Faceted Metadata for Information Architecture and Search

Marti A. Hearst
School of Information
UC Berkeley
hearst@sims.berkeley.edu

Preston Smalley
eBay User Experience & Design
psmalley@ebay.com

Corey Chandler
eBay User Experience & Design
cchandler@ebay.com

CHI 2006 Course

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 06/0004.

The Association for Computing Machinery
1515 Broadway
New York, New York 10036

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

Printed in the USA

Contents

Course Agenda

Session I

Introduction and Goals

Faceted Metadata
- Definition
- Advantages

Interface Design Using Faceted Metadata
- The Chess Analogy
- The Nobel Prize Example
- Results of Usability Studies
- Software Tools

Design Issues

Discussion

Session II

Highlights from Session I

Interactive Exercise

Evolution of Information Architecture at eBay

Demo of latest eBay design

Lessons learned at eBay

Discussion

Instructor Biographical Sketches

Marti Hearst

Dr. Marti Hearst is an associate professor in the School of Information Management at UC Berkeley, with an affiliate appointment in the Computer Science Division. Her primary research interests are user interfaces and visualization for information retrieval, empirical computational linguistics, and text data mining. She received BA, MS, and PhD degrees in Computer Science from the University of California at Berkeley, and she was a Member of the Research Staff at Xerox PARC from 1994 to 1997. Prof. Hearst is on the editorial boards of ACM Transactions on the Web and ACM Transactions on Computer-Human Interaction, and was formerly on the boards of Computational Linguistics, ACM TOIS, and IEEE Intelligent Systems, and was the program co-chair of HLT-NAACL '03 and SIGIR '99. She has received an NSF CAREER award, an IBM Faculty Award, an Okawa Foundation Fellowship, and two student-initiated Excellence in Teaching awards.

Incorporating faceted hierarchical metadata has gained considerable traction and acceptance in the information architecture community in the past few years, spurred in part by the writings about and demos of the Flamenco project led by Professor Hearst. The approach has been refined in a series of usability studies by her group, supported by a grant from the National Science Foundation (for which Hearst was PI). The Flamenco project has been described in the CHI proceedings by one full paper and one late-breaking paper.

Preston Smalley

Preston Smalley is familiar with the challenges faced in design eCommerce sites and is tasked with designing the search products which enable buyers to access eBay's growing marketplace of over 20 million items for sale with nearly that many visitors a day. As Lead UI Designer, Smalley has directed the evolution of eBay's search offerings for the past 2 years. He played a key role in building support for the investment in improving faceted metadata search and continues to drive the product strategy for the area. He's also explored the use of tagging thru folksonomies with the launch of eBay's Reviews & Guides feature. Smalley contributed to CHI 2005 on his work "Creating a System to Share User Experience Best Practices at eBay" with his colleague Jeff Herman. Finally, he has five patents pending for his innovations in faceted search, search error recovery, and reputation systems while at eBay.

Corey Chandler

Corey Chandler has been a User Interface Designer at eBay for over two years. He has been involved with a number of key projects, most notably designing the search interface for eBay Express, which is eBay's first implementation of faceted search. Before joining eBay, Chandler worked with Microsoft's Meeting and Presentation Services group. Chandler has a degree in Cognitive Science from UC Berkeley, where he worked closely with the Group for User Interface Research to publish his student poster, "Low-fidelity Prototyping for Multimodal Applications," at CHI 2002. Chandler now has four patents pending based on work at eBay.

Course Goals

The purpose of this course is to introduce and explain a systematic approach to designing information architecture for web sites consisting of large collections of items. The main goals of the approach are to produce websites with multiple different views to reflect differences in user's preferred search and browsing methods, to incorporate search uniformly throughout the design of the site, and to do this in a manner that makes use of standard technology and allows non-experts to be able to add to the content of the site without disturbing the other properties.

The main objective of the course is to instruct attendees about how to integrate navigation and search for large collections in a seamless, flexible manner that helps users find things quickly and browse items comfortably. The course will present a method for developing such interfaces and steer attendees away from common pitfalls. It will also present justifications for the approach in the form of usability results and real-world application of the ideas. Finally, it will point out the limitations of the approach.

This course is intended for practitioners in the field of information architecture and website design, although it should also be of interest to researchers on search user interfaces. The instructors have designed an approachable, reproducible methodology for the design of highly usable, highly searchable information-centric web sites. The emphasis is on sites that provide access to large collections of information, as opposed to sites that focus on performing actions. Examples of appropriate types of sites include ecommerce sites that contain large catalogs of products, government sites that provide information on services and resources for citizens, and sites providing access to large collections of information such as images or medical documents.

The main idea is to take advantage of faceted hierarchical metadata in a systematic way; to specify how to organize this metadata to reflect different user tasks, and how to create a navigation structure that flexibly responds to changes in user's information seeking patterns while at the same time retains consistency in the views of the information structure. Crucial to this approach is the separation of the information structure from the navigation structure in the process of information architecture design.

The methodology arises from a view of search as being tightly coupled with browsing. In this worldview, the system should steer the user through a set of next choices rather than requiring specification of complex combinations of terms. It should also allow users to type in arbitrary terms and then organize the results to reflect the underlying information architecture in a way that logically reflects what a user would expect based on the search terms and the kind of content available at the site. It should provide information organizations that reflect user tasks, but without requiring editors to change the content. Finally, it should reflect users' preferences over time, but in a systematic and predictable way. The use of metadata systems and rules linking metadata types to one another helps
to ensure these goals.

This method does not introduce radically new ideas; rather it selectively chooses among existing ideas, motivated by observations of regularities of best practices, and puts them in a systematic, reproducible framework whose parameters are being validated by the application of the results of scientifically conducted usability studies.

Suggested Readings and Presentations

Taxonomies to Tags: From Trees to Piles of Leaves, David Weinberger, *Release 1.0*, 23(2), February 2005. (fee required)
http://www.release1-0.com/release1/abstracts.cfm?Counter=4579726

How to make a faceted classification and put it on the web, William Denton, Nov 2003
http://www.miskatonic.org/library/facet-web-howto.pdf

Faceted Metadata for Image Search and Browsing, Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst, in the proceedings of ACM CHI 2003. (enclosed in these notes)

Finding the Flow in Web Site Search, Marti Hearst, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ping Yee, Communications of the ACM, 45 (9), September 2002, pp.42-49. (enclosed in these notes)

Clustering vs. Faceted Categories for Information Exploration, Marti Hearst, to appear in CACM. (enclosed in these notes)

Faceted Navigation: Best of Browse and Search, Tom Reamy, KMWorld - November, 2005 (ppt, this talk has links to many other relevant resources)
http://www.kapsgroup.com/presentations.shtml

Suggested Website Collections

SearchTools.com's page on Faceted Metadata Search
http://www.searchtools.com/info/faceted-metadata.html

Keith Instone's pages on Faceted Browsing
http://user-experience.org/uefiles/facetedbrowse/

IAWike's page on FacetedClassification:
http://www.iawiki.net/FacetedClassification

Links to Software and Demos

The Flamenco project has demos and open source course available:
http://flamenco.berkeley.edu

FacetMap has demos and software available:
http://facetmap.com

ebay's faceted demo:
http://express.ebay.com

Faceted Metadata for Information Architecture and Search

CHI Course - April 24, 2006
Session I

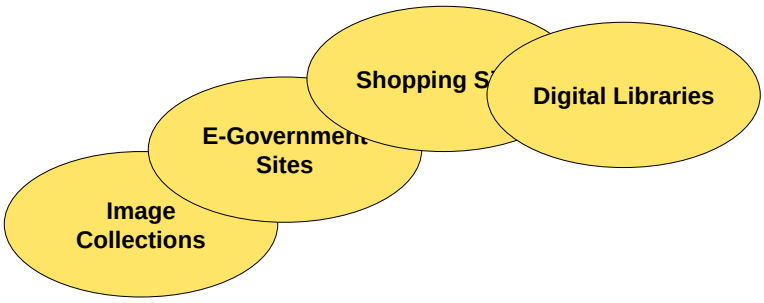
Marti Hearst, School of Information, UC Berkeley
Preston Smalley & Corey Chandler, eBay User Experience & Design

Session I: Agenda

- Intro and Goals (5 min)
- Faceted Metadata (15 min)
  - Definition
  - Advantages
- Interface Design using Faceted Metadata (40 min)
  - The Chess Analogy
  - The Nobel Prize Example
  - Results of Usability Studies
  - Software Tools
- Design Issues (15 min)
- Q&A (15 min)

2

Focus: Search and Navigation of Large Collections



3

Problems with Site Search

- Study by Vividence in 2001 on 69 Sites
  - 70% eCommerce
  - 31% Service
  - 21% Content
  - 2% Community
- Poorly organized search results
  - Frustration and wasted time
- Poor information architecture
  - Confusion
  - Dead ends
  - "back and forthing"
  - Forced to search

4

What we want to Achieve

- Integrate browsing and searching seamlessly
- Support exploration and learning
- Avoid dead-ends, “pogo’ing”, and “lostness”

5

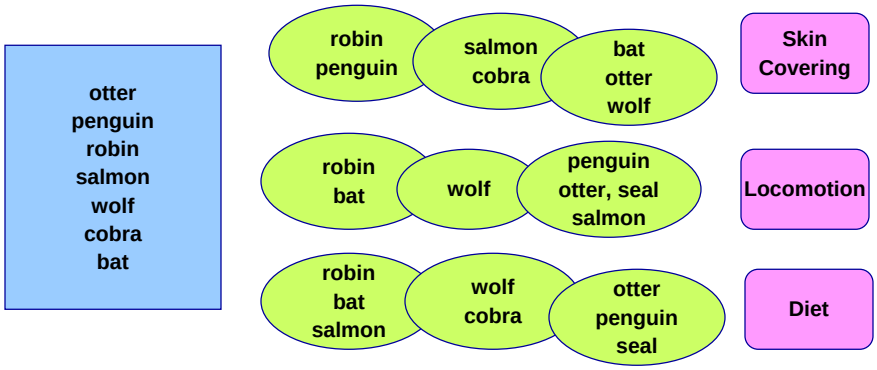
Main Idea

- Use hierarchical faceted metadata
- Design the interface to:
  - Allow flexible navigation
  - Provide previews of next steps
  - Organize results in a meaningful way
  - Support both expanding and refining the search

6

The Problem With Categories

- Most things can be classified in more than one way.
- Most organizational systems do not handle this well.
- Example: Animal Classification



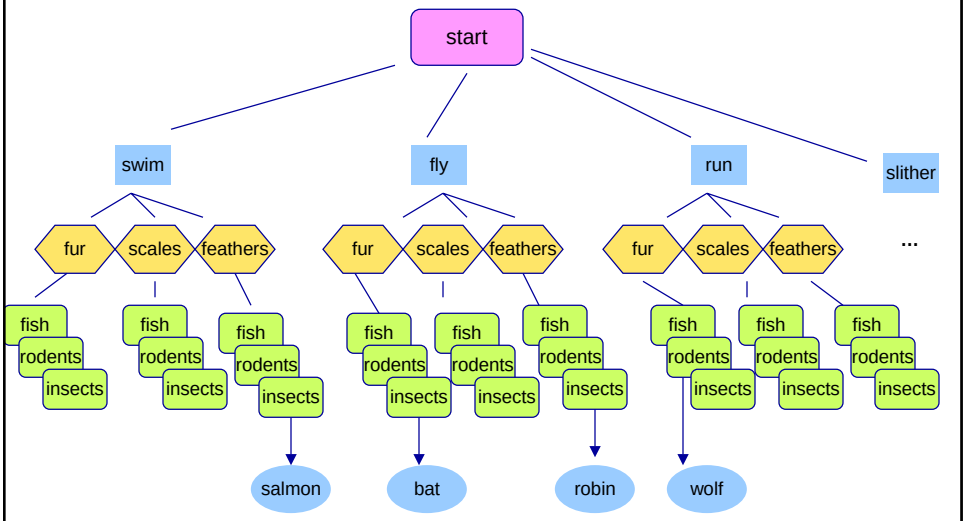
7

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 interace 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

8

The Problem With Hierarchy



9

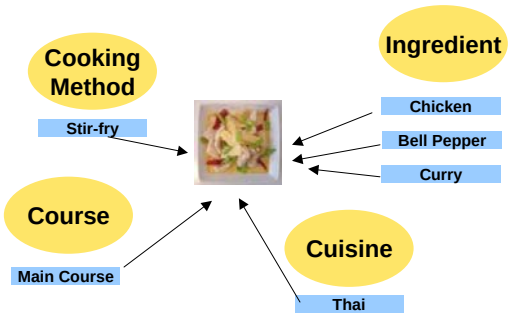
The Idea of Facets

- 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

10

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



11

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

12

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
  - Sorbet
  - Flan

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

Fruit > Pineapple
Dessert > Cake
Preparation > Bake

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

13

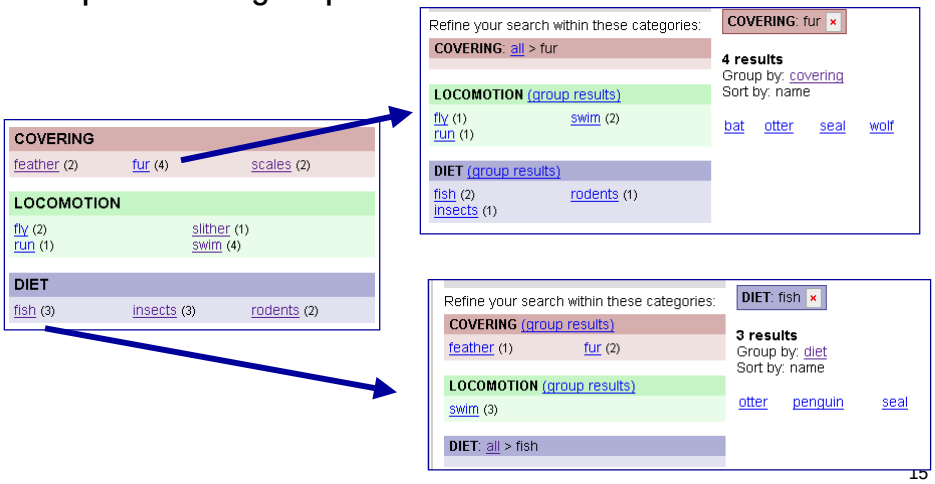
Using Facets

- 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.

