Semantic Quality

Karl Aberer
School of Computer and Communication Sciences
EPFL
How to Detect a Semantic Agreement?

SwissProt peers
authors, titles, organism, ...

other peers
authors, ...

EMBLChange peers
species, ...

Check what is preserved in cycles!
Approach

where did the problems occur?
Experimental Evaluation

- **Creation topology of** \( n \) **peers**
  - Peers share \( |C| \) concepts, but use distinct names to refer to them
  - Each peer connected through translation to other peers (e.g. using small world graphs)

- **Generation of mappings for every translation link**
  - Correct mappings
  - Erroneous mappings (eRate)

- **Semantic gossiping techniques applied iteratively to detect and rectify erroneous translations**
  - At every step peer randomly selects one name and issues a query about it
  - Query propagated in Gnutella-like fashion with TTL value

- **Peers evaluate the correctness of current mapping (maximum-likelihood techniques)**
  - Peer adopts most probably correct mapping if probability of being correct is above 50%
Cycle analysis - Sensitivity to the initial error rate

- $N=25$; $\|C\|=4$; $TTL=5$; $l=4$
Results Analysis – Sensitivity to number of documents
Combined results
Results Analysis - Scalability

% wrong mappings

©2003, Karl Aberer, School of Computer and Communication Sciences
Some Conclusions

• Semantic interoperability can be understood as a problem of *mapping quality*

• Assessing quality of mapping comes down to a process of *establishing agreement and trust* based on experience (see trust and reputation management)

• Notions of quality are *context and task dependent*

• There is no absolute truth: *different consistent interpretations* may exist

• Agreed upon semantics in a large network is an *emergent state* of resulting from the dynamic behavior of the participating agents/peers

• Experiences from complex systems show that these states follow certain *basic laws* (e.g. power-laws), should be no different for semantics