SIMILE: objectives, status & demo

Stefano Mazzocchi - MIT Libraries

“Practices of Knowledge Sharing” Workshop
April 20-22 2004, Heraklion, Greece
Objectives
SIMILE Goals

• Make semantic interoperability of metadata a reality for digital libraries by:
  • providing reusable software for browsing, searching and mapping heterogenous metadata
  • using semantic web technologies
  • identifying issues, gaps and best practices
SIMILE Vision

• tools should help humans focus on their abilities, amplifying, not replacing them!

• metadata quality depends on its heterogeneity

• serendipitous discovery is a value that should not get lost

• empower recombinant metadata
SIMILE Participants

• MIT Libraries (MacKenzie Smith)
• MIT CSAIL (David Karger)
• HP Labs (Mick Bass)
• W3C (Eric Miller)
Status
Longwell

- faceted metadata browser
- aimed at end users
- goal is to show max functionality with min complexity (maximize usability)
Knowle

- RDF browser
  - aimed at semantic web specialists
  - goal is to enable cognitive estimations of complex models
Datasets

- ARTStor
- MIT OpenCourseWare
- Wikipedia
- CIA World Fact Book (in progress)
Schemata

- Dublin Core
- VRA
- LOM
- SKOS
- SIMILE’s own glue ones
- LoC TGM (in progress)
Achieved Results

• Usable implementation of both Longweel and Knowle as web applications
• Passed the 0.5 Megatriples wall
• Successful use of XSLT2 as XML->RDF bridge
• Use of the Levenshtein distance on literals to evaluate potential mappings between datasets
Demo
Open Questions
Scalability

• How more complex can the model grow before saturating our computational capacity?

• How can we design a distributed architecture and still be fast enough to be useable?
Connectivity

• How can we increase the connectivity when merging models with reasonable costs and without compromising perceived metadata quality?
Provenance

• How should provenance influence the reasoning on aggregated models?
Evolution

• How can we deal with the evolution of models and their impact on previous inferenced interpretations?

• Can time be another provenance or we need a different dimension?
Disagreement

• How well can the semantic web model cope with disagreement?

• How do we distinguish disagreement from mistakes?
Thanks!