Content-retrieval techniques:

- classification rules to relate a returned documents to queries (extensional VS intentional expression of concepts)

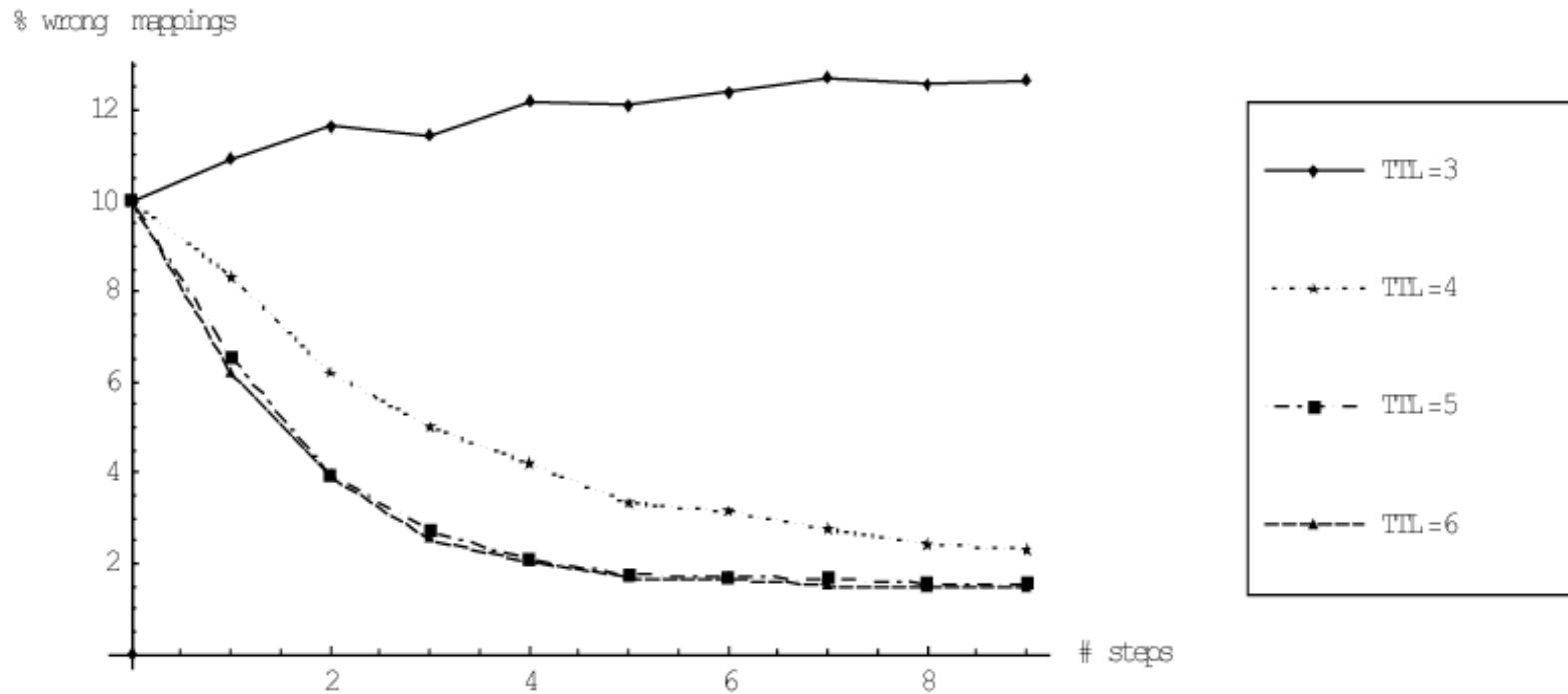


Realizing semantic interoperability

- Evaluations based on Chatty Web simulations.
- Automatic correction of erroneous mappings based on evidences gathered.
- Small-world graph