14

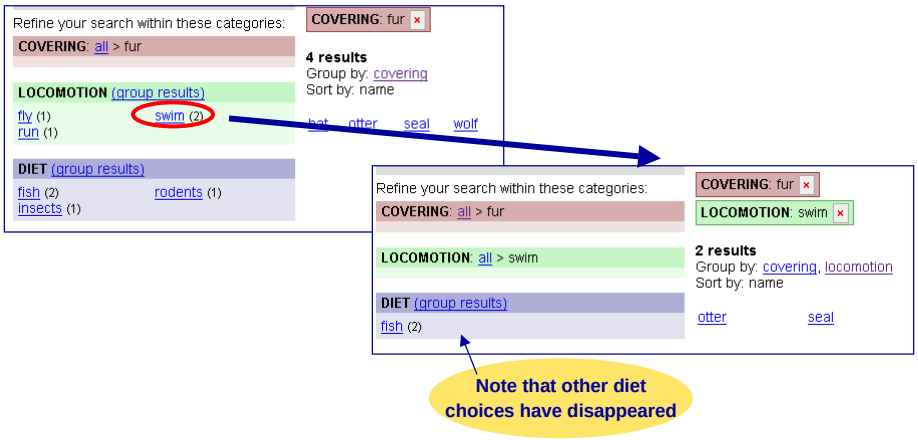
The Advantage of Facets

- Lets the user decide how to start, and how to explore and group.



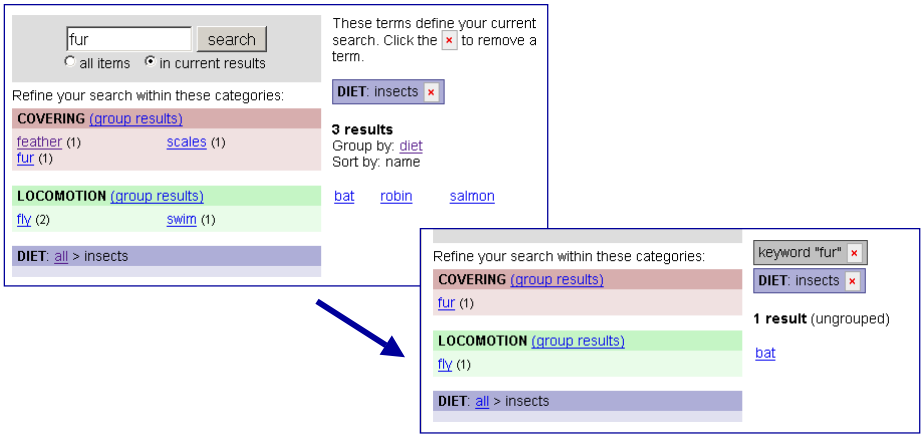
The Advantage of Facets

- After refinement, categories that are not relevant to the current results disappear.



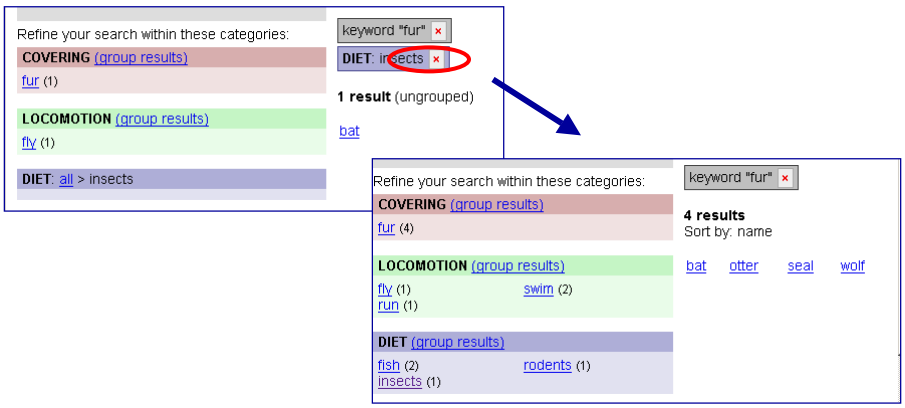
The Advantage of Facets

- Seamlessly integrates keyword search with the organizational structure.



The Advantage of Facets

- Very easy to expand out (loosen constraints)
- Very easy to build up complex queries.



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)

19

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

20

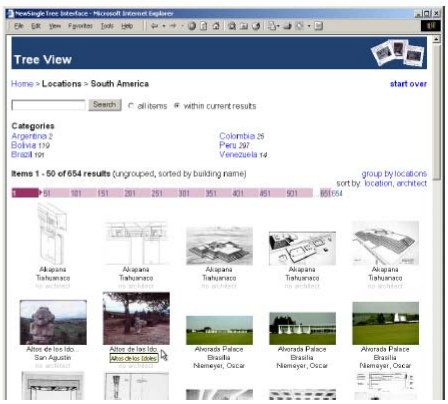
Information previews

- Use the metadata to show where to go next
  - More flexible than canned hyperlinks
  - Less complex than full search
- Help users see and return to previous steps
- Reduces mental work
  - Recognition over recall
  - Suggests alternatives
- More clicks are ok *only if* (J. Spool)
  - The "scent" of the target does not weaken
  - If users feel they are going towards, rather than away, from their target.

21

Facets vs. Hierarchy

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



Limitation of Facets

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

Aquamarine
Red
Orange

Door
Doorway
Wall

- Which color associated with which object?

23

Terminology Clarification

- Facets vs. Attributes
  - Facets are shown independently in the interface
  - Attributes just associated with individual items
    - E.g., ID number, Source, Affiliation
    - However, can always convert an attribute to a facet
- Facets vs. Labels
  - Labels are the names used within facets
  - These are organized into subhierarchies
- Synonyms
  - There should be alternate names for the category labels
  - Currently (in Flamenco) this is done with subcategories
    - E.g., Deer has subcategories “stag”, “faun”, “doe”

24

The Chess Analogy

25

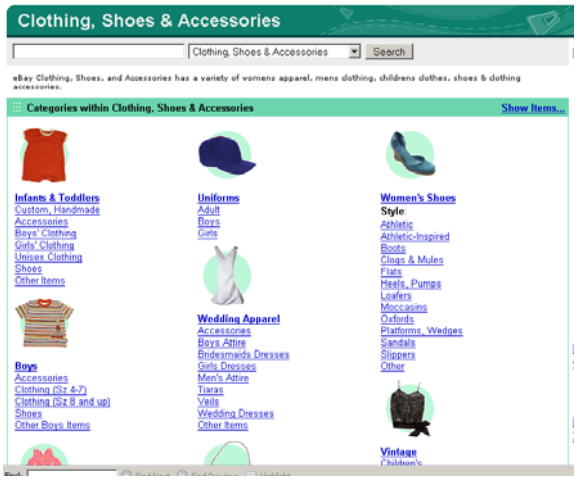
Analogy: Chess

- Chess is characterized by a few simple rules that disguise an infinitely complex game
- The three-part structure of play
  - Openings:
    - many strategies, entire books on this
  - Endgame:
    - well-defined, well-understood
  - Middlegame:
    - nebulous, hard to describe
- Our thought: information navigation has a similar structure, and the middlegame is critically underserved.

26

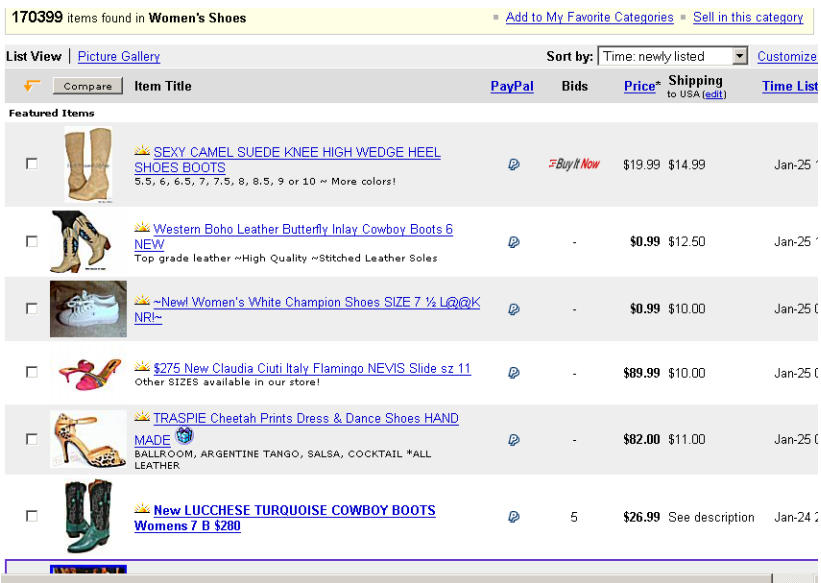
The Opening

- Usually exposes top-level hierarchy or top-level facets
- Usually also has a search component



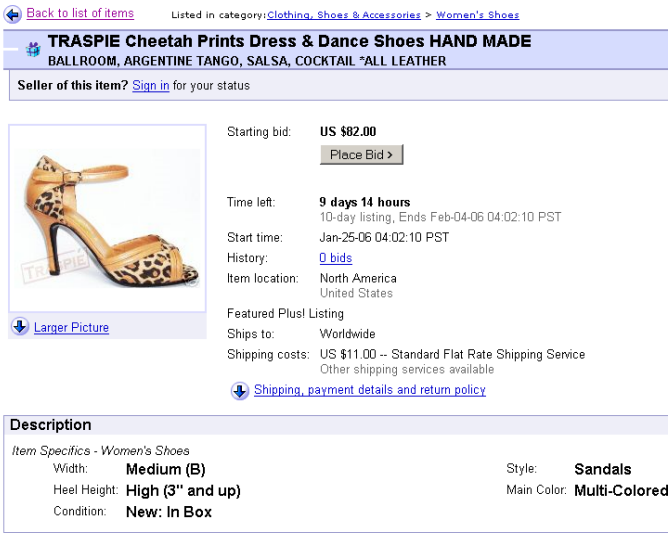
27

The Endgame – Penultimate Pages



28

The Endgame – Content Pages



29

The Middlegame

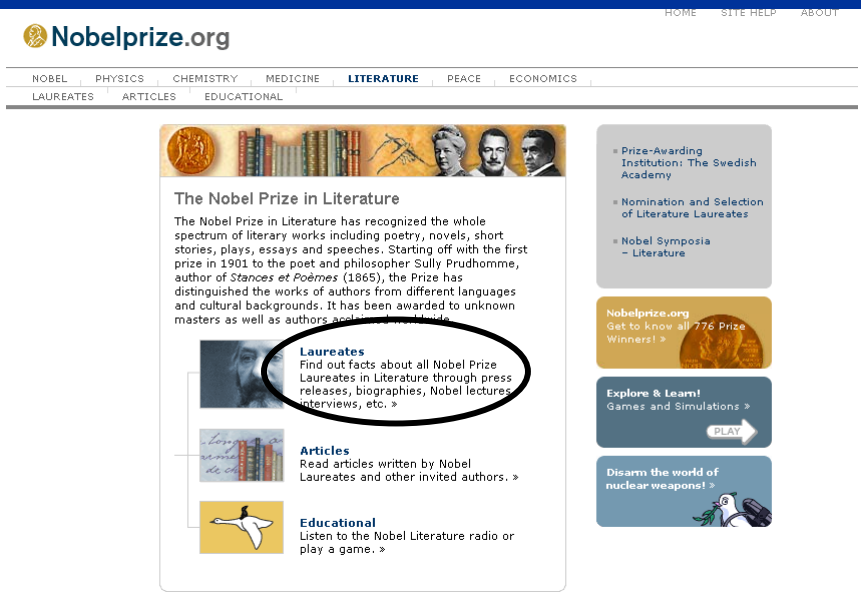
- The heart of the navigation experience
- There is a strategic advantage to having a good middlegame
  - Standard Web search doesn't handle this well
- This is where the flexible faceted metadata approach can work best.

