Social Network Representations

Peter Mika
Researcher, Data Architect
Yahoo! Research
Yahoo! Research Barcelona

- Established January, 2006
- Led by Ricardo Baeza-Yates
- Research areas
  - Web Mining
    - content, structure, usage
  - Distributed Web retrieval
  - Multimedia retrieval
  - NLP and Semantics
Why bother?

• Yahoo’s SearchMonkey application framework, a part of Yahoo’s Open Strategy

BEFORE

**Raspberry Chocolate Mousse Cake Recipe - AcmeRecipe**
This award winning Recipe is simple to make with fresh ingredients. This cake can be served topped with chocolate shavings, raspberries and mint....
www.acmercipe.com/desserts/01223 - 49k - Cached

AFTER

**Raspberry Chocolate Mousse Cake Recipe - AcmeRecipe**
This award winning recipe is simple to make with fresh ingredients. This cake can be served topped with chocolate shavings, raspberries and mint....
www.acmercipe.com/desserts/01223 - 49k - Cached

• Application reuse requires data compatibility
Decentralization: how to make it work

- Shared syntax (RDF, RDFa)
- API compatibility (OpenSocial)
- Authentication (OpenID, OAuth)
- Incentives to participate
- Interlinking, identity management
- Decentralized vocabulary management
  - Requires social agreements!
## Microformats vs. RDF/RDFa

### Microformats
- Ease of use
- Syntax + vocabulary
- HTML compatible
- Fixed schema, fixed combinations
- Adoption
  - hCard: 1.9% e.g. twitter, linkedin, hi5
  - XFN: 0.3% e.g. twitter, hi5, mybloglog
  - hResume: 0.03% e.g. linkedin, xing

### RDF/RDFa
- Learning curve
- Syntax only
- Requires XHTML (no one cares)
- Extendible, open
- Adoption
  - Up-and-coming
- Identification by reference (URIs)

**BOTH ARE USEFUL AND WILL BE USED**

**RDF HAS THE POTENTIAL OF A LINGUA FRANCA**

**HOW DO WE MAKE THEM CONNECT?**
Other issues

• Vocabularies are often a mismatch to data
  – Data-driven vs principled design

• Doing it right vs making it simple
  – People in general don’t understand the differences
    • Object vs. Document about the object
    • Resource vs. Literal
    • Type vs. Datatype
  – Complexity leads to abuse

• Mobile readiness
  – e.g. values as image/sound/video

• Representation of trust, time, provenance
Questions?

• pmika@yahoo-inc.com