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

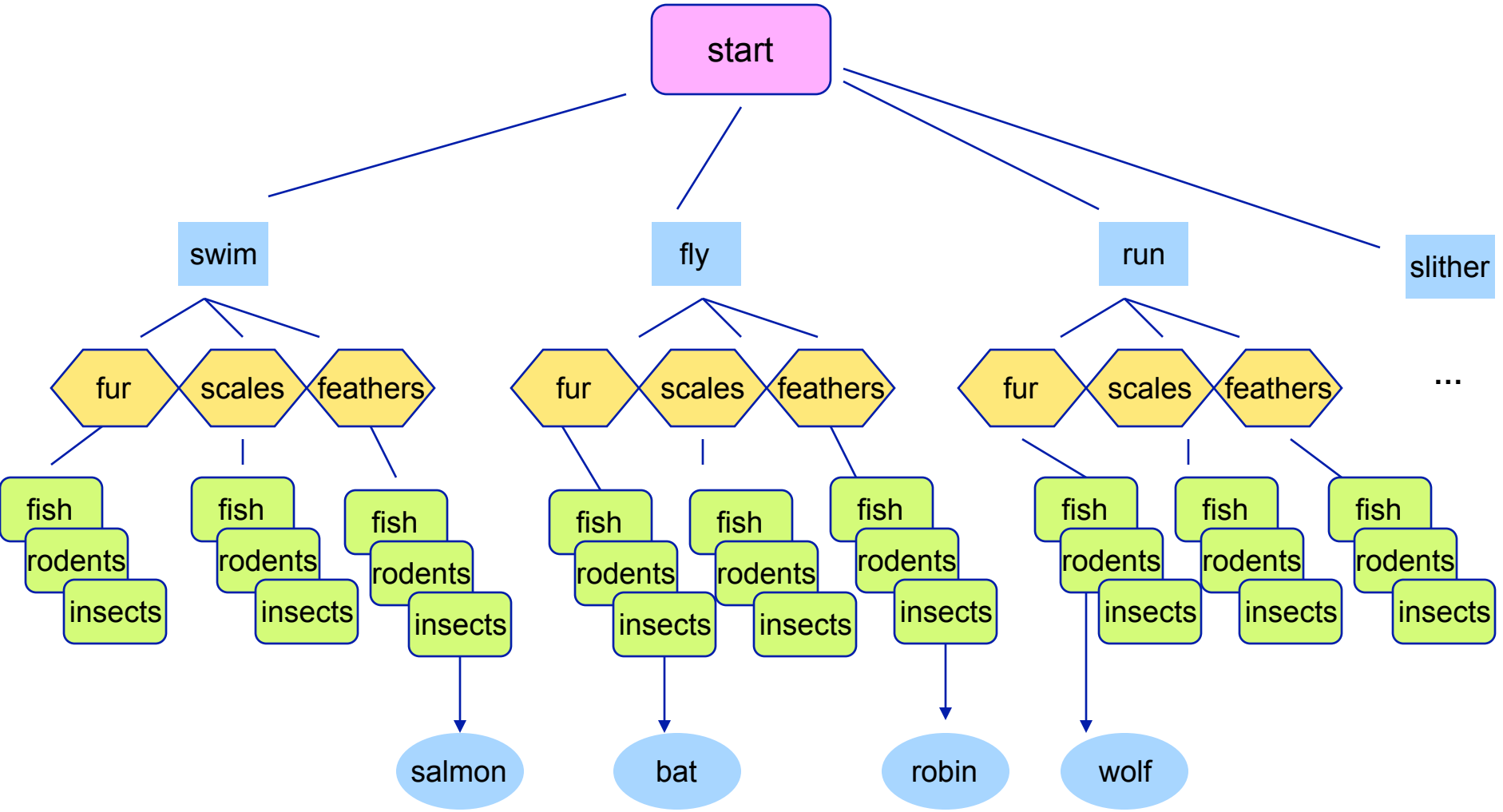
Faceted Navigation Solves a UI Problem

- 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

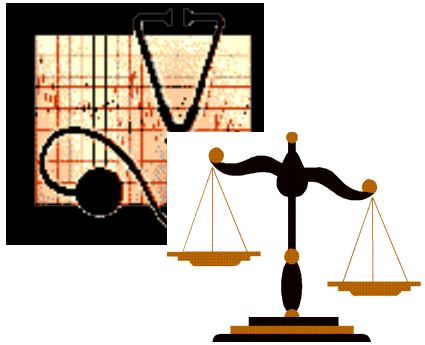


The Idea of Facets

- 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?