30

Example:
Nobel Prize Winners Collection
(Before and After Facets)

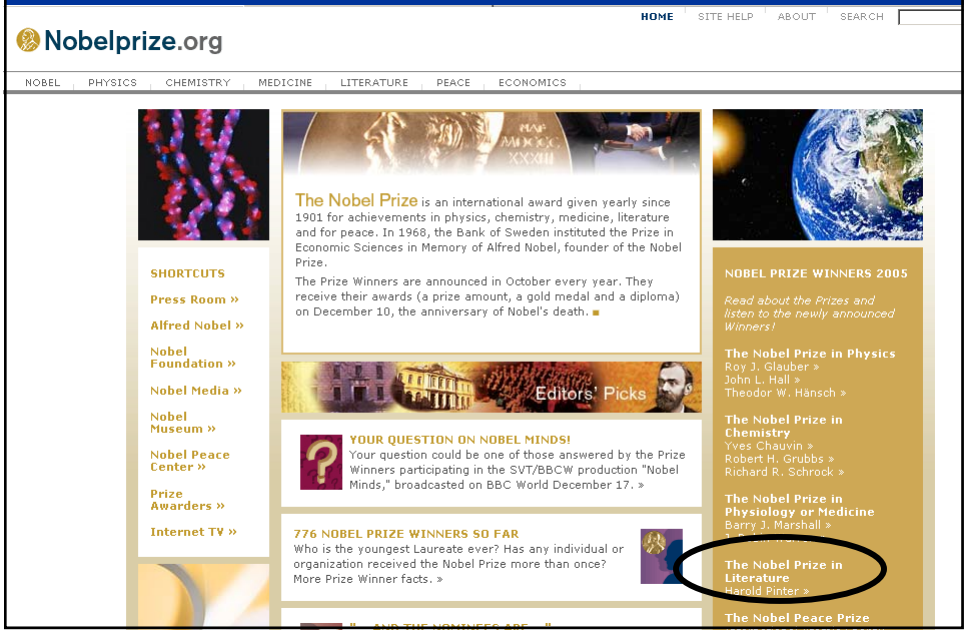
31

Only One Way to View Laureates

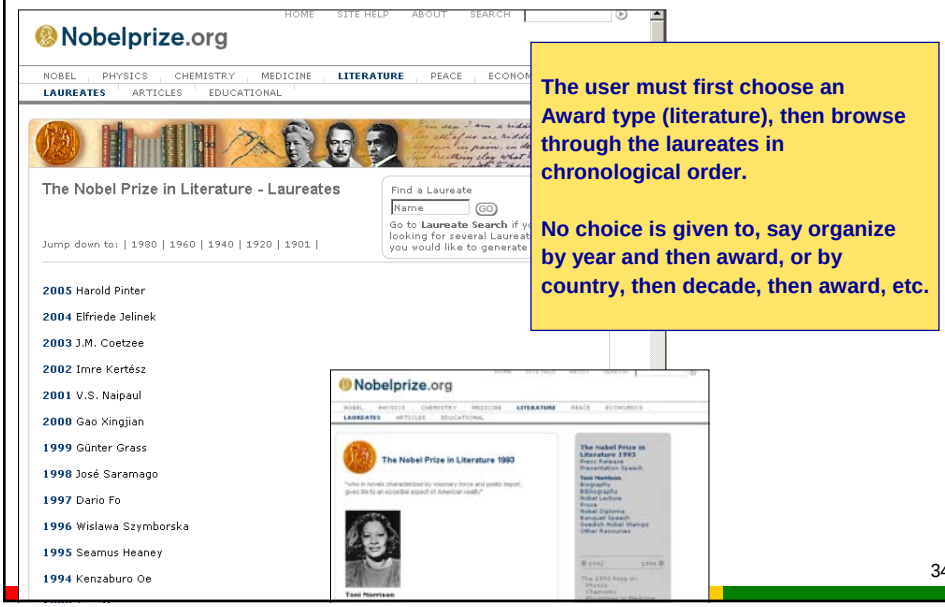


32

First, Choose Prize Type



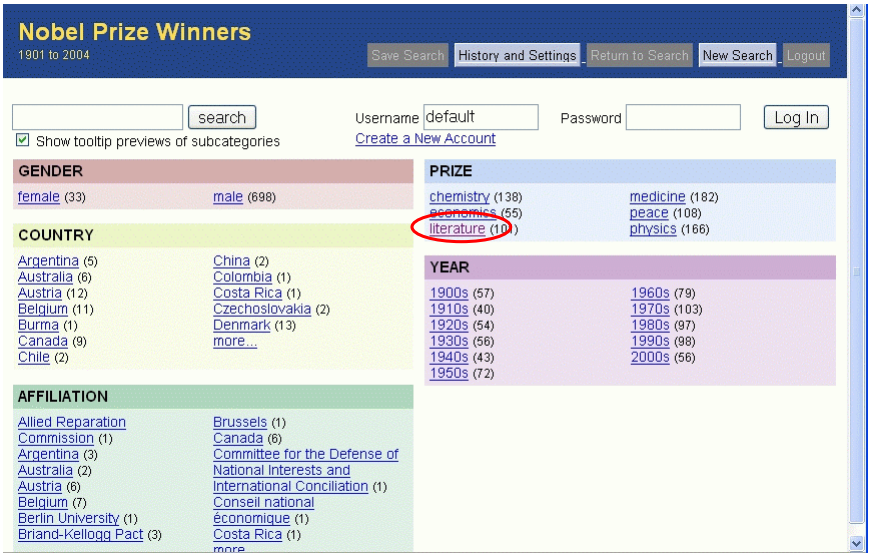
Next, view the list!



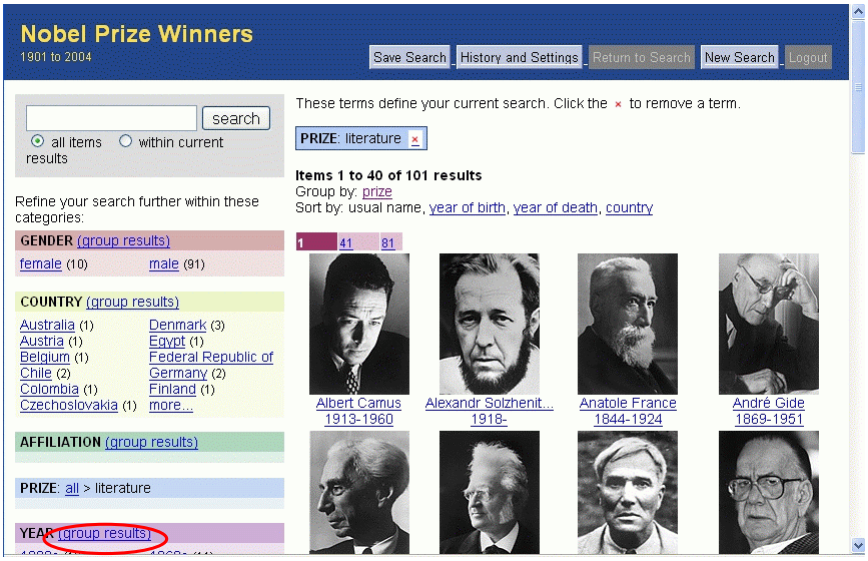
Using Hierarchical Faceted Metadata

35

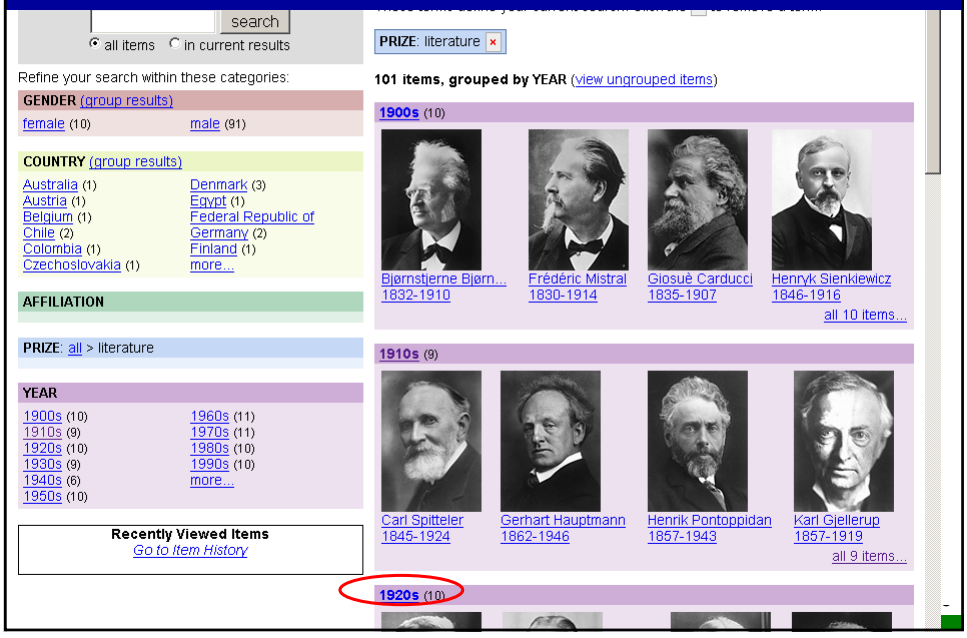
Opening View
Select literature from PRIZE facet



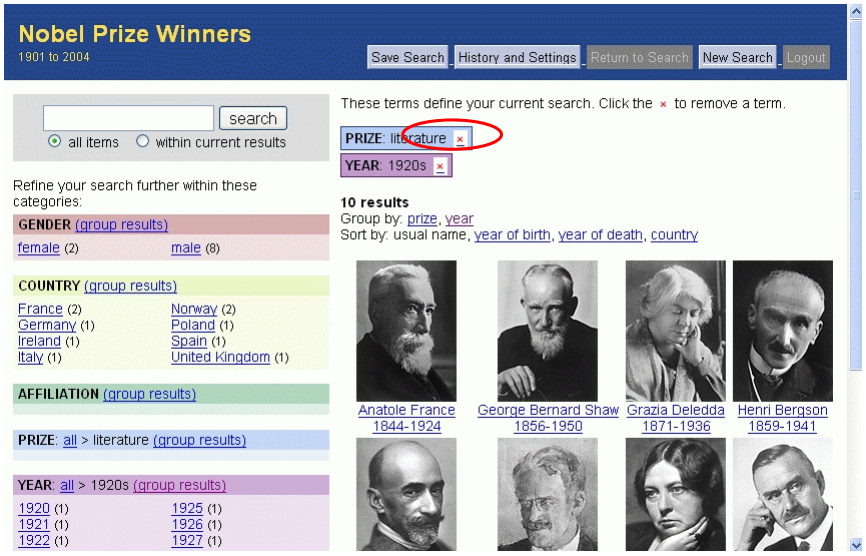
Group results by YEAR facet



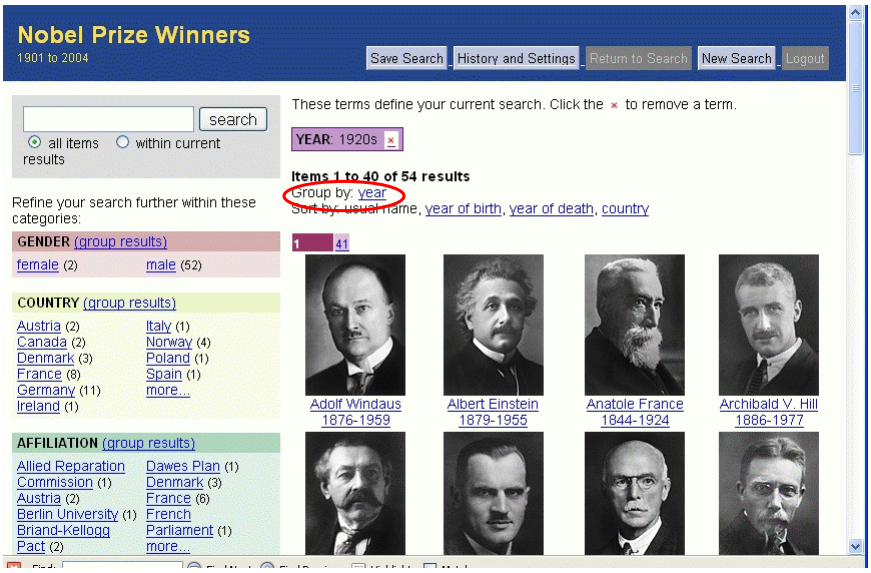
Select 1920's from YEAR facet



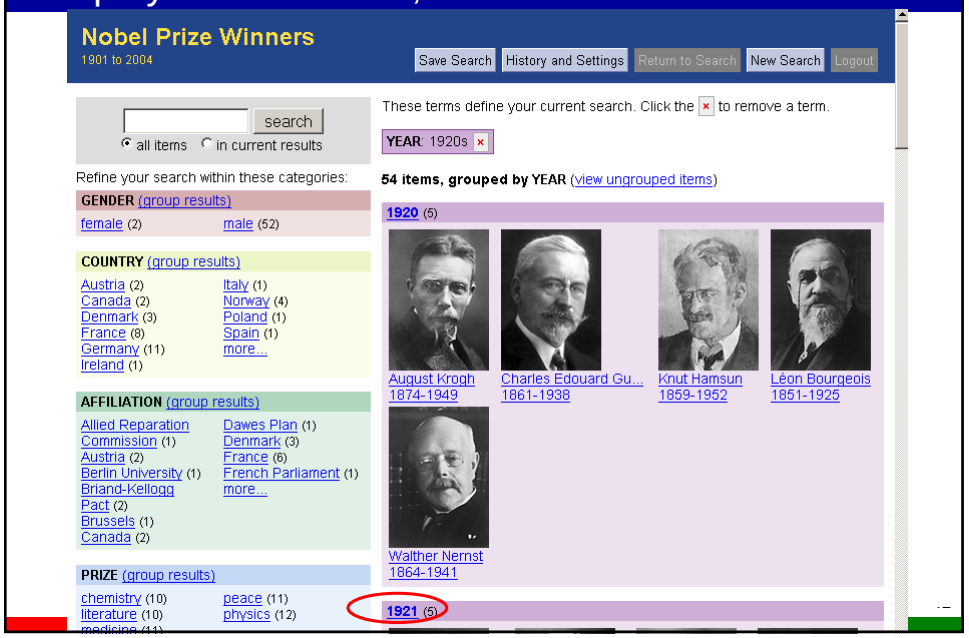
Current query is PRIZE > literature AND YEAR: 1920's. Now remove PRIZE > literature



