Breakout session on

Methods/Software as Standards
e.g., LDA

Lead: All

Participants: Andre Skupin, Margaret Clements, Katy Borner, Ying Ding, Stasa Milojevic
State of the Art:
Please List Existing Standards

LDA for topic identification, dimensionality reduction

Quality:
Can human users make sense of topics
Can you use topics to support retrieval and predictions
Identify Most Needed Standards

Methods/Software to
• Disambiguate, ddupe (http://www.cs.umd.edu/projects/linqs/ddupe)
• Extract networks
• Dimensionality reduction LDA (http://mallet.cs.umass.edu) NNF
• Layout algorithms (KK, FR)
• Rendering (GIS, Pajek)

LDA computes from probability distribution,

Non negative factorization does matrix analysis,
http://en.wikipedia.org/wiki/Non-negative_matrix_factorization

Arrive at same result
Processes to Get these Standards

De facto standards: ISI, .net data formats

WC3 Standards – submit candidate standards proposals, review, approval. Democratization of standards to those than can pay membership.

Bottom up – to get emergent areas
Top-down – to impose structure, resolve mapping issues, good labels. Upper-level ontologies to meet in middle.
Processes to Update/Maintain/Promote/Align These Standards

Map must expose sockets
Any artifact should have sockets.
Need to know how to do proper cosine similarity calculation.
• Identity (PaperID)
• similarity (vis n-dim vector calculation)
• semantic sockets (same semantic class/category)

Take novel record and see probability landscape of where the new record should go.

---

Drop in county data to state US ma. Map then aggregated to states. –
Semantic socket.
Yahoo geocoding. Yahoo gets address, delivers lat/long, goes to map that has Mercator projected system, gives image back.
In geography, we can take maps from different projections and drag and drop.

Hard to design sockets (resource) and easy to design plugs (users).
Joseph, Katy et al build a giant socket system (standard) and some prototypical plugs in anticipation of future plugs/plugins.