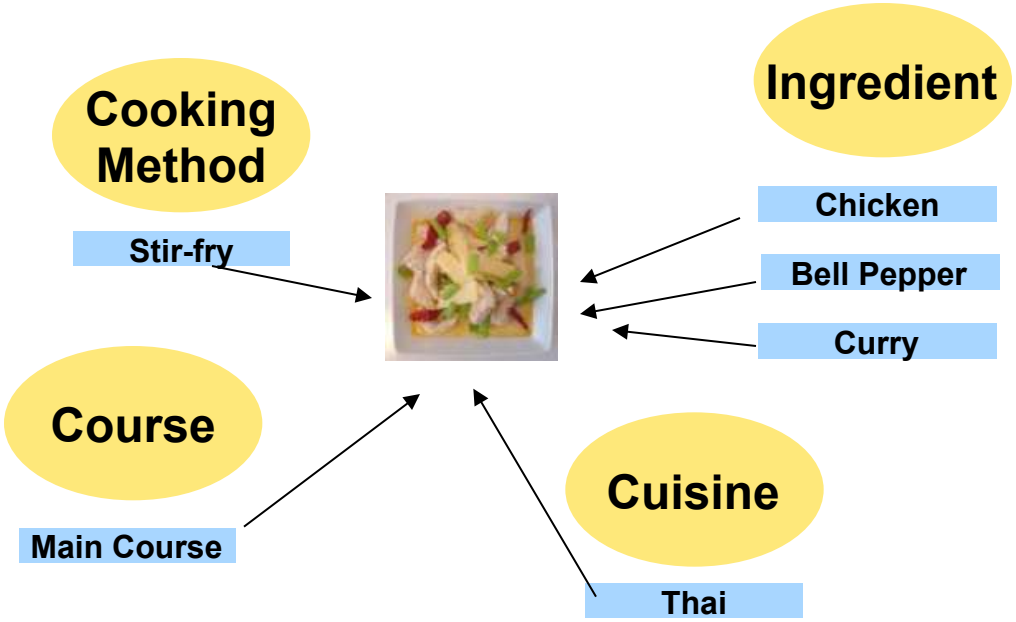
- Sets of categories, each of which describe a different aspect of the objects in the collection.
- Each of these can be hierarchical.
- (Not necessarily mutually exclusive nor exhaustive, but often that is a goal.)



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

Preparation Method
- Fry
- Saute
- Boil
- Bake
- Broil
- Freeze

Desserts
- Cakes
- Cookies
- Dairy
  - Ice Cream
  - Sorbet
  - Flan

Fruits
- Cherries
- Berries
  - Blueberries
  - Strawberries
- Bananas
- Pineapple

Using Facets

- Now there are multiple ways to get to each item

Preparation Method
- **Fry**
- **Saute**
- **Boil**
- **Bake**
- **Broil**
- **Freeze**

Desserts
- **Cakes**
- **Cookies**
- **Dairy**
  - **Ice Cream**
  - **Sherbet**
  - **Flan**

Fruits
- **Cherries**
- **Berries**
  - **Blueberries**
  - **Strawberries**
- **Bananas**
- **Pineapple**

Fruit > Pineapple
Dessert > Cake
Preparation > Bake

Dessert > Dairy > Sherbet
Fruit > Berries > Strawberries
Preparation > Freeze

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

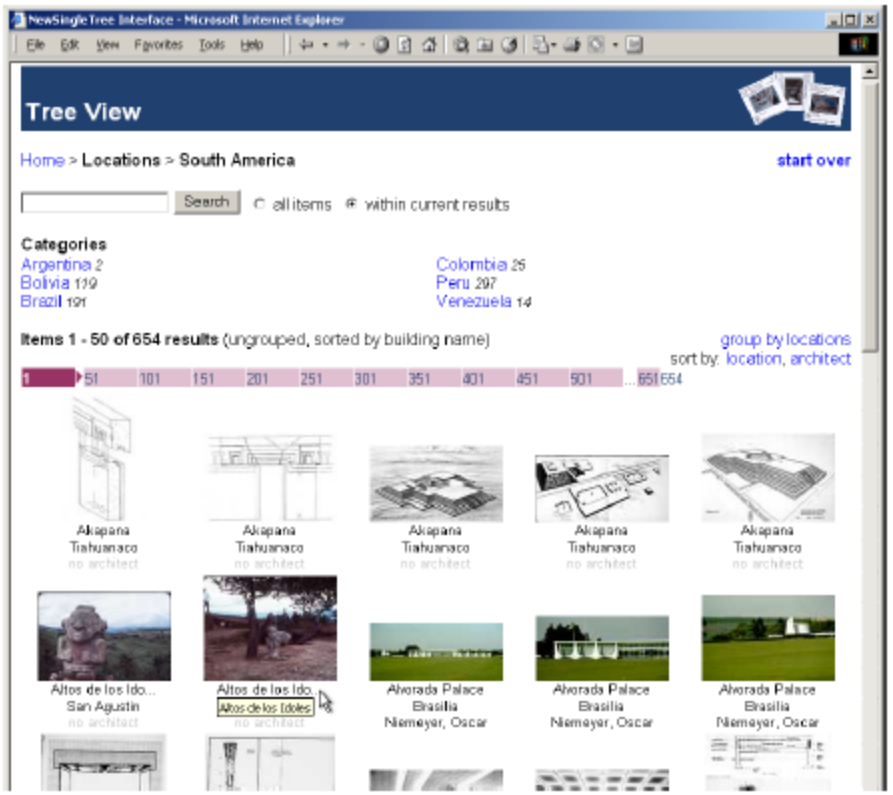
- Can’t end up with empty results sets
  - (except with keyword search)
- Helps avoid feelings of being lost.
- Easier to explore the collection.
  - Helps users infer what kinds of things are in the collection.
  - Evokes a feeling of “browsing the shelves”
- Is preferred over standard search for collection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