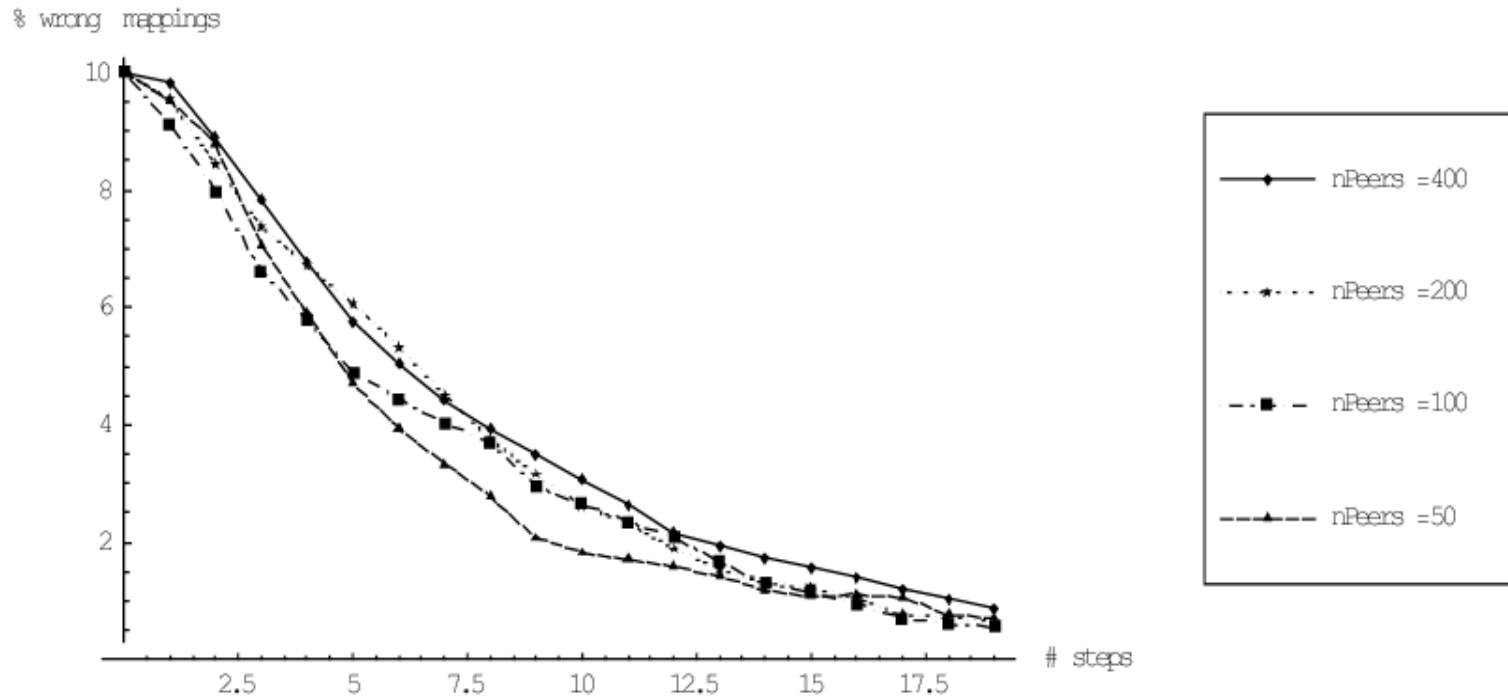
=> Self-repairing semantic networks

Some Results (1)



Sensitivity to TTL
(cycle analysis only, 25 peers, 4 concepts)

