How can Search Interfaces Enhance the Value of Semantic Annotations (and Vice Versa?)

Keynote Talk

ESAIR’13:
Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval

Marti Hearst, UC Berkeley
October 28, 2013
Talk Structure

- Four Interlocking Points about State-of-the-Art
- Together: Roadmap for a New Direction:
  - Richer, usable search that makes use of semantic information in a realistic way.
Four Interlocking Points

- Faceted Navigation solves a Search UI problem
  - Now users are ready for the next evolution

- Auto-suggest is a good Search UI paradigm
  - Is currently limited by data/tasks applied to

- Behavior Log Analysis has made great strides
  - Results are not being fully utilized in Search UIs

- Current Knowledge Base use focuses on head queries
  - The real value will come in addressing harder queries
Faceted Navigation Solves a UI Problem

- Used in thousands of websites and tools now.
- Successful because:
  - Conforms to user expectations, even if they don’t fully understand the underlying model
  - Helps navigate better than keyword search alone
  - Reduces likelihood of empty results
  - Suggests new ways to go
  - The category structure is logical, understandable
The Problem with Hierarchy

- Inflexible
  - Force the user to start with a particular category
  - What if I don’t know the animal’s diet, but the interface makes me start with that category?

- Wasteful
  - Have to repeat combinations of categories
  - Makes for extra clicking and extra coding

- Difficult to modify
  - To add a new category type, must duplicate it everywhere or change things everywhere
The Problem With Hierarchy

- start
  - swim
    - fur
    - scales
    - feathers
    - fish
    - rodents
    - insects
    - salmon
  - fly
    - fur
    - scales
    - feathers
    - fish
    - rodents
    - insects
    - bat
  - run
    - fur
    - scales
    - feathers
    - fish
    - rodents
    - insects
    - robin
    - wolf
  - slither
  - ...
The Idea of Facets

- Facets are a way of labeling data
  - A kind of Metadata (data about data)
  - Can be thought of as properties of items

- Facets vs. Categories
  - Items are placed INTO a category system
  - Multiple facet labels are ASSIGNED TO items
What are facets?

- Sets of categories, each of which describe a different aspect of the objects in the collection.
- Each of these can be hierarchical.
- (Not necessarily mutually exclusive nor exhaustive, but often that is a goal.)

GeoRegion + Time/Date + Topic + Role
The Idea of Facets

- Create INDEPENDENT categories (facets)
  - Each facet has labels (sometimes arranged in a hierarchy)
- Assign labels from the facets to every item
  - Example: recipe collection
The Idea of Facets

- Break out all the important concepts into their own facets
- Sometimes the facets are hierarchical
  - Assign labels to items from any level of the hierarchy

<table>
<thead>
<tr>
<th>Preparation Method</th>
<th>Desserts</th>
<th>Fruits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fry</td>
<td>Cakes</td>
<td>Cherries</td>
</tr>
<tr>
<td>Saute</td>
<td>Cookies</td>
<td>Berries</td>
</tr>
<tr>
<td>Boil</td>
<td>Dairy</td>
<td>Blueberries</td>
</tr>
<tr>
<td>Bake</td>
<td>Ice Cream</td>
<td>Strawberries</td>
</tr>
<tr>
<td>Broil</td>
<td>Sorbet</td>
<td>Bananas</td>
</tr>
<tr>
<td>Freeze</td>
<td>Flan</td>
<td>Pineapple</td>
</tr>
</tbody>
</table>
Now there are multiple ways to get to each item
Using Facets

- The system only shows the labels that correspond to the current set of items
  - Start with all items and all facets
  - The user then selects a label within a facet
  - This reduces the set of items (only those that have been assigned to the subcategory label are displayed)
  - This also eliminates some subcategories from the view.
Advantages of Facets

- Can’t end up with empty results sets
  - (except with keyword search)
- Helps avoid feelings of being lost.
- Easier to explore the collection.
  - Helps users infer what kinds of things are in the collection.
  - Evokes a feeling of “browsing the shelves”
- Is preferred over standard search for collection browsing in usability studies.
  - (Interface must be designed properly)
Advantages of Facets