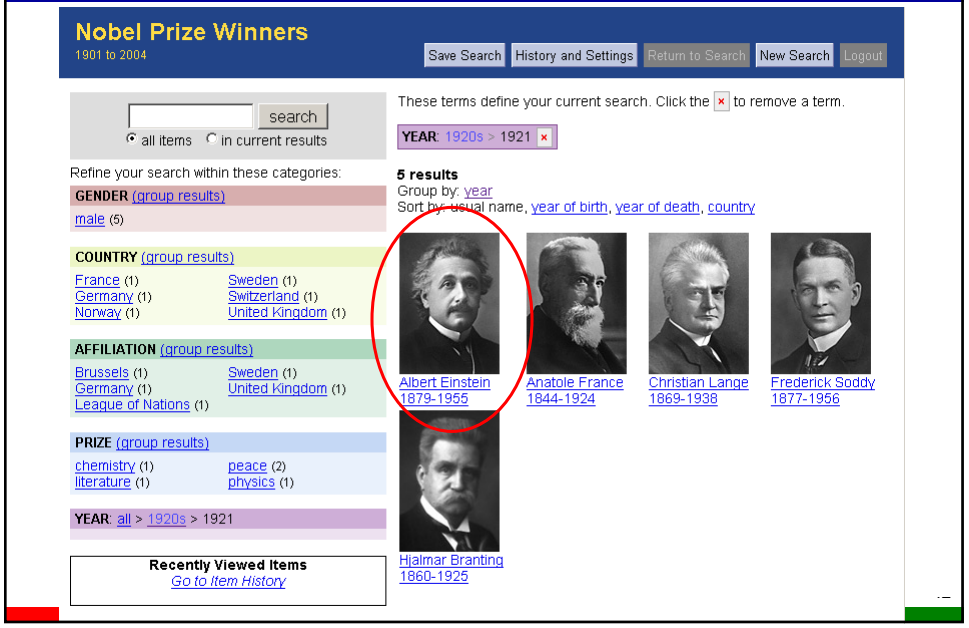
Now Group By YEAR > 1920's



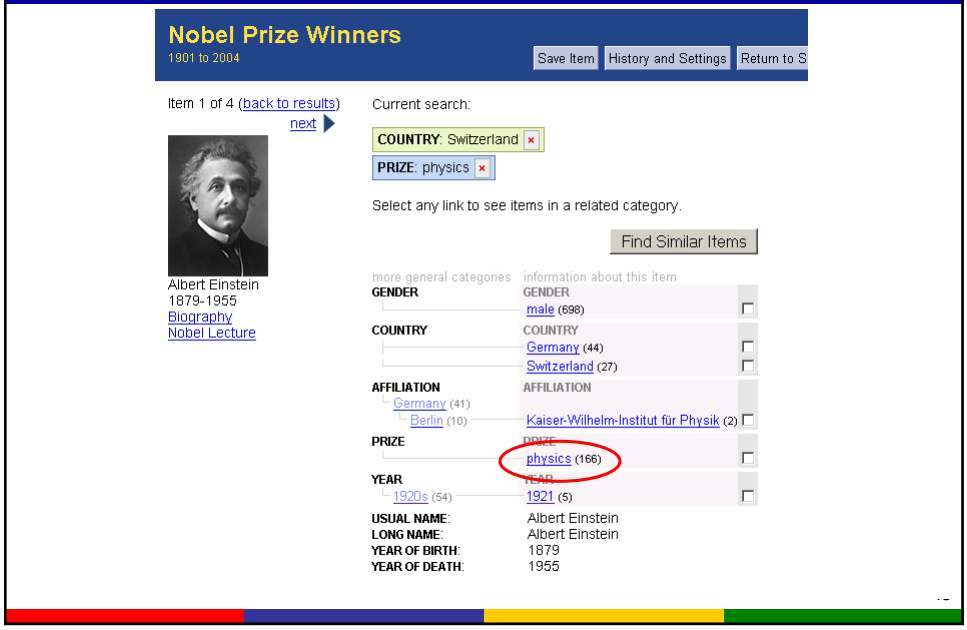
Hierarchy Traversal:
Group By YEAR > 1920's, and drill down to 1921



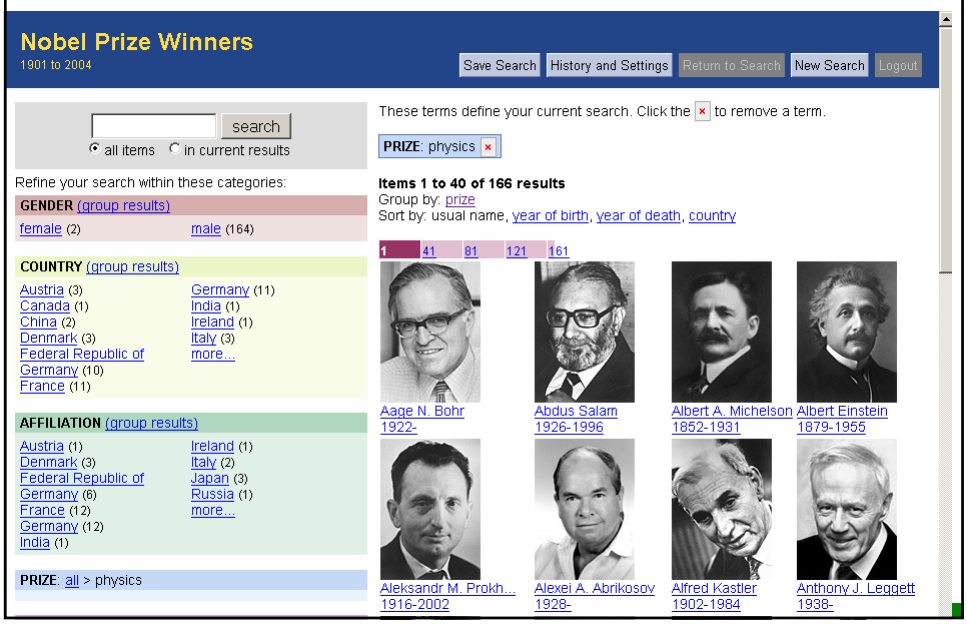
Select an individual item



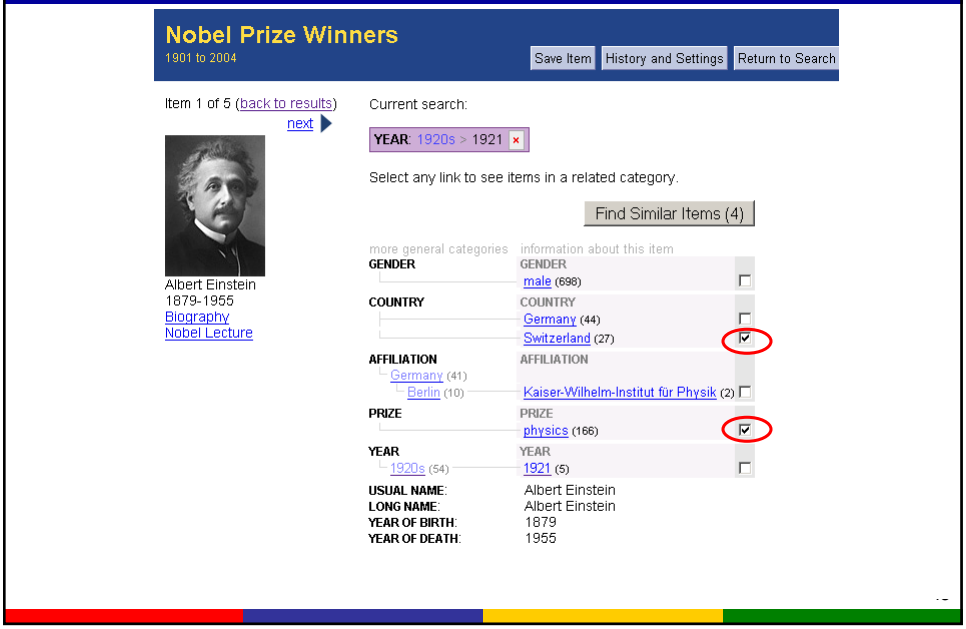
Use Endgame to expand out



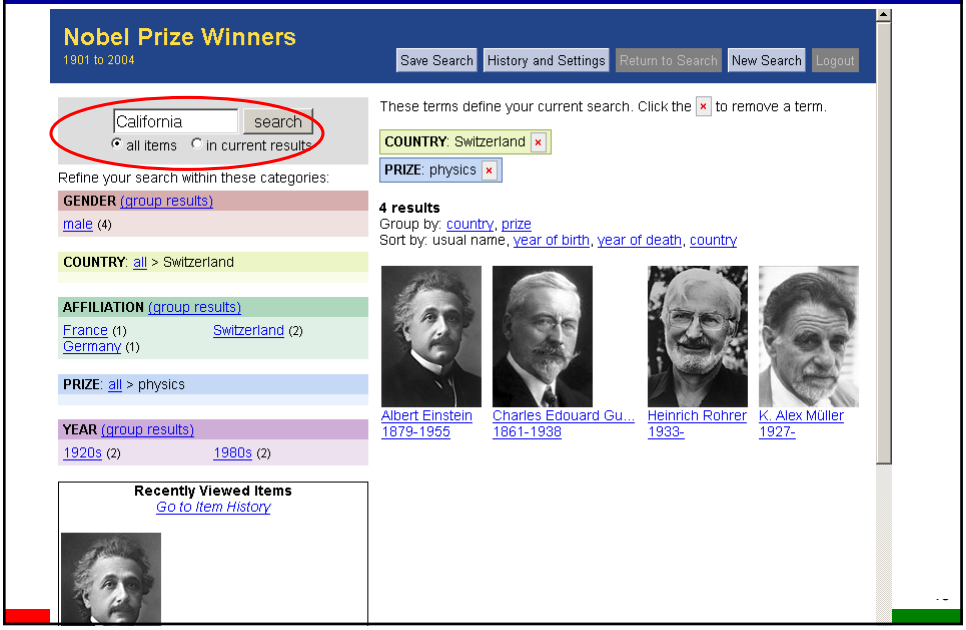
Use Endgame to expand out



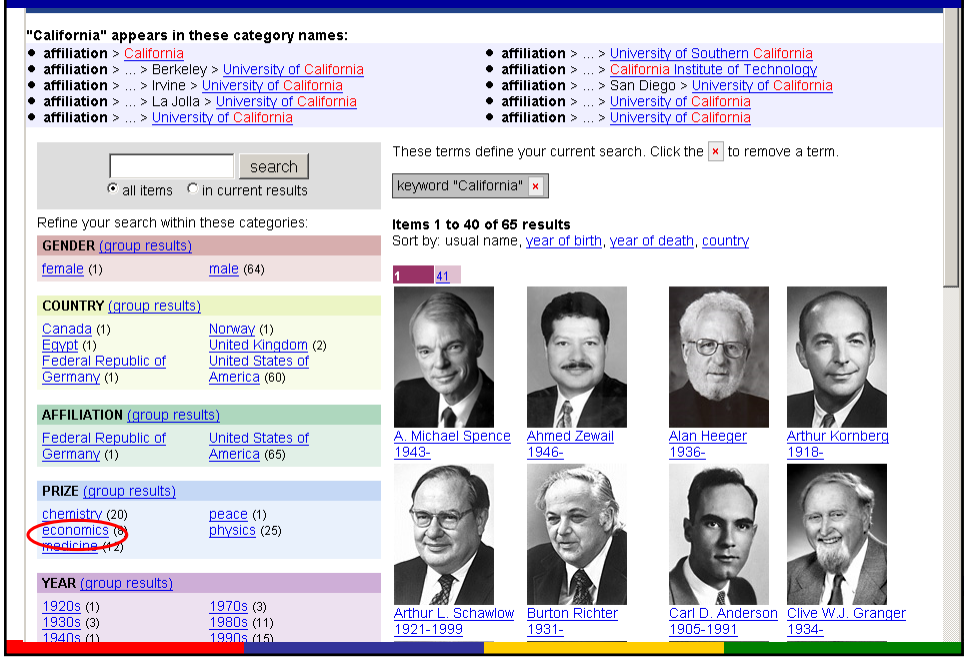
Or use "More like this" to find similar items



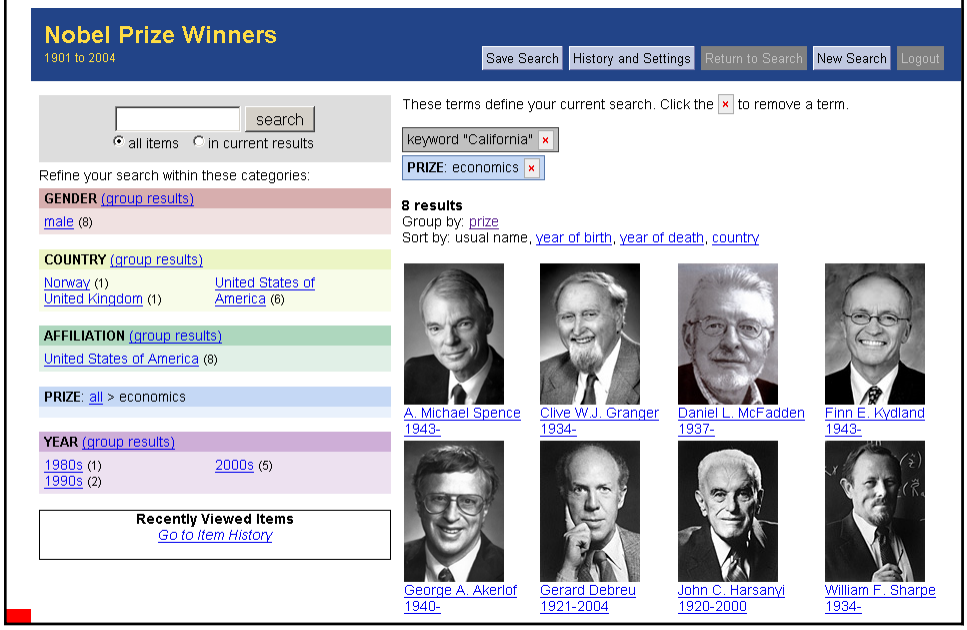
Start a new search using keyword "California"



Note that category structure remains after the keyword search



The query is now a keyword ANDed with a facet subhierarchy



The Challenges

- Users generally do not adopt new search interfaces
- How to show a lot more information without overwhelming or confusing?
  - Most users prefer simplicity unless complexity really makes a difference
  - Small details matter
- Next we describe the design decisions that we have found lead to success.

