Data for the People, by the People

Mor Naaman
Rutgers SCILS
Flashback: 2007

Discuss:

- The Semantic Web is Dead
- Welcome to the Emerging-Semantics Web!
A Semantic Web (could be) Alive!

http://flickr.com/photos/cenz/16128560/
Academics on Flickr

More crazy people

At the banquet. Wine was flowing, and food was not. This is the result.

Tags
- Michael Fleischman
- Jurgen Scheible
- Brian Bailey
- ACM MM
- ACM MM 2007
- ACM Multimedia 2007
- Banquet
- Germany
- zonetag
- cameraphone
- Augsburg
- zip:86199
- geotagged
- celltagged
Guess the Tags

Zion  Hiking  Mountains  Landscape
Nature  Valley

http://flickr.com/photos/oneeighteen/1610814928/
Guess the Tags

Dog  Puppy  White  Animal
Pet  Sad    Cute   Nepal

http://flickr.com/photos/klash/858533852/
Flickr Tag Affordances
So, Why We Tag?

- 13 ZoneTag users (23-45, 9m, 4f)
- All “taggers”
- Semi-Structured interviews
A whole bunch of reasons!

Function

Organization

* Retrieval, Directory
* Search

Communication

Sociality

Self
Also by the Numbers

- Motivations operationalized into survey, shown correlation with number of tags

### Stated Motivations
- Self
- Family & Friends
- Public

### Social Presence Indicators
- Groups
- Contacts

### Control
- Photos

**Tags (R² = .571)**
Why *Not* Tag (other’s photos)?

- 99% of the tags on Flickr are added by the photo owner

- Not Collected
- Not Identified
- Not prominent
- Not aggregated
Tagging Systems are not Created Equal

- Source and type of object
- Tagging rights
- Tagging support/suggestions
- Aggregation (set/bag)?
- Display/functionality
Is Facebook Different?

• Yes!

• Social constructs encourage ("semantic") people tagging

• No other tags to speak of
Tagging, communities, vocabulary, evolution (*toread*)

- Sen et al., CSCW 2006

**Movie Tags** *(more about tags)*
Add and edit tags here

- **My Tags** [edit]
  - phone booth
  - dark
  - carrie-anne moss in tight latex pants
  - power of myth
  - sufficiently explodey to be good

[add new tags]

**Popular tags:**
Click on this icon (_PUT_) to add a tag to your list!
- Great heroics (1)
- smart (1)
- lots of kicking (1)
- Brilliant (1)
- Pure action (1)
Sen et al. Controlled Tagging Experiment

- **Treatments:**
  - Private tags
  - Shared tags (various treatments)

<table>
<thead>
<tr>
<th>Tag Group</th>
<th>Subjective</th>
<th>Factual</th>
<th>Personal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unshared</td>
<td>24%</td>
<td>38%</td>
<td>39%</td>
</tr>
<tr>
<td>Shared</td>
<td>9%</td>
<td>82%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Data, People, Semantics? Mor Naaman - Rutgers SCILS
SSS Factors in Social Media

- **Self**
  - Psychological, cognitive, motivational

- **Social**
  - Social psychology, social network, communication, interaction

- **System**
  - Application design, system affordances, functionality
“research by Oded Nov of the Polytechnic Institute of New York University is beginning to ascertain why Wikipedians post entries and what motivates their activity; the psychological drivers that are revealed will help us understand how to encourage people to contribute to the Semantic Web.”

TBL.
Where am I going with this?

- The motivations
- The data
- The multimedia
- A semantic web?
Where am I going with this?

- The motivations
- The data
  - Semantics from unstructured data
  - Tag Maps / World Explorer
- The multimedia
- A semantic web?
Throw Location into the Mix

More crazy people

At the banquet. Wine was flowing, and food was plentiful.

View Flickmor's map

Tags
- Michael Fleischman [x]
- Jurgen Scheible [x]
- Brian Bailey [x]
- ACM MM [x]
- ACM MM 2007 [x]
- ACM Multimedia 2007 [x]
This is not an Arch

- Errors
- Noise
- Bias
- …
Tag Patterns
Tag Patterns
Tag Patterns

Data, People, Semantics?

Mor Naaman - Rutgers SCILS
Folkpolygonos
Also: Time

Data, People, Semantics?