- Seamless to add new facets and subcategories
- Seamless to add new items.
- Helps with “categorization wars”
  - Don’t have to agree exactly where to place something
- Interaction can be implemented using a standard relational database.
- May be easier for automatic categorization
Facets vs. Hierarchy

- Early Flamenco studies compared allowing multiple hierarchical facets vs. just one facet.
- Multiple facets was preferred and more successful.
Limitation of Faceted Metadata

- Do not naturally capture MAIN THEMES
- Facets do not show RELATIONS explicitly

Which color associated with which object?
Description: 19th c. paint horse; saddle and hackamore; spurs; bandana on rider; old time cowboy hat; underchin thong; flying off.

Nature
Animal
Mammal
Horse

Occupations
Cowboy

Clothing
Hats
Cowboy Hat

Media
Engraving
Wood Eng.

Location
North America
America
Motivation

Description: 19th c. paint horse; saddle and hackamore; spurs; bandana on rider; old time cowboy hat; underchin thong; flying off.

By using facets, what we are not capturing?

The hat flew off;
The bandana stayed on.

The thong is part of the hat.

The bandana is on the cowboy (not the horse).
The saddle is on the horse (not the cowboy).
Old Yahoo restaurant guide combined:
- Region
- Topic (restaurants)
- Related Information
  - Other attributes (cuisines)
  - Other topics related in place and time (movies)
Biba
272 Boylston St
Boston, MA 02116
617/426-7878

**Find Nearby**
- Movie Theaters
- Bars, Pubs, and Clubs
- Cafes
- More A&E

**Features**

**Descriptions** Brought to you by...

*CuisineNet*

Biba's brash interior includes a bar mural depicting debauchery and an upstairs "food hall" with geometric accents. Chefs Lydia Shire and Susan Regis excite... more

*Fodor's*

Arguably Boston's best restaurant, and surely one of the most original and high-casual restaurants in America. Biba is a place to see and be seen,... more
Other Possible Combinations

- Region + A&E
- City + Restaurant + Movies
- City + Weather
- City + Education: Schools
- Restaurants + Schools
- ...

\[\]
Creating Tasks from HFM

- Recipes Example:
  - Click Ingredient > Avocado
  - Click Dish > Salad
  - Implies task of “I want to make a Dish type d with an Ingredient i that I have lying around”
  - Maybe users will prefer to select tasks like these over navigating through the metadata.
Taking Facets to the Next Level

- Allow queries to express combinations of facets
  - Be smart about recognizing those already expressed
- Go beyond just co-occurrence of nouns
  - Also allow relations between concepts
- Be smart about recognizing information needs:
  - Tasks that are combinations of concepts
  - Complex relations among concepts.
Query Auto-suggest

- sequoia
- sequoia national park
- sequoia restaurant
- sequoia trees
- sequoia capital
- sequoia fund
- sequoia high school
- sequoia yah

Sequoia National Park
Sequoia National Park is a national park in the...
Query auto-suggest

faceted navigation
faceted navigation seo
faceted navigation sharepoint 2013
faceted navigation example

About 2,510,000 results (0.15 seconds)

Ad related to faceted navigation

Faceted Navigation - smartlogic.com
www.smartlogic.com/federated-search
Drive enterprise, faceted and federated search with Semaphore
Contact us for more info - Enterprise Search

Design Patterns: Faceted Navigation - An A List Apart
alistapart.com/article/design-patterns-faceted-navigation
Apr 20, 2010 - Faceted navigation may be the most significant search innovation of the past decade. It features an integrated, incremental search and browse ...