49

Usability Study Results

50

Search Usability Design Goals

1. Strive for Consistency
2. Provide Shortcuts
3. Offer Informative Feedback
4. Design for Closure
5. Provide Simple Error Handling
6. Permit Easy Reversal of Actions
7. Support User Control
8. Reduce Short-term Memory Load

From Shneiderman, Byrd, & Croft, Clarifying Search, DLIB Magazine, Jan 1997. www.dlib.org

51

Usability Studies

- Usability studies done on 3 collections:
  - Recipes (epicurious): 13,000 items
  - Architecture Images: 40,000 items
  - Fine Arts Images: 35,000 items
- Conclusions:
  - Users like and are successful with the dynamic faceted hierarchical metadata, especially for browsing tasks
  - Very positive results, in contrast with studies on earlier iterations.

52

Most Recent Usability Study

- Participants & Collection
  - 32 Art History Students
  - ~35,000 images from SF Fine Arts Museum
- Study Design
  - Within-subjects
    - Each participant sees both interfaces
    - Balanced in terms of order and tasks
  - Participants assess each interface after use
  - Afterwards they compare them directly
    - Data recorded in behavior logs, server logs, paper-surveys; one or two experienced testers at each trial.
    - Used 9 point Likert scales.
    - Session took about 1.5 hours; pay was $15/hour

53

The Baseline System

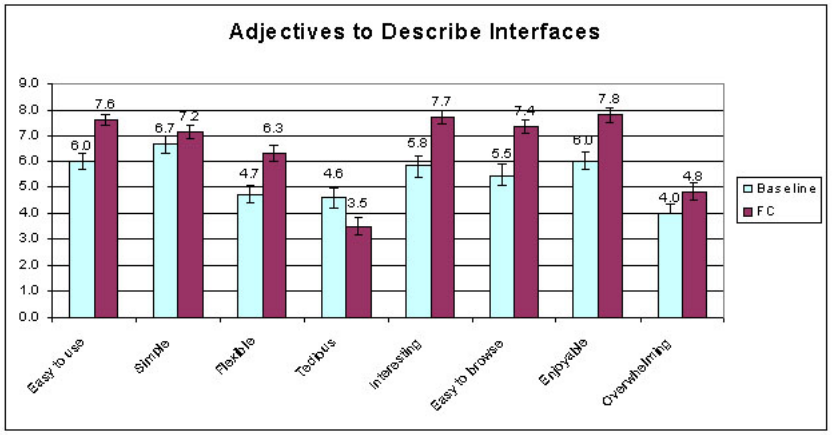
- Floogle (takes the best of the existing keyword-based image search systems)

54


Shasta Image Search
Search
all items
within current results
Current Search: sword
1 2 3 4 5 6 7 8 9 10
Next 10
Items 1 - 40 of 522 results
(One From) from ...
Callot
1633
(One From) from ...
Callot
1633
(One From) Les E...
Callot
1635
(One from) Lux C...
Callot
1628


Shasta Image Search
Search
all items
within current results
Current Search: sword
item 6 of 522 (back to results)
previous
next
(One of) Five Of The Set Of Cardinal Virtues - Justicia
Details about this image:
Media: print
Location: Germany
Date: 16th century
Themes: sword
People: woman
Artists: Beham, Barthel, 1502 - 1540
Record Number: 55649
Object Title: (One of) Five Of The Set Of Cardinal Virtues - Justicia
Description: Before a low masonry wall a winged woman stands, unsheathed sword in her right hand, and a balance in her
Artist Name: Barthel Beham, 1502 - 1540
Media Detail: Engraving
Perform search on relevant keywords:
16th century
Beham, Barthel, 1502 - 1540
Germany
print
sword

Post-Interface Assessments



All significant at p<.05 except “simple” and “overwhelming”

57

Post-Test Comparison

Which Interface Preferable For:	Baseline	Faceted
Find images of roses	15	16
Find all works from a given period	2	30
Find pictures by 2 artists in same media	1	29
Overall Assessment		
More useful for your tasks	4	28
Easiest to use	8	23
Most flexible	6	24
More likely to result in dead ends	28	3
Helped you learn more	1	31
Overall preference	2	29

58

Software Tools

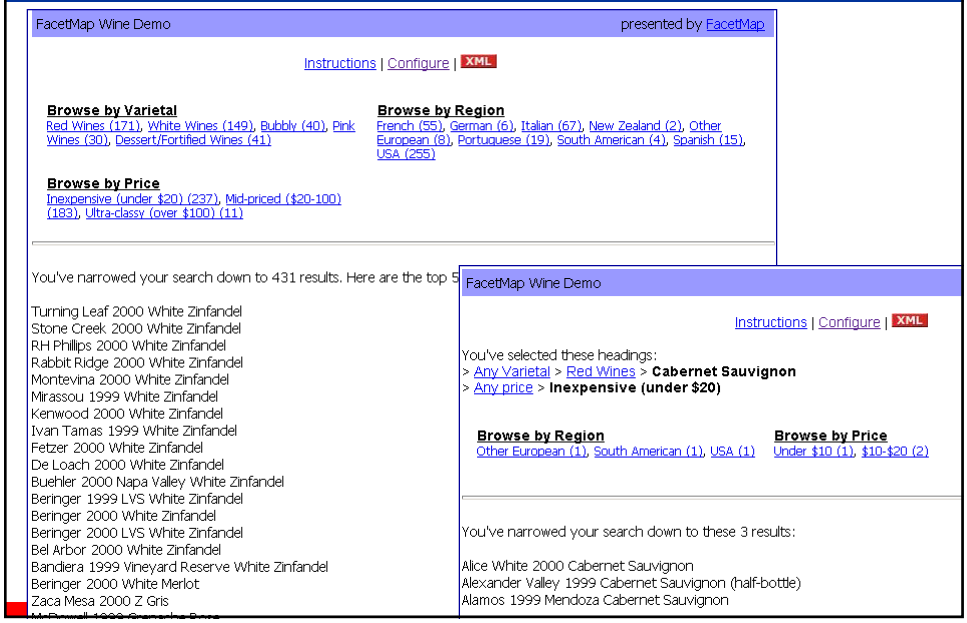
59

Flamenco (flamenco.berkeley.edu)

- Demos, papers, talks are online
  - Nobel example uses this toolkit
- Open source software is now available!
  - Requires Apache and a DBMS (MySQL)
  - You format your data in simple text files
    - (We may add XFML support later)
  - Our programs convert to appropriate DBMS tables
- Check it out:
  - http://flamenco.berkeley.edu

60

FacetMap (facetmap.com)



Commercial Implementations

- (Not an exhaustive list)
  - endeca.com
  - siderean.com
  - www.dieselpoint.com

62

Design Issues

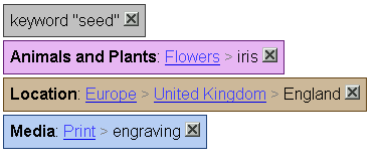
63

Small Details Matter

- With text, it's very difficult to avoid a cluttered look
- Must carefully design visual details
  - White space
  - Font style and weight contrast
  - Color that distinguishes and doesn't clash

BEFORE

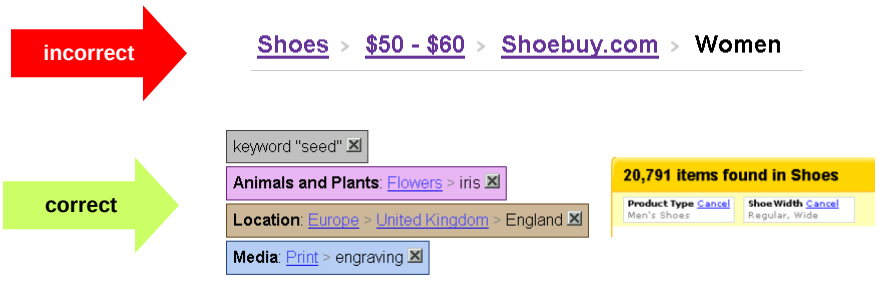
AFTER



64

“Breadcrumb” Design

- Chains should only be used within hierarchy
- Need to separate the facets
  - This allows both expanding within a facet and removing one facet while retaining the rest of the navigation.



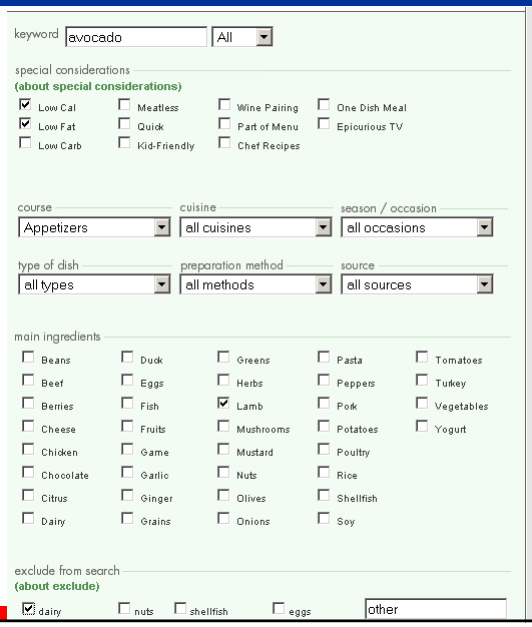
65

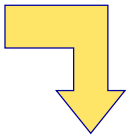
Checkboxes vs. Hyperlinks

- People LOVE checkboxes *in principle*
- However, they are *dangerous* because, when ANDED, they lead to empty results which people HATE
- They also often have confusing semantics
  - Combine AND, OR, keyword search, etc.
  - See Advanced Search at eat.epicurious.com

66

Checkboxes vs. Hyperlinks





Sorry, your search returned no results for: **avocado, Low Cal + Low Fat + Appetizers + Lamb + not dairy**

Perhaps one of the options below can help:
- Try a keyword search (e.g. ham and eggs) in the recipe search box at the top right of the page.
- Try an **advanced search**.
- **Browse** recipes by category.
- See helpful **search tips**.

(Advanced search from epicurious.com)

67

Handling Disjunction (ORs)

- The faceted queries are really a combination of ANDs and ORs
- The facet hierarchies actually do this
  - Example: select
    Animal > Feline AND
    Location >Continent > North America
    - This actually does a query as follows:
      AND(OR (panther, jaguar, lion),
      OR (US, Canada, Mexico))
- Nevertheless, sometimes you want to select just a subset of a facet's labels

68

Handling Disjunction (ORs)

- Using checkboxes with ORs can work
- However, if allowed everywhere they clutter the screen
- eBay shows how to do it:
  - Focus on one facet
  - Select multiple labels
  - Treat as an OR
  - Won't get empty results

69

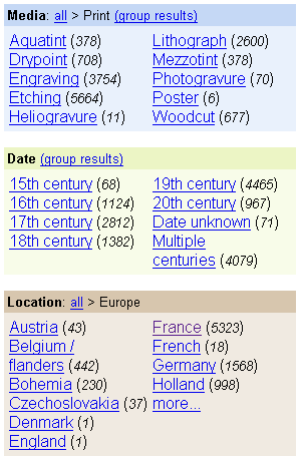
How many facets?

- Many facets means more choice, but more scanning and more scrolling
- An alternative (by eBay)
  - initially show the few most important facets
  - allow user to choose a label from one
  - then show an additional new facet (next most important)
- The right choice depends on the application
  - Browsing art history vs. shopping

70

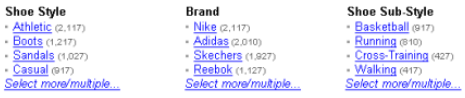
Revealing Hierarchy

- One approach (Flamenco): keep all facets present, show deeper level as you descend.