Mor Naaman - Rutgers SCILS
Extracting “Semantics”

- Examine all the photos associated with a tag
- Does the tag describe a place? An event? Neither?
- Using scale-structure clustering and measures of entropy
Where are we going?

- The motivations
- The data
  - Semantics from unstructured data
- Tag Maps / World Explorer
- The multimedia
- A semantic web?
A Better Representation of Place
Intuition

More “activity” in a certain location indicates the importance of that location

Tags that are unique to a certain location can be used to represent the location

Ahern et al. [JCDL 2007]
Algorithm

- Document are clusters of photos
- Simple, tf-idf scoring of each tag based on clusters (document)
+++40,000,000 photos:
Side note from 1976

We can do Paris too!


"Tag Maps: World Explorer"
Live demo: World Explorer

http://flickr.com/photos/wynandvanpoortvliet/264171197/

http://tagmaps.research.yahoo.com

Ahern et al. [JCDL 2007]
Where am I going with this?

- The motivations
- The data
- The multimedia
- A semantic web?
Where am I going with this?

- The motivations
- The data
- The multimedia
  - Flickr landmark representation
  - YouTube audio synchronization (if time)
- A semantic web?
Multimedia Opportunity!

- We used social media to identify semantics of various tags
- Can we apply targeted image analysis methods in a robust way?
Challenges

- Content is still hard...
- Unstructured data (no semantics)
- Tags, not ground truth labels
- Noise
- Scale
  - Computation
  - Long tail means no supervised learning
Representative Content

- We identified the "landmark" tags
- We know where their photos are taken
- Can we select the most representative photos?
Discover Views Clustering Location Summary (Representative Photos)
Rank Clusters by "Representativeness"

Tagged Photos

Discarded Views (Non-representative Photos)

Kennedy et al. [MM 2007, WWW 2008]
Technology: Clustering

- Visual clusters predict similar views
  - Color, texture
  - Simple k-Means
Technology: Ranking

- Repeated features suggest relevane/representativeness
- SIFT
Ranking Considerations

- Number of users
- Temporal spread
- Inter vs. Intra-cluster similarity
- Visual coherence
- Visual connectivity
Sample Results - GG

Tag only
Tag + Location
Tag + Location + Visual

Data, People, Semantics?
Mor Naaman - Rutgers SCILS
Onto YouTube

- Tons of content
- Used context tools to crawl all videos taken at the same event
- Audio Fingerprints to find/sync overlapping videos
Better Metadata

- Grouping based on link graph
  - Emerging concepts
  - Significant moments of interest
  - Quality audio segments
Where am I going with this?

- The motivations
- The data
- The multimedia
- A semantic web?
Machine tags – Emerging Semantics

“Machine tags are not RDF... but they can play RDF on TV”

– Aaron Cope, Hackr

- Lastfm:event=352444
- Camera:lens=35mm
- ...

Data, People, Semantics? Mor Naaman - Rutgers SCILS
Folk semantics

- xCal, Location, vCard, microformats
- GeoNames, openSocial, Facebook IDs, Email (identity, kind of)
- Twitter #
- Addressable leading platforms of content: Wikipedia, Amazon, iTunes
Driven By Major Players

- Yahoo! Search Monkey
- Outside.in location-based aggregator
- What if Google did it?
- Thinking about motivations...
Summary

- Self-Social-System motivation
  - When building new user-driven applications and designing new systems

- Can apply content analysis to social media when domain is clear despite scale and noise
Notes

- All photos CC or with permission
  - http://flickr.com/photos/cenz/16128560/
  - http://flickr.com/photos/oneeeighteen/1610814928
  - http://flickr.com/photos/klash/858533852/
  - http://flickr.com/photos/statico/2359648441/

- Didn’t talk about: ZoneTag, Zurfer (for your Nokia phone) – just look for them...
That Last Slide

- Thanks: Lyndon Kennedy, Tye Rattenbury, Alex Jaffe, Shane Ahern, Simon King, Rahul Nair, Jeannie Yang @ Y!

mor@scils.rutgers.edu

http://mornaaman.com

- Faculty: SCILS is looking for you:
  - Health information, Global Democracy, Social Media
  - http://scils.rutgers.edu

- PhD students: yes, you too.

- HTTP://MULTIMEDIAGRANDBALLENGE.COM