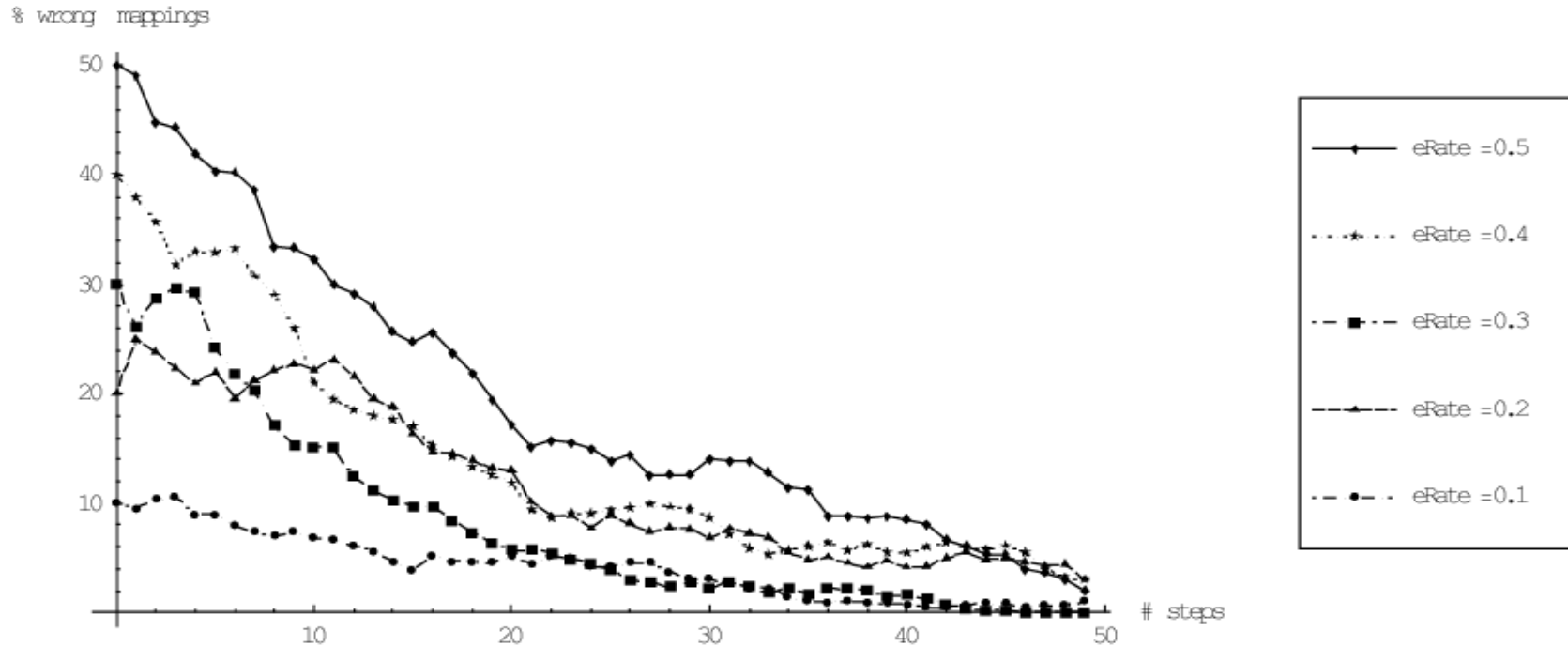
Some Results (2)



Scalability

(results analysis only, 4 concepts, TTL=3, misclassification rate=0.1, 2 documents/peer on avg.)