Revealing Hierarchy

- Another approach (eBay): show only one level at a time; if a facet is chosen that has subhierarchy, show the next level as an additional facet.
  - Example:
    - In Shoes, user selects Style > Athletic
    - Now show a new facet that shows types of Athletic shoes
      - Hiking, Running, Walking, etc.



Reversibility

- Make navigation urls consistent and persistent
  - This way the Back button always works
  - Allows for bookmarking of pages

73

Choosing Labels

- Labels must be short – to fit!
  - Tricky with terminology: “endoplasmic reticulum”
- Labels must be evocative
  - It’s very difficult to find successful words
    - Depends on user familiarity with the domain
  - Use card-sorting exercises
  - Associate synonyms with labels
- Beware the context of label use!
  - The “kosher salt” incident

74

Creating Facets

- Need to balance depth and breadth
  - Avoid long "skinny" hierarchies
    - Example from the Art and Architecture Thesaurus:
    - 7 clicks before you get to anything interesting

Top of the AAT hierarchies
.... Objects Facet
..... Built Environment
..... Built Complexes and Districts
..... complexes
..... <complexes by function>
..... <commercial complexes>
..... shopping centers

75

Summary

- Flexible application of hierarchical faceted metadata is a proven approach for navigating large information collections.
  - Midway in complexity between simple hierarchies and deep knowledge representation.
  - Currently in use on e-commerce sites; spreading to other domains
- We have presented design issues and principles.

76

Session II: Agenda

- Highlights from Session 1 (5 min)
- Interactive exercise (20 min)
- Evolution of IA at eBay (10 min)
- Demo of latest eBay design (5 min)
- Lessons learned at eBay (35 min)
- Discussion and Q&A (15 min)

77

Discussion

78

Faceted Metadata for Information Architecture and Search

CHI Course - April 24, 2006
Session II

Marti Hearst, School of Information, UC Berkeley
Preston Smalley & Corey Chandler, eBay User Experience & Design

Session II: Agenda

- Highlights from Session 1 (5 min)
- Interactive exercise (20 min)
- Evolution of IA at eBay (10 min)
- Demo of latest eBay design (5 min)
- Lessons learned at eBay (35 min)
- Discussion and Q&A (15 min)

80

Highlights from Session I

81

Terminology Clarification

- Facets vs. Attributes
  - Facets are shown independently in the interface
  - Attributes just associated with individual items
    - E.g., ID number, Source, Affiliation
    - However, can always convert an attribute to a facet
- Facets vs. Labels
  - Labels are the names used within facets
  - These are organized into subhierarchies
- Synonyms
  - There should be alternate names for the category labels
  - Currently (in Flamenco) this is done with subcategories
    - E.g., Deer has subcategories “stag”, “faun”, “doe”

82

Interactive Exercise

- 20 minute interactive exercise

83

Evolution of IA at eBay

Flat Structure
(2000 and earlier)

Clothing, Shoes & Accessories
- **Shoes**
  - Women's Shoes
    - Boots
    - Pumps
    - Sandals

84

Evolution of IA at eBay

Flat Structure
(2000 and earlier)

Clothing, Shoes & Accessories
- **Shoes**
  - Women's Shoes
    - Boots
    - Pumps
    - Sandals

Issues with approach:

- **Products had to be categorized in just one way.**

  Ex: Where are all the red Women's shoes?

- **Adding more descriptors meant creating a deep and complicated category structure.**

  Ex: Shoes > Women's > Boots > Black > Size 8

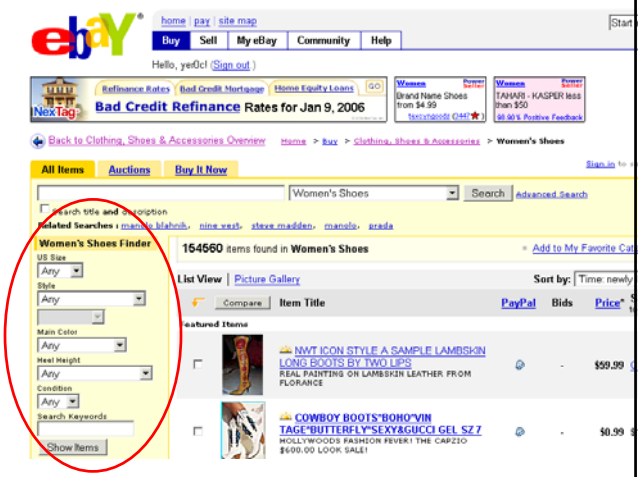
85

Evolution of IA at eBay

+ Product Facets
(2001 – 2005)

Clothing, Shoes & Accessories
- **Shoes**
  - Women's Shoes
    - Style (Boots, Pumps, Sandals…)
    - Size (6, 6.5, 7, 7.5…)
    - Color (Black, Red, Tan…)
    - Condition (New, Used…)

Added Facets (flat)



86

Evolution of IA at eBay

+ Product Facets

(2001 – 2005)

Clothing, Shoes & Accessories

- **Shoes**
  - Women's Shoes
    - Style (Boots, Pumps, Sandals…)
    - Size (6, 6.5, 7, 7.5…)
    - Color (Black, Red, Tan…)
    - Condition (New, Used…)

Issues with approach:

- **Encourages over-constrained queries (Values "ANDED" together)**
- **Placing facets behind dropdowns reduces the exposure of the values to the user**
- **Left-Navigation Placement is only used a minority of the time by users**
- **While effective within a product domain their still is a need for facets above that level**

  Ex: Everything Coach makes that is Red.

87

Evolution of IA at eBay

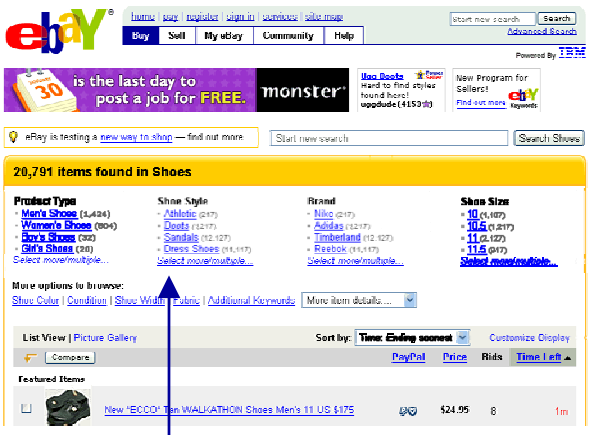
Faceted Metadata

(May 2005 Magellan Test)

Clothing, Shoes & Accessories

- **Shoes**
  - Women's Shoes
    - Style (Boots, Pumps, Sandals…)
    - Size (6, 6.5, 7, 7.5…)
    - Color (Black, Red, Tan…)
    - Condition (New, Used…)
    - Brand (Nine West, Coach…)
- **Brands**
  - Coach
  - Louis Vuitton
- Materials
  - Cotton
  - Leather

Added Hierarchical Facets



88

Demo of latest eBay design

- Try multi-faceted search yourself with the launch of **eBay Express** in Spring 2006.

  See http://express.ebay.com for details.

89

Methodology

Qualitative:

- Rapid Iterative Testing & Evaluation (RITE) Method (2 days testing, 1 day to iterate design)
- n = 48 users (over 9 months)
- 10 versions of the design
- 3 domains: Shoes, TVs, and Collective Glass

Quantitative:

- A/B Test on the live site for 3 weeks
  [n = 73k searches in test environment compared to current site]

90

Lessons Learned at eBay

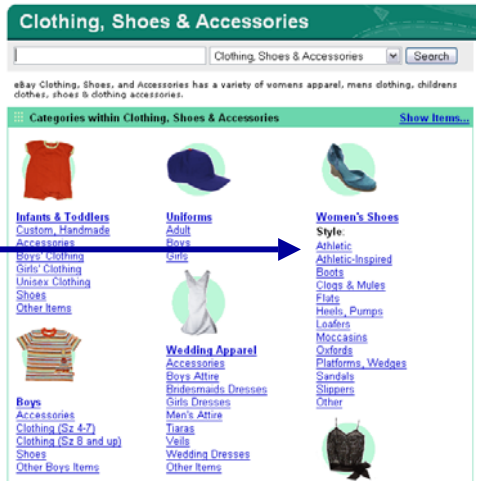
- Data Design
  - Facets
  - Dependencies
  - Flexibility of Facets vs. Hierarchy
- Presentation
  - Integrating "browse" and "search"
  - Control Placement
  - Facet Presentation
  - Breadcrumbs

91

Facets

Lesson: Users desire facets above the domain

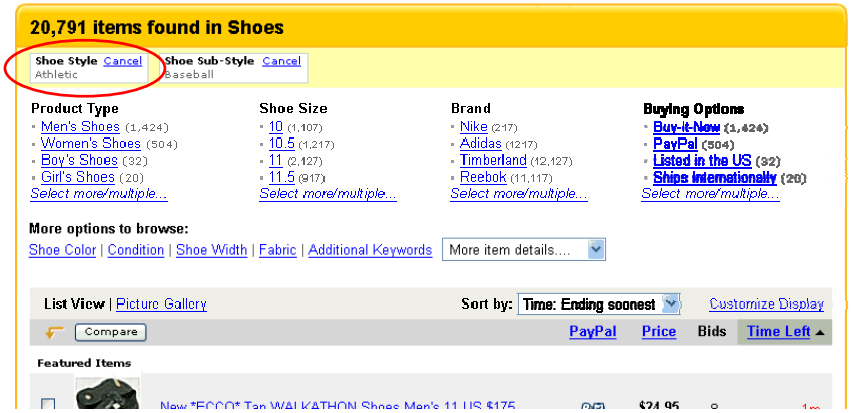
Users also want...
Brands (Coach, Louis Vuitton)
Materials (Leather, Cotton)



92

Dependencies

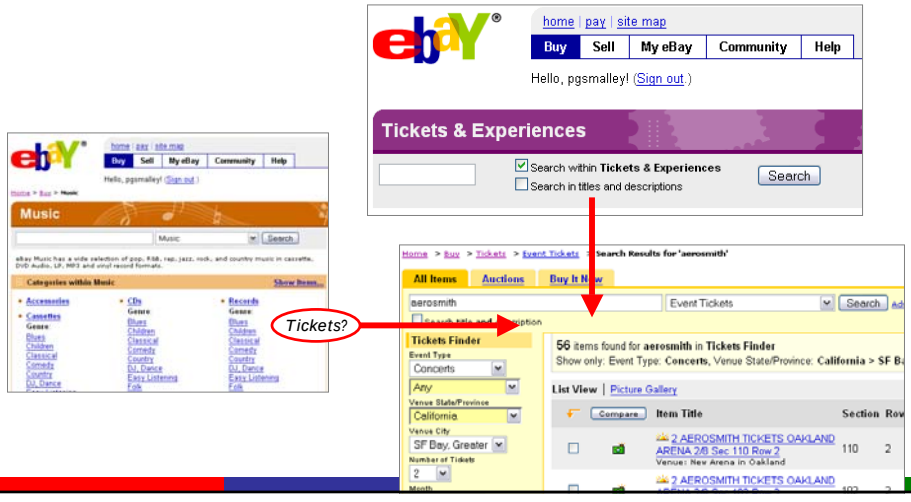
Lesson: Users understand result of removing a parent facet (dependant facets also removed)



93

Flexibility of Facets vs. Hierarchy

Lesson: Users expect multiple entry points into a domain (tickets under sports)



94

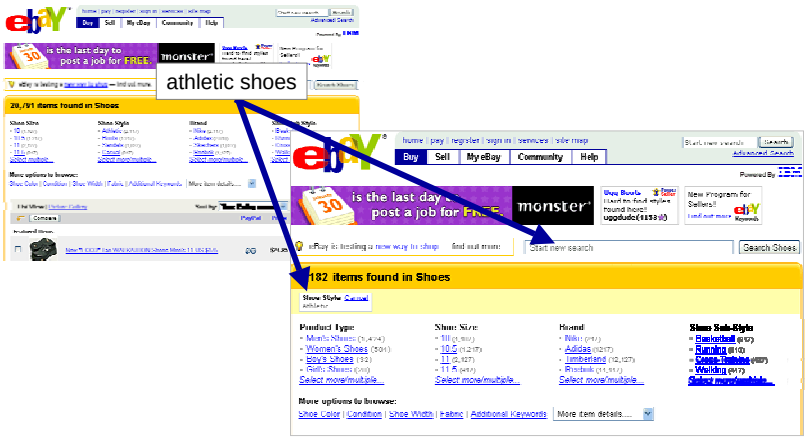
Lessons Learned at eBay

- Data Design
  - ✓ Facets
  - ✓ Dependencies
  - ✓ Flexibility of Facets vs. Hierarchy
- Presentation
  - Integrating "browse" and "search"
  - Control Placement
  - Facet Presentation
  - Breadcrumbs

95

Integrating "browse" and "search"

Lesson: "Parsing" feels natural to users (and the text in the search box is not sacred)