Aquamarine
Red
Orange

Door
Doorway
Wall

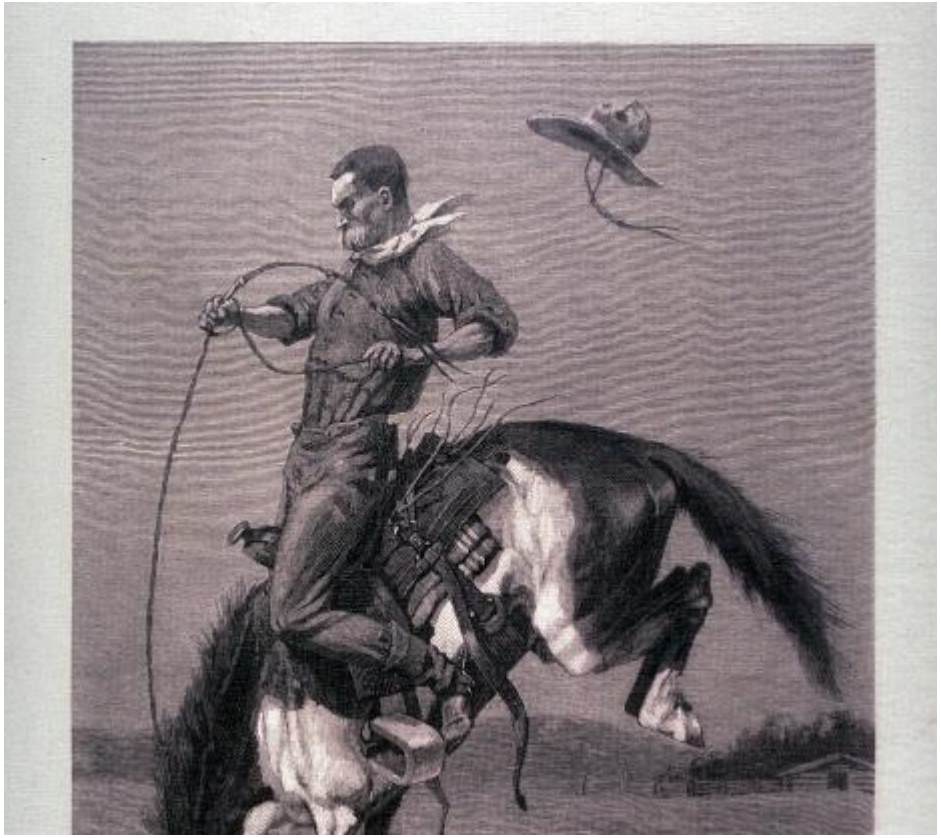
- Which color associated with which object?

Photo by J. Hearst, jhearst.typepad.com

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

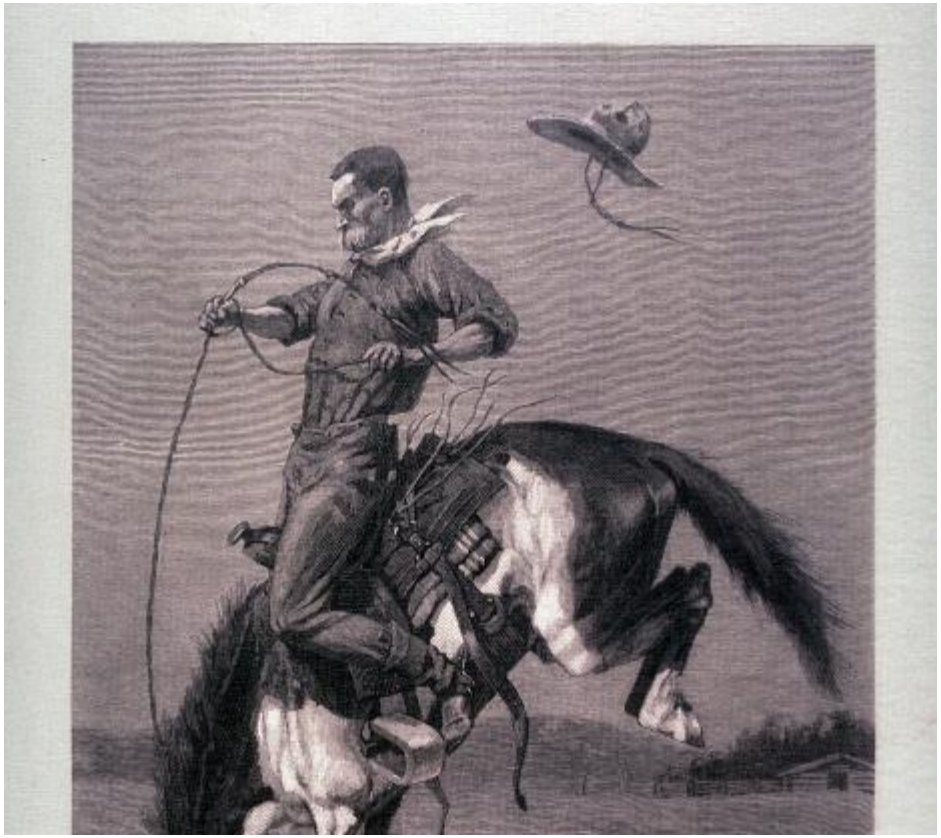


Clothing
Hats
Cowboy Hat

Media
Engraving
Wood Eng.

Location
North America
America

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?