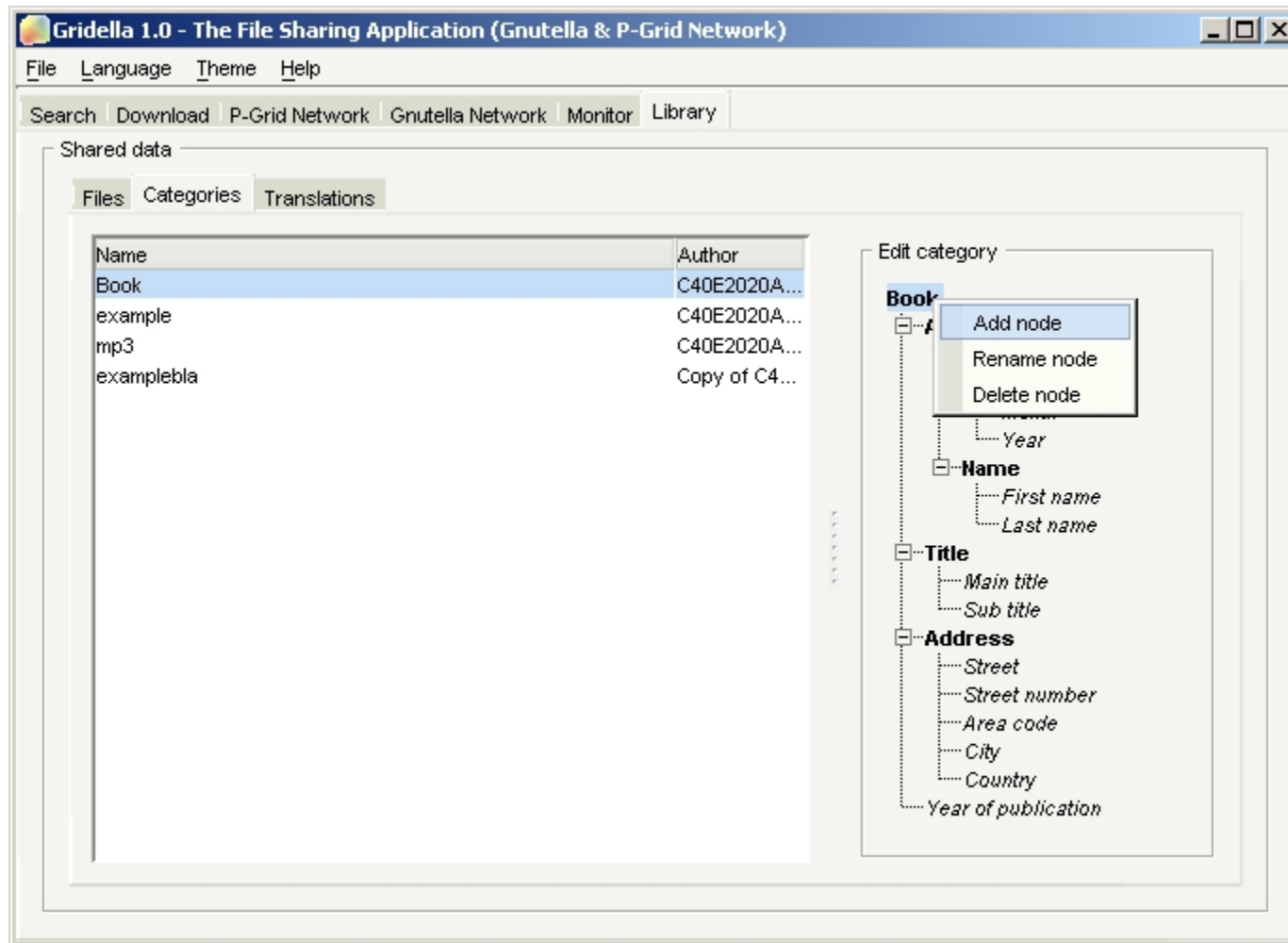
Some Results (3)



Combined results

(25 peers, 4 concepts, TTL=6 | 3, misclassification rate=0.1, 2 documents/peer on avg.)

P-Grid Implementation (ongoing work)





References

- **Start making sense: The Chatty Web approach for global semantic agreements,**
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth
1st issue of Journal of Web Semantics.
- **The Chatty Web: Emergent Semantics Through Gossiping**
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth
Proceedings of the Twelfth International World Wide Web Conference (WWW2003), 20-24 May 2003, Budapest, Hungary.
- **A Framework for Semantic Gossiping**
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth
SIGMOD Record, 31(4), December 2002.
- **<http://www.p-grid.org/>**



MMGPS Workshop, London

The Chatty Web approach for global semantic agreements

Philippe Cudré-Mauroux, Karl Aberer

Distributed Information Systems Laboratory (LSIR)
Swiss Federal Institute of Technology, Lausanne (EPFL)