96

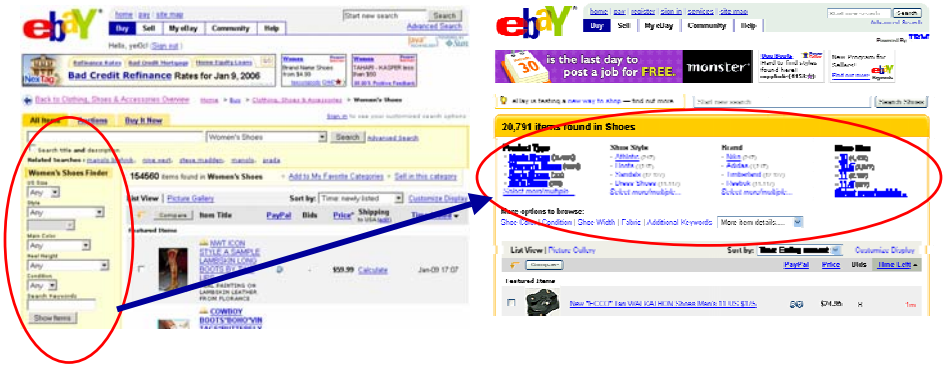
Integrating “browse” and “search”

Lesson: People browse using the facets more when they are not familiar with the domain

97

Control Placement

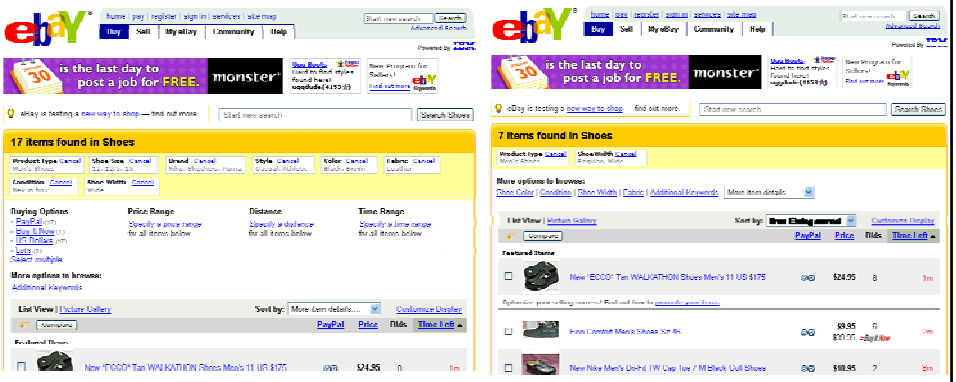
Lesson: Controls placed along the top of the page are used more than when on the left side



98

Facet Presentation

Lesson: Users stop using refinements when
a) not useful, and b) item count low enough



99

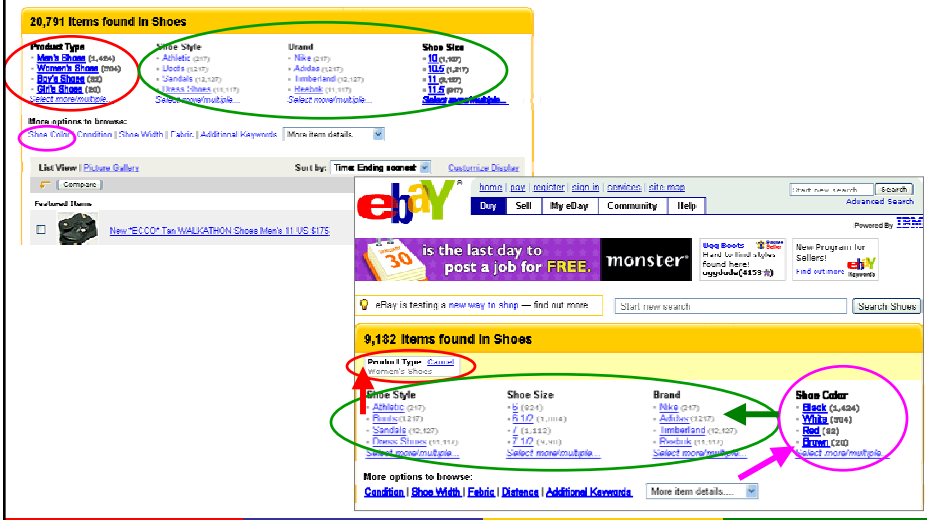
Facet Presentation

Lesson: Prominently showing 4 facets is sufficient (but prioritization is important)

100

Facet Presentation

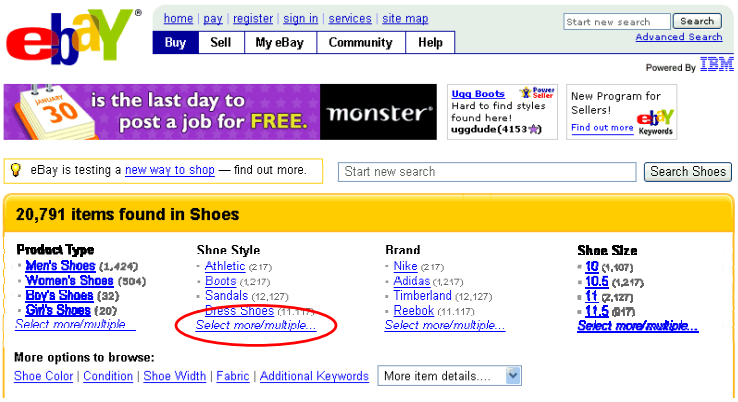
Lesson: Shifting columns doesn't disturb people



101

Facet Presentation

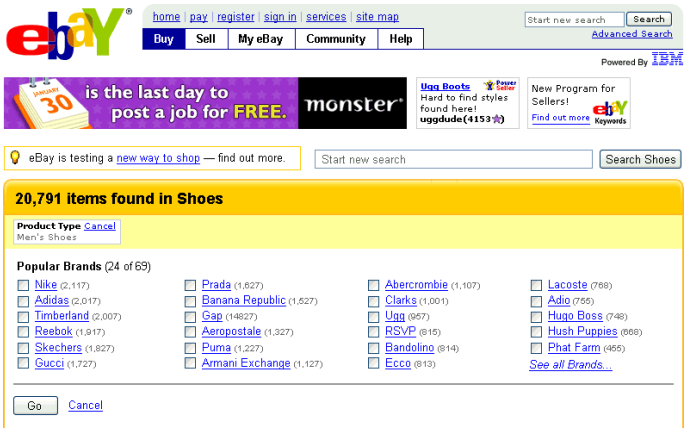
Lesson: Truncated list of values per facet is okay (users know how to access the rest)



102

Facet Presentation

Lesson: Showing sample values help users understand facets and can expose breadth

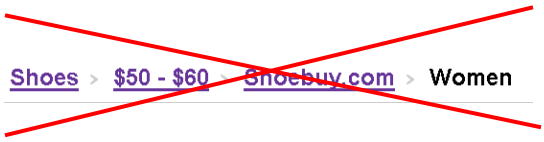


Facet Presentation

Lesson: Users often want to select multiple facet labels and are pleased when they can (treated as an OR by search engine)

Breadcrumbs

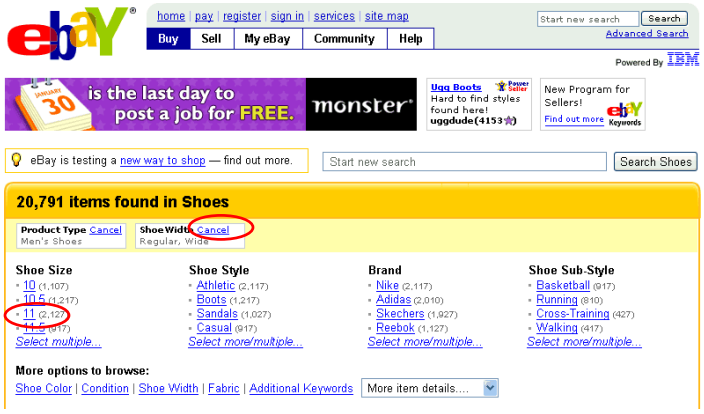
Lesson: Traditional breadcrumbs don't work here



105

Breadcrumbs

Lesson: Users understand the idea of applying and removing facets using this modified breadcrumb without instruction



106

Lessons Learned at eBay

- Data Design
  - ✓ Facets
  - ✓ Dependencies
  - ✓ Flexibility of Facets vs. Hierarchy
- Presentation
  - ✓ Integrating “browse” and “search”
  - ✓ Control Placement
  - ✓ Facet Presentation
  - ✓ Breadcrumbs

107

Discussion and Q&A

- Your chance to make a comment on the subject or ask a question of the presenters.

108

Acknowledgements

- Flamenco Team
  - Brycen Chun, Ame Elliott, Jennifer English, Kevin Li, Rashmi Sinha, Emilia Stoica, Kirsten Swearingen, Ka-Ping Yee
  - This work supported in part by NSF (IIS-9984741)
- eBay Product Team
  - Corey Chandler, Sam Devins, Elaine Fung, Jean-Michel Leon, Michelle Millis, Louis Monier, Michael Morgan, Hill Nguyen, Kenny Pate, Melissa Quan, James Reffell, Suzanne Scott, Seema Shah, Preston Smalley, Anselm Baird-Smith, Luke Wroblewski

109

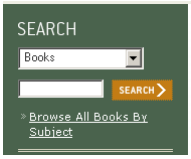
Exercise

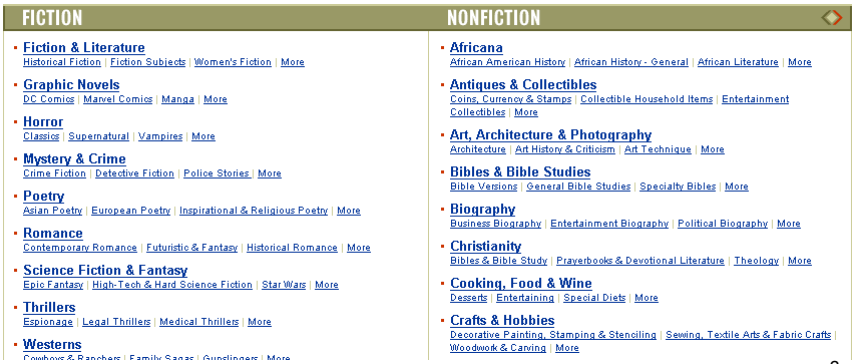
- In the following slides are shown screenshots from an online bookstore that is using faceted navigation.
- In our opinion, this site makes many strong choices
- However, there are at least 3 poor design decisions, from our perspective.
- Can you spot them?

Screenshots taken from browse.barnesandnoble.com, Feb 2, 2006

1

Start with General Browse
Select Nonfiction (not shown)



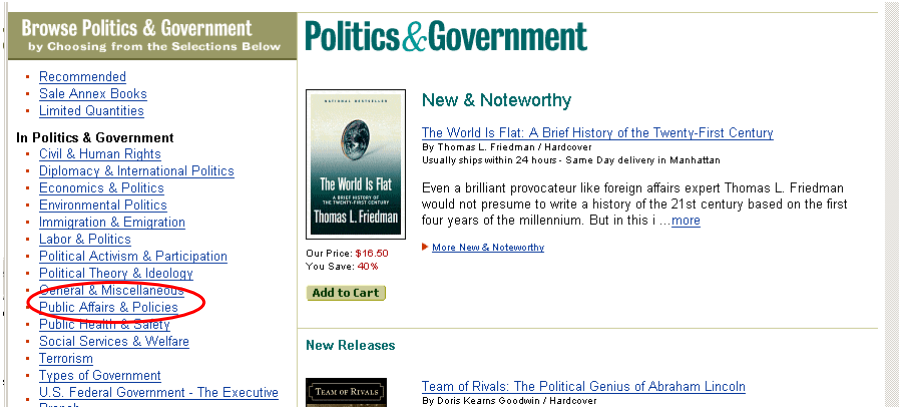


2

Narrow within Nonfiction

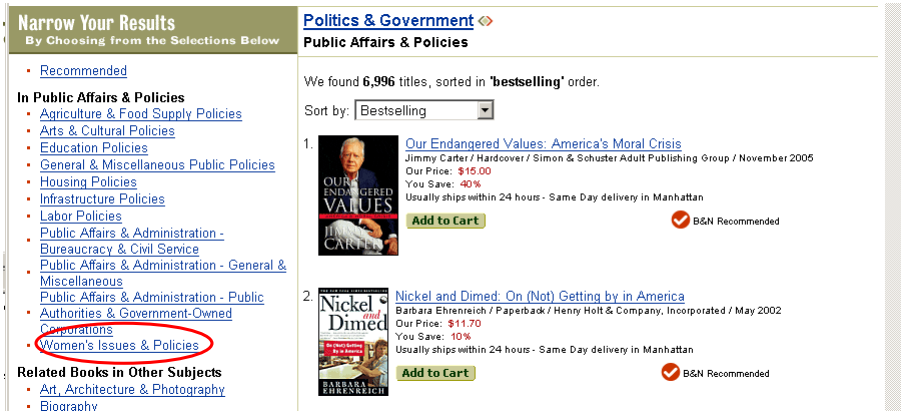
3

Narrow within Politics



4

Narrow within Public Affairs



The resulting screen

Scroll down to see more choices

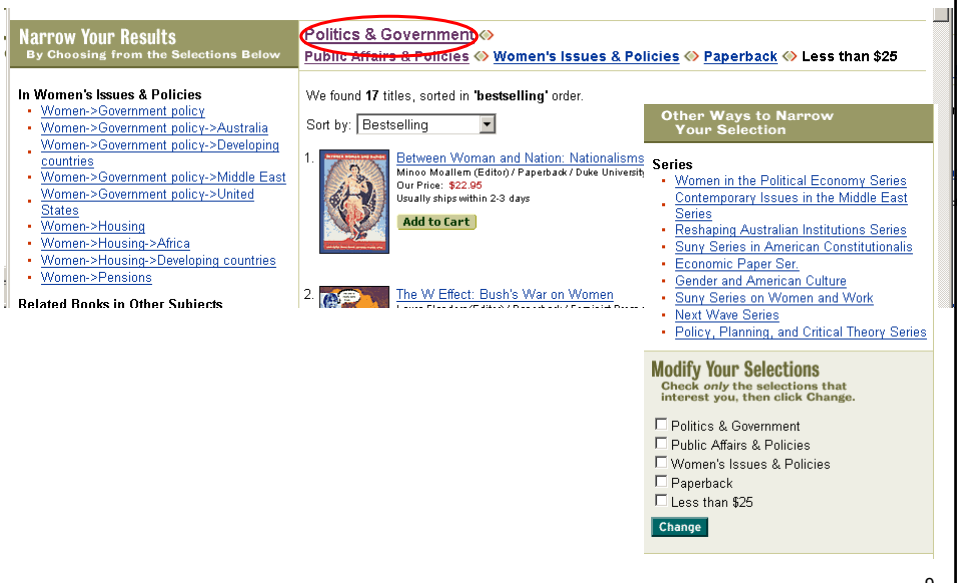
Scroll down to see more choices
Select within these

7

Narrow again on another facet

8

Now broaden out.