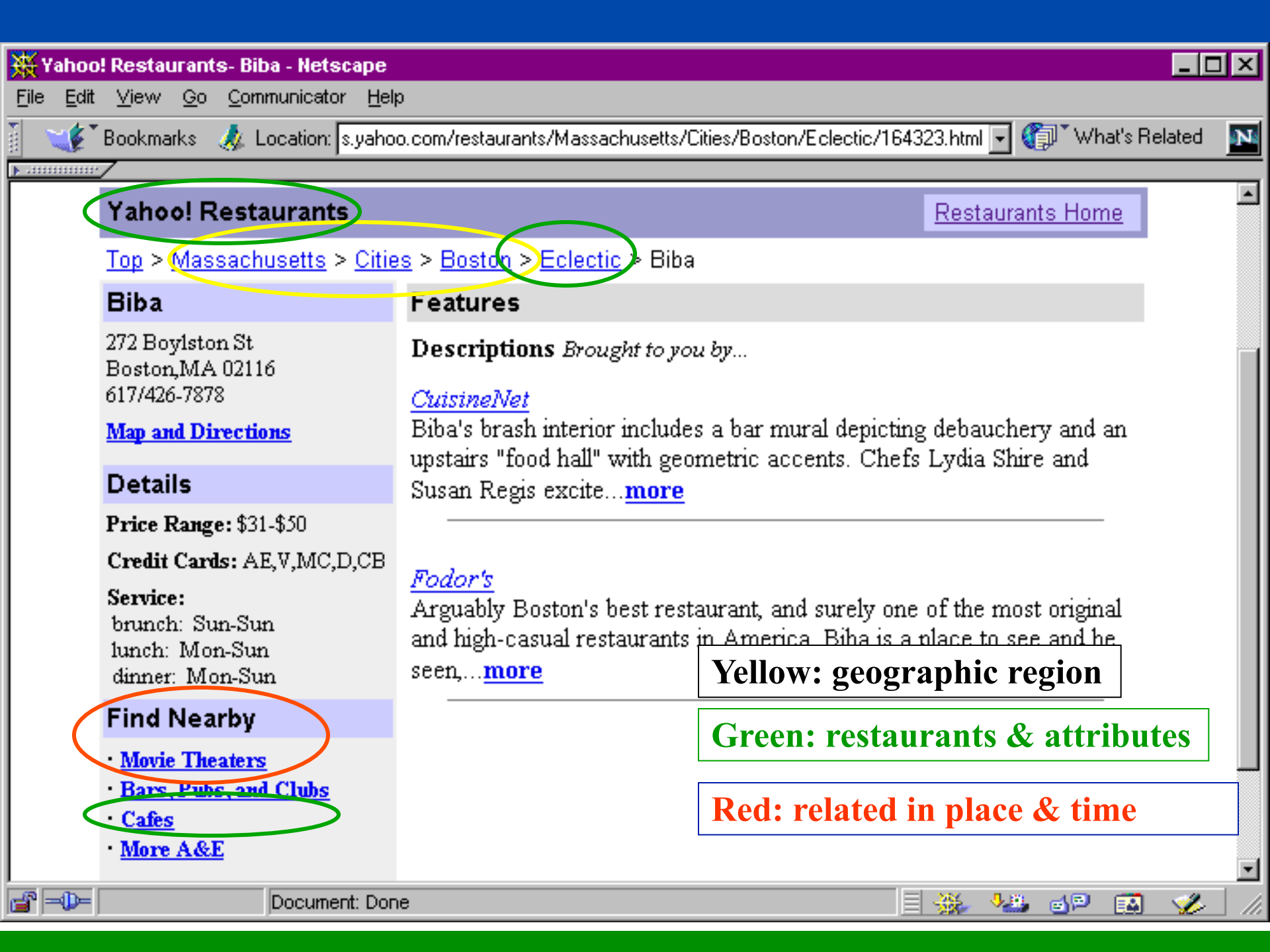
The hat flew off;
The bandana stayed on.

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).

Linking Metadata Into Tasks

- Old Yahoo restaurant guide combined:
  - Region
  - Topic (restaurants)
  - Related Information
    - Other attributes (cuisines)
    - Other topics related in place and time (movies)


Yahoo! Restaurants- Biba - Netscape
File Edit View Go Communicator Help
Bookmarks Location: s.yahoo.com/restaurants/Massachusetts/Cities/Boston/Eclectic/164323.html What's Related
Yahoo! Restaurants
Restaurants Home
Top > Massachusetts > Cities > Boston > Eclectic > Biba
Biba
272 Boylston St
Boston,MA 02116
617/426-7878
Map and Directions
Details
Price Range: $31-$50
Credit Cards: AE,V,MC,D,CB
Service:
brunch: Sun-Sun
lunch: Mon-Sun
dinner: Mon-Sun
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
Yellow: geographic region
Green: restaurants & attributes
Red: related in place & time
Document: Done