Faceted search - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Faceted_search
Faceted search, also called faceted navigation or faceted browsing, is a technique for accessing information organized according to a faceted classification ...
Development - Projects - Mass market use - Online retail

Images for faceted navigation - Report images
Auto-suggest at the Next Level
Taking Auto-Suggest to the Next Level

- It’s gotten more conservative lately on the web
- It can try to organize possible results
  - As seen in the earlier NextBio interface
    - Group by gene, compound, tissue
  - As seen in the art museum database
    - Group by exhibition, artwork, video
- The trick is knowing which categories are relevant
Knowledge bases: popular head queries

Sequoia - Discover the 2014 Sequoia Online - Toyota.com
www.toyota.com/Sequoia
View MPG, Pricing, Specs & More.
Toyota USA has 163,138 followers on Google+

Sequoia Capital: Venture Capital: US
www.sequoiacap.com/
Investment focus: Seed stage, early stage and growth stage emerging technology companies across the US, China, India and Israel.
Jobs - Venture - About Us - Writing a Business Plan

Sequoia - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Sequoia
Sequoia or Sequoyah or Sequoya may refer to: Sequoyah
What do people most want to know about Albert Einstein?
- His theories of physics?
- His brain?
- His formulae?
- His views on life and god and the dropping of the bomb.

What do you get?
- His high school.
- The names of his wife and kids.
Albert Einstein | Albert Einstein Official Site
Albert Einstein official Web Site and Fan Club, featuring biography, photos, trivia, rights representation, licensing, contact and more.
Bio - Photo Gallery - Quotes - Links

Albert Einstein - Wikipedia, the free encyclopedia
Albert Einstein (ˈæl伯t ˈaɪnstaɪn; German: [ˈalbert ˈʔaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist who ...
Hans Albert Einstein - Eduard Einstein - Mileva Marić - Elsa Einstein

Albert Einstein - Biographical - Nobelp伫e.org
Albert Einstein - Biographical. Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879. Six weeks later the family moved to Munich, where ...

Albert Einstein Biography - Facts, Birthday, Life Story...
This Can Be Better!

- If the searcher doesn’t know what they want...
  - The tools should lead them to new discovery
- If they do have some ideas
  - The tools should show good next paths
    - But get beyond faceted choices
    - While still not being chaotic and unpredictable
Long “long tail” queries

- 30.4% of the unique queries in the two-month period were not issued to any search engine in the previous 6 months
- 16.8% were both long (4 or more words) and unseen
  - “Mount Rainier’s scenic hiking trails”

Focuses on matching Open Directory categories

- Only matches 1 in 5 times

Potentially would work better with richer representations

Mining Past Query Trails to Label Long and Rare Search Engine Queries, Bailey, White, Liu, Kumaran, ACM TWEB 4(4) 2009
Association Rules on Queries

- **Goal:** identify underperforming queries
- **Approach:** use lexical and topical attributes and association rule mining
  - They do this to classify queries as underperforming
  - I suggest this as a way to determine what parts of the knowledge base to show for these queries

### Table 1: Sample rule and example queries. Label indicates if query leads to satisfaction (SAT) or dissatisfaction (DSAT).

<table>
<thead>
<tr>
<th>No.</th>
<th>Query</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robocop remake poster</td>
<td>DSAT</td>
</tr>
<tr>
<td>2</td>
<td>pics from Robocop remake</td>
<td>DSAT</td>
</tr>
</tbody>
</table>

Behavior Logs + Knowledge Bases

- Figure out which concept combinations co-occur
- Use these to encourage longer relational queries
Review of the Main Points

- Faceted Navigation solves a Search UI problem
  - Now users are ready for the next evolution

- Auto-suggest is a good Search UI paradigm
  - Is currently limited by data/tasks applied to

- Behavior Log Analysis has made great strides
  - Results are not being fully utilized in Search UIs

- Current Knowledge Base use focuses on head queries
  - The real value will come in addressing harder queries
Putting It All Together

- KBs can improve harder queries. Here’s how:
- Create Search UIs that:
  - Encourage longer queries by:
  - Matching different meaning components
    - Like faceted navigation did, but more richly,
    - Based on behavior log analysis,
    - Match concepts against with knowledge base relations, and
  - Scaffold with auto-suggest interface to help get it right.
Envisionment
Thank you!

Marti Hearst, UC Berkeley
October 28, 2013