Other Possible Combinations

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

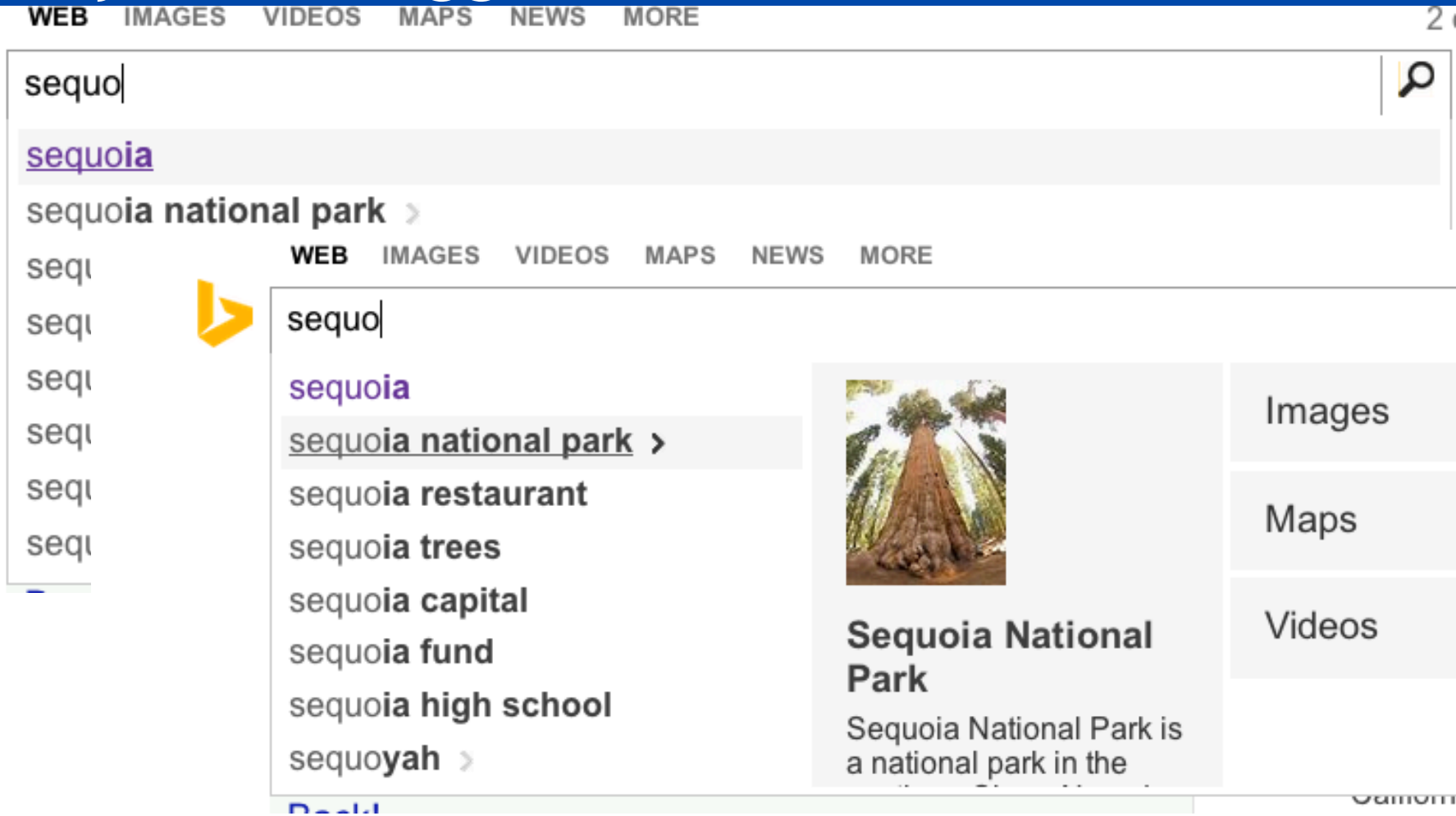
Creating Tasks from HFM

- 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.

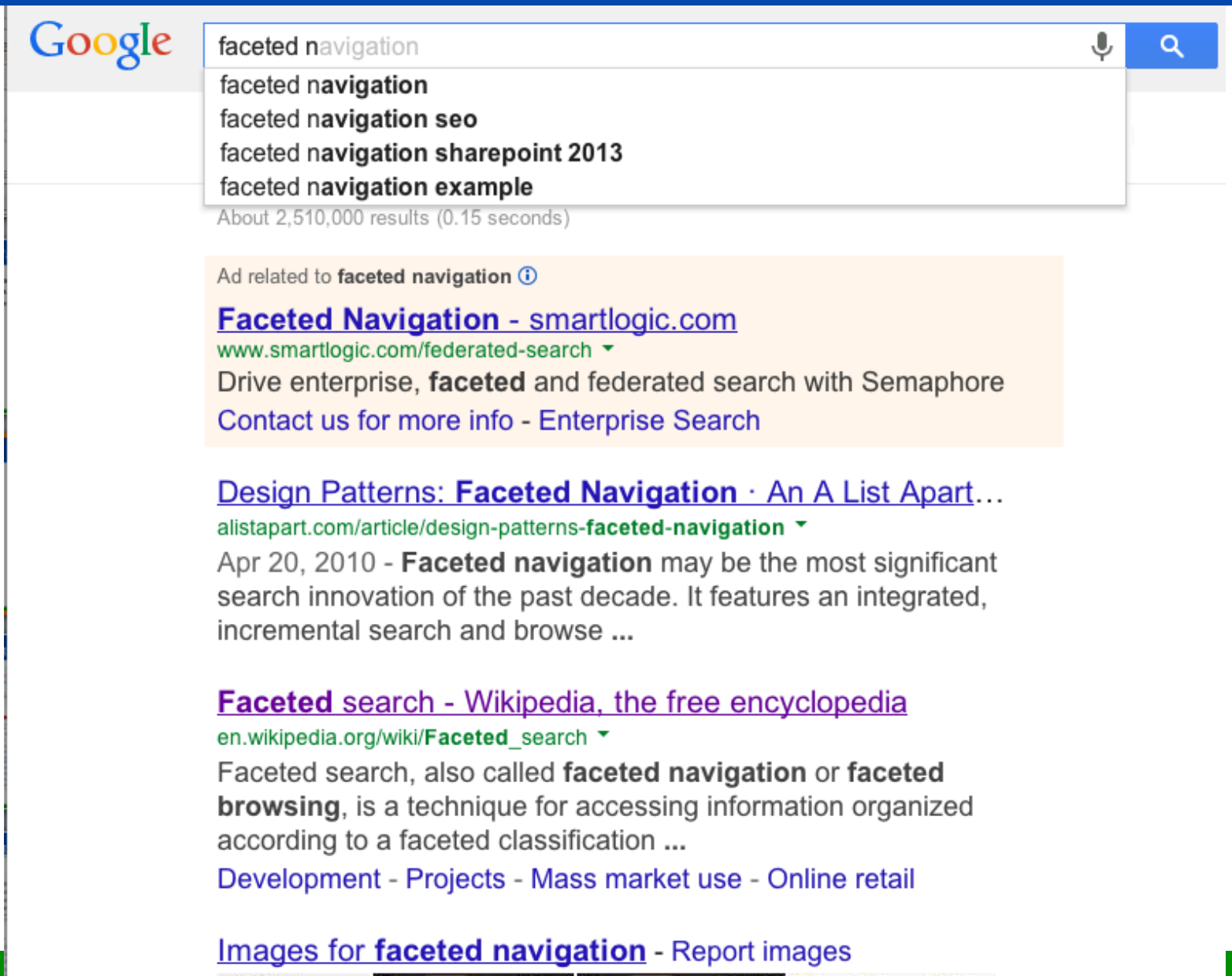
Taking Facets to the Next Level

- 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.

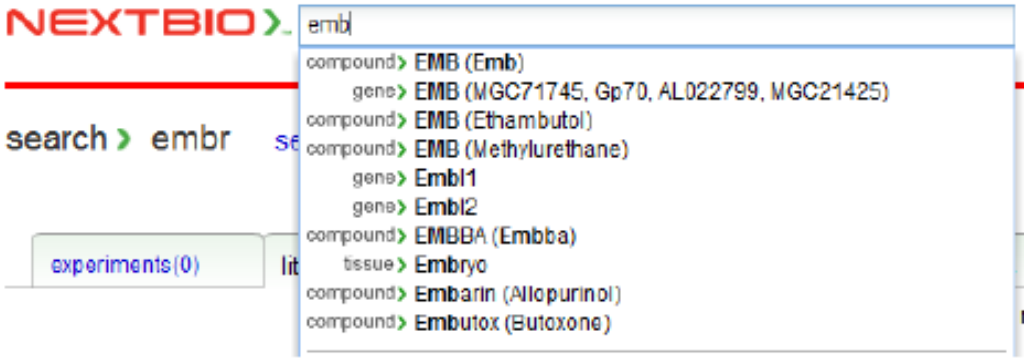
Query Auto-suggest

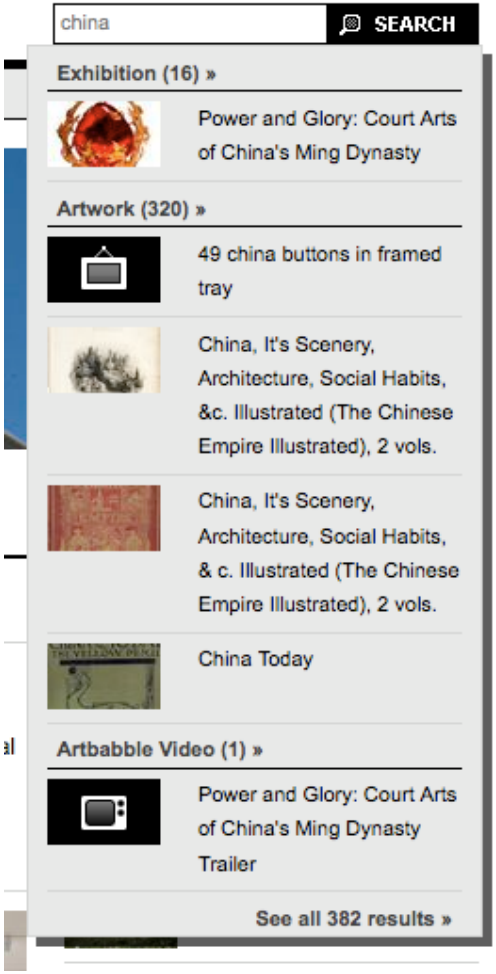


Query auto-suggest



Auto-suggest at the Next Level

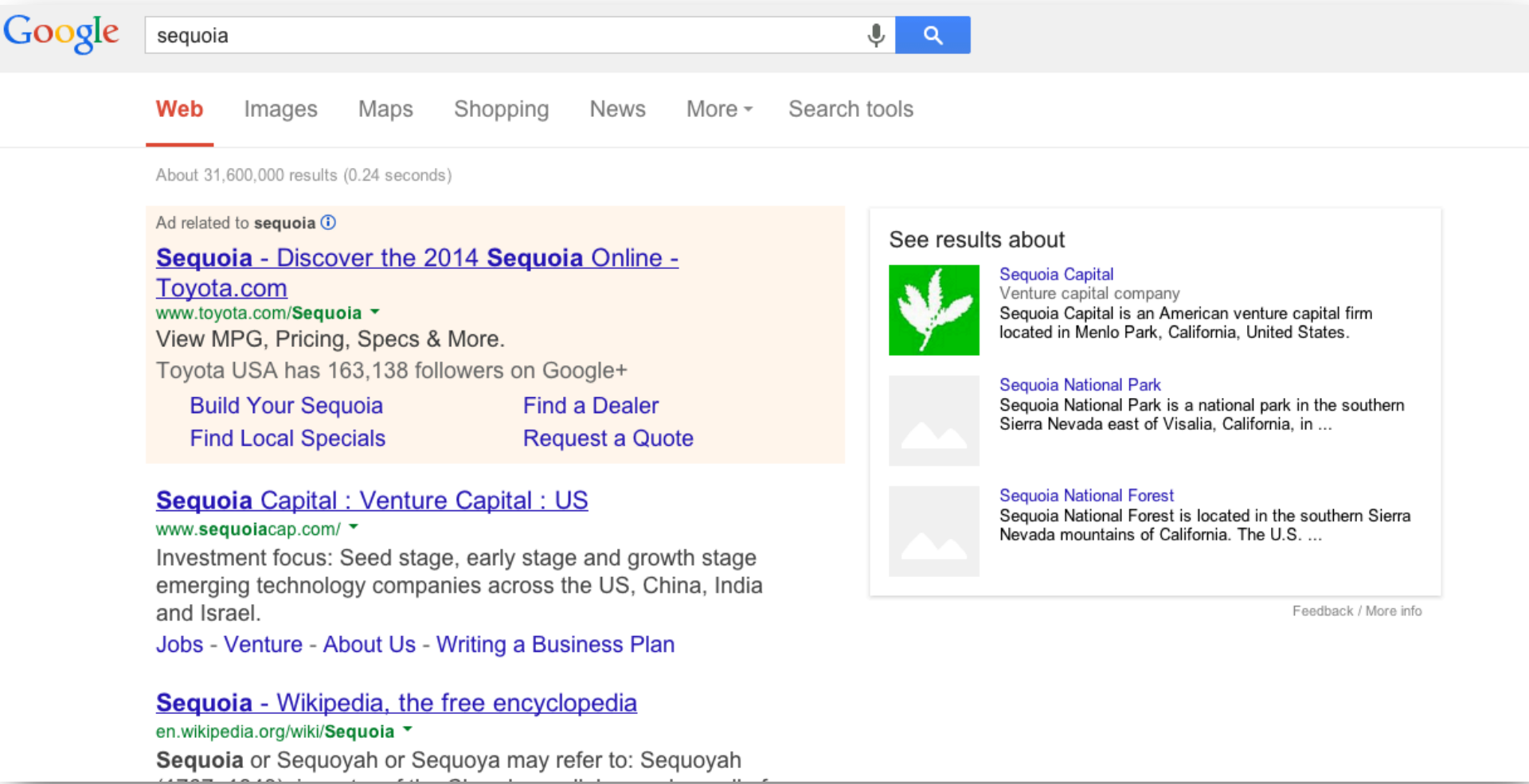




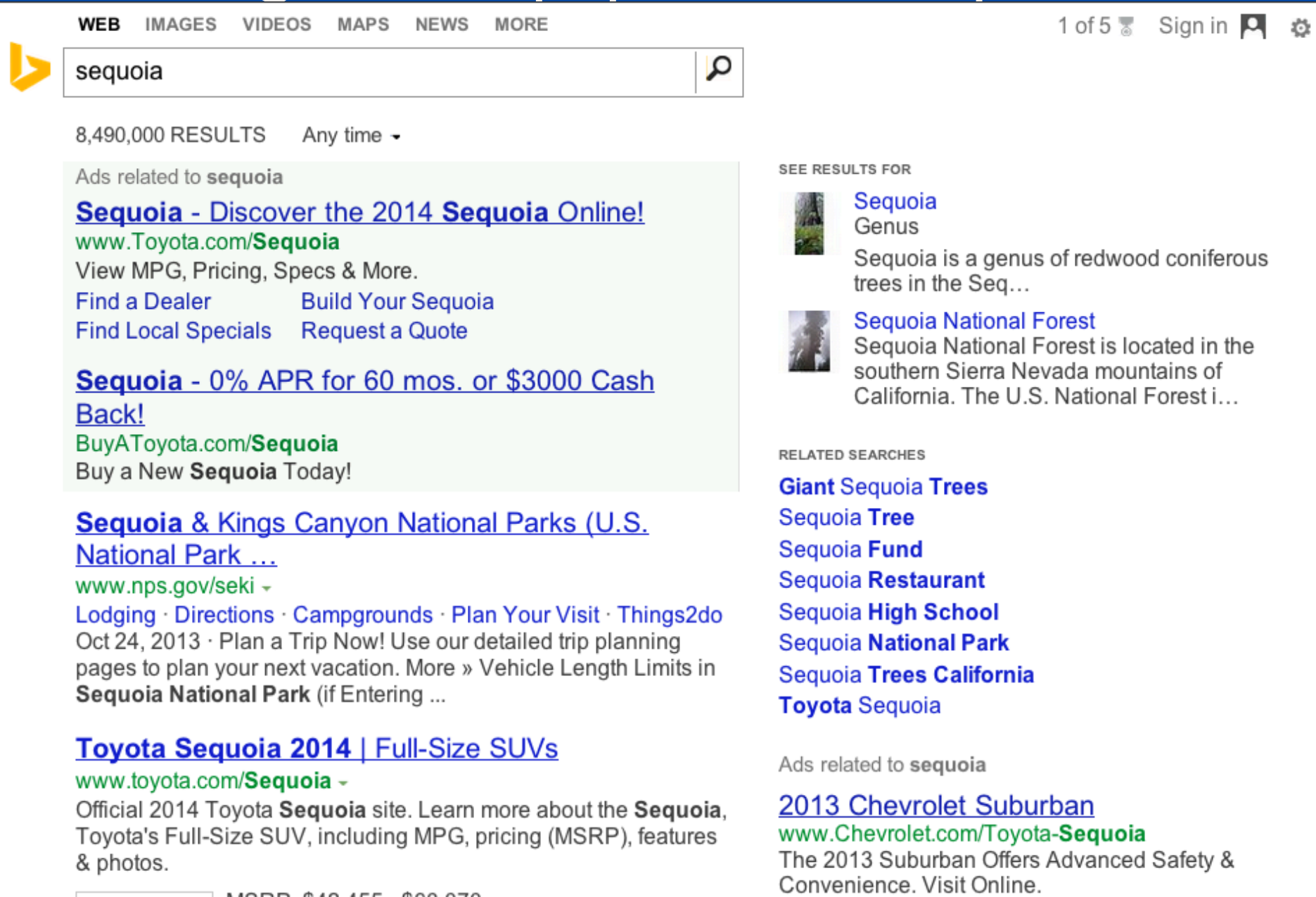
Taking Auto-Suggest to the Next Level

- 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



Knowledge base: popular head queries



Knowledge Base: Relevant Information

- 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.

Knowledge Base Results

Google albert einstein

Web Images Maps Shopping Videos More Search tools

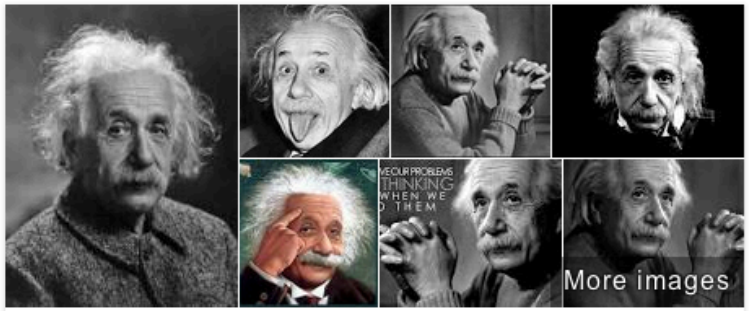
About 30,400,000 results (0.38 seconds)

Albert Einstein | Albert Einstein Official Site
einstein.biz/
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
en.wikipedia.org/wiki/**Albert_Einstein**
Albert Einstein (/ˈælbərt ˈaɪnstaɪn/; German: [ˈalbɐt ˈaɪnʃtaɪn] (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 - Nobelprize.org
www.nobelprize.org/nobel_prizes/physics/laureates/.../**einstein**-bio.html
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…



Albert Einstein

Theoretical Physicist

Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, one of the two pillars of modern physics.
Wikipedia

Born: March 14, 1879, Ulm, Germany
Died: April 18, 1955, Princeton, NJ
Children: Eduard Einstein, Hans Albert Einstein, Lieserl Einstein
Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)
Education: University of Zurich (1905), ETH Zurich (1901), Aargau Cantonal School (1895–1896), Luitpold Gymnasium
Awards: Nobel Prize in Physics, Copley Medal, Franklin Medal, More

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

Query and Behavior Log Analysis

- 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).

Association Rule		
{Movie, Art, "Robocop remake"} ⟹ {DSAT}		
No.	**Query**	**Label**
1	Robocop remake poster	DSAT
2	pics from Robocop remake	DSAT

Playing by the Rules: Mining Query Associations to Predict Search Performance, Kim, Hassan, White, Wang, WSDM 2013.

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

image of cowboy

image of cowboy
image of cowboy **hat**
image of cowboy **boots**
image of cowboy **hat and boots**

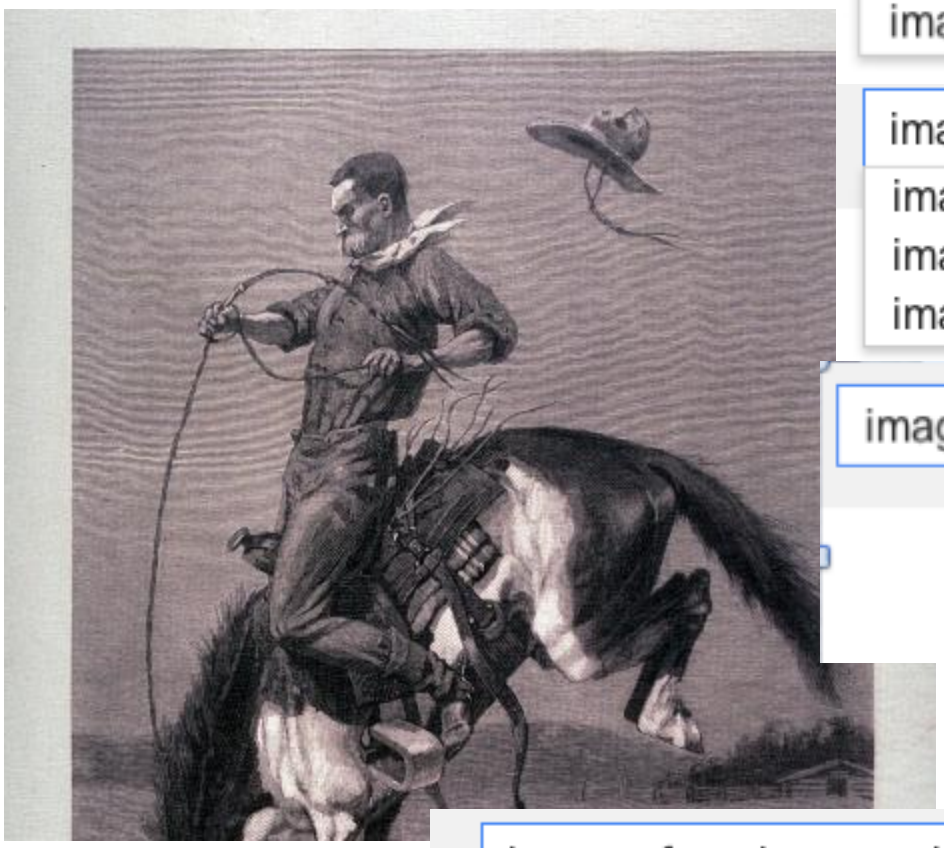


image of cowboy on a

image of a cowboy on a **horse**
image cowboy a **colorier**
image cowboy a **cheval**

image of cowboy on a bucking bron

bron**co**
bron**co chandelier**
bron**ze**

image of cowboy on a bucking bronco with his hat flying off

Press Enter to search.

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

Thank you!

Marti Hearst, UC Berkeley

October 28, 2013