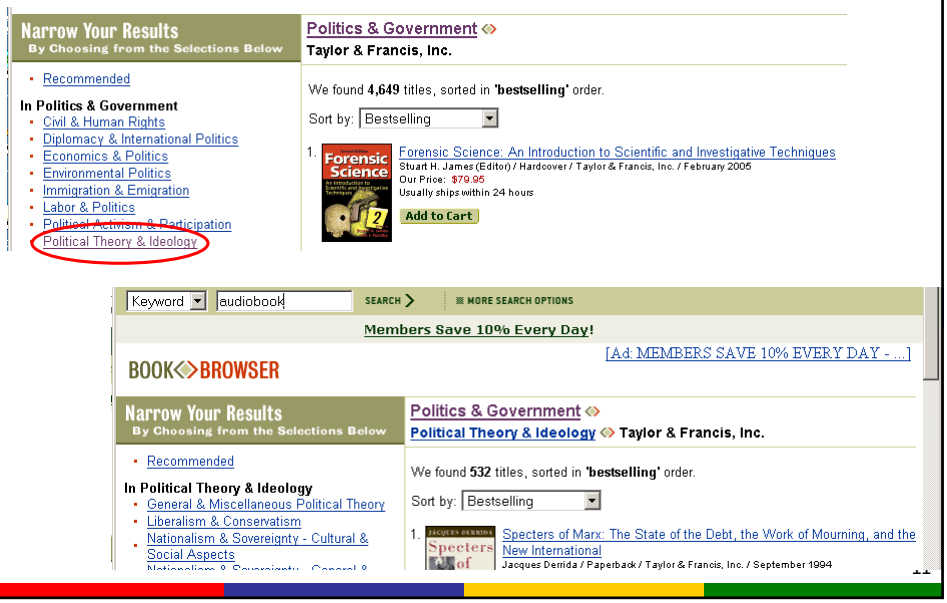


9

Scroll down to see more facet choices
Select one

10

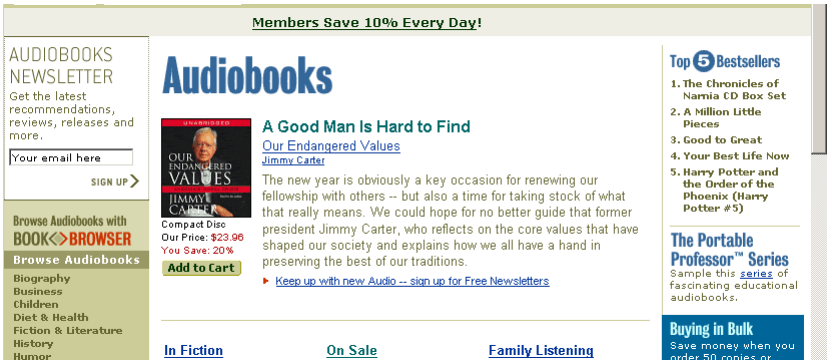
Now Narrow by an existing facet (within Politics & Government)



Now want to search within these

12

Results of the Search



13

How can this design be improved?

14

{By Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee}

FINDING THE FL SITE SEAR

What constitutes a good user interface for search? It depends on the type of answers that users are pursuing. It can be helpful to think of these types of answers as lying along a conceptual continuum, ranging from directed search to informal browsing to text mining and analysis. For example, consider the following questions a user might ask of a large text collection or the Web:

1. How tall is the average female giraffe?
2. What are some good design ideas for landscaping my client's yard?
3. What are some promising untried treatments for Raynaud's disease?

The kinds of answers that best respond to these questions differ qualitatively. For the giraffe question, a single short phrase can be an acceptable response (for example, "4.4 meters") and a standard search interface probably suffices; the user should be able to enter a list of keywords ("giraffe female height") or a natural language question ("What is the height of the average female giraffe?") and the system should simply list the answers, along with links to additional information. For questions of this flavor, Web search engines (such as Google) and automated question answering systems are becoming increasingly successful. This is due in part to a recent swell in commercial and research efforts in this direction, and in part to the redundancy of information available on the Web, which makes systems like these likely to find good answers.





For the landscaping question, a simple list of results is not the best response. It is a more open-ended task; designers tend to look through images searching for inspiration from designs done by others. Thus an interface for this task should allow a designer to browse through a collection and view images relevant to the climate, shape, and existing foliage of the client's yard. The system should also allow a fluid shift from one idea to other related ideas. For example, a view of a garden containing a small cactus in the corner might inspire a designer to change direction and start looking at desert landscapes. This kind of shift should be supported in a manner that does not interrupt the chain of thought, enabling the designer to smoothly steer from one direction to the next, without getting lost and without getting stuck. A direct search method should be a part of such an interface, but it should be tightly integrated with browsing support so as not to interrupt the flow of exploration. An interface framework that supports this type of task is described here.

The question regarding new treatments for Raynaud's represents an analysis problem on the far end of the search task continuum. Trying to discover potential causes of rare diseases by finding links across the biomedical literature is best termed a "text mining" or "knowledge discovery" task [4]. Although it has both a search and a browsing component, this task also requires the ability to track trails of reasoning, perform comparisons, summarize, and otherwise process the information in detail. Designing an interface to support such a task

OW IN WEB CH

Designing a search system and interface may best be served (and executed) by scrutinizing usability studies.

is a fascinating problem, but text mining interfaces are in their infancy.

In this article we focus on the middle part of the answer type spectrum by posing the question of how to design a search system and interface that provide a "browsing the shelves" sensation for large collections of information items. We first summarize what is known from usability results about how to design good search user interfaces. We then illustrate these principles with a browse-and-search interface framework we have developed that has been successful in preliminary usability studies.

Search Interface Desiderata

How does one build an interface that successfully supports both direct search and browsing? The press is rife with accounts of failed searches and unhappy users. For example, a recent report by Forrester Research found that while 76% of firms rated search as "extremely important" only 24% consider their Web site's search to be "extremely useful" [6].

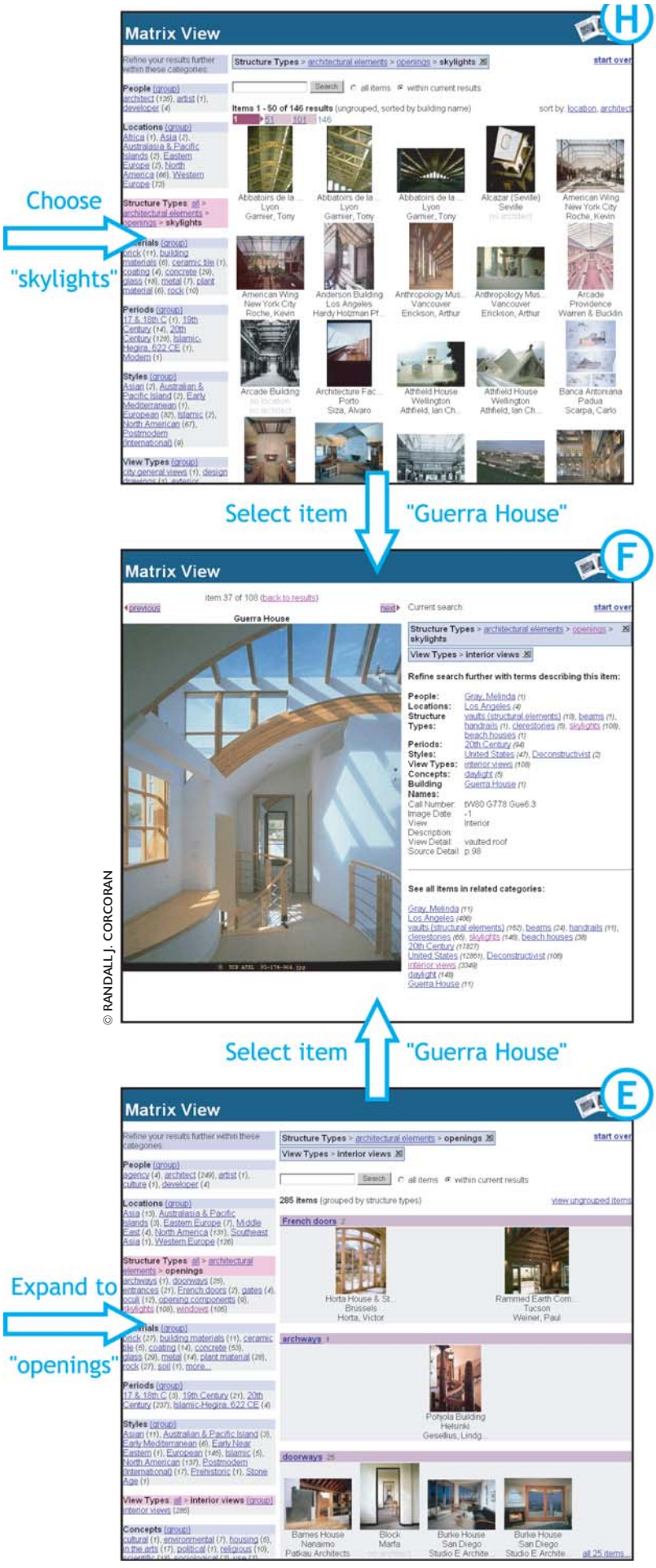
In our view, the way to do things correctly is to use the evidence found in the results of usability studies of search systems. Unfortunately, most studies of search behavior are inconclusive about how to improve the system (for example, [12]), but some consistencies do emerge about what works. Here, we summarize which search features tend to work well, and which fail, in practice. Throughout this article, the assumption is that the user population consists of people who do not specialize in search and who have only basic knowledge of how to use computers.

First and foremost, most users engaged in directed searches are not interested in search for its own sake; thus systems that make users focus on the operations for performing search are seldom successful [1]. For browsing tasks, users are engaged with the data, but again are not focused on the mechanisms of the search system. Users can tell the difference between these two cases. In a small study we conducted on a recipe Web site [3], we found that users preferred a browsing-oriented interface for a browsing task, and a direct search interface when they knew precisely what they wanted.

Features found to work well across studies are color highlighting of search terms in result listings (also known as "keywords-in-context"); sorting of search results along criteria such as date and author; and grouping search results according to well-organized category labels [5].

Certain features are helpful in principle, but only work in practice if the underlying algorithms are highly accurate and if the interface is carefully designed. Some examples of such features include spelling correction, automated term expansion, and simple relevance feedback (also known as "more like this"), in which the user selects one item and the system shows items that are similar in scope along several dimensions.

"FLAMENCO" ALLOWS USERS TO MOVE THROUGH LARGE INFORMATION SPACES IN A FLEXIBLE MANNER WITHOUT FEELING LOST.



Two simple features are underappreciated by search researchers—exposing metadata in the interface, and making use of hyperlinks and the interactive nature of user interfaces. Other reports have found that hyperlinks outperform search on most Web sites [6]. Our view is the two should be tightly integrated for access to content within Web sites or large information collections.

Specific problems most often named in the literature include empty result sets (zero results); disorganized result lists; results that make the user feel lost or overwhelmed; difficulty with using the correct terminology; and difficulty with forming queries where special syntax is required (for example, specifying Boolean expressions) [5].

The incorporation of visualization into search interfaces has yet to be favorably received by users in general [11]. Similarly, text clustering is not found to be valuable for ordinary users who prefer organization according to categories that have predictable, understandable meanings [9]. These tools are more likely to be effective for knowledge discovery tasks, like the Raynaud's treatment question.

Shneiderman et al. [10] specify eight design desiderata for search user interfaces: strive for consistency; offer informative feedback; offer simple error handling; permit easy reversal of actions; support user control; reduce short-term memory load; design for closure; and provide shortcuts for experts. The browsing interface described here attempts to incorporate most of these design elements.

Search Interfaces That Flow

We have created a search interface framework called "Flamenco" whose primary design goal is to allow users to move through large information spaces in a flexible manner without feeling lost (see Figure 1). A key property of the interface is the explicit exposure of hierarchical faceted metadata, both to guide the user toward possible choices, and to organize the results of keyword searches. The interface

Figure 1. Flamenco: A search interface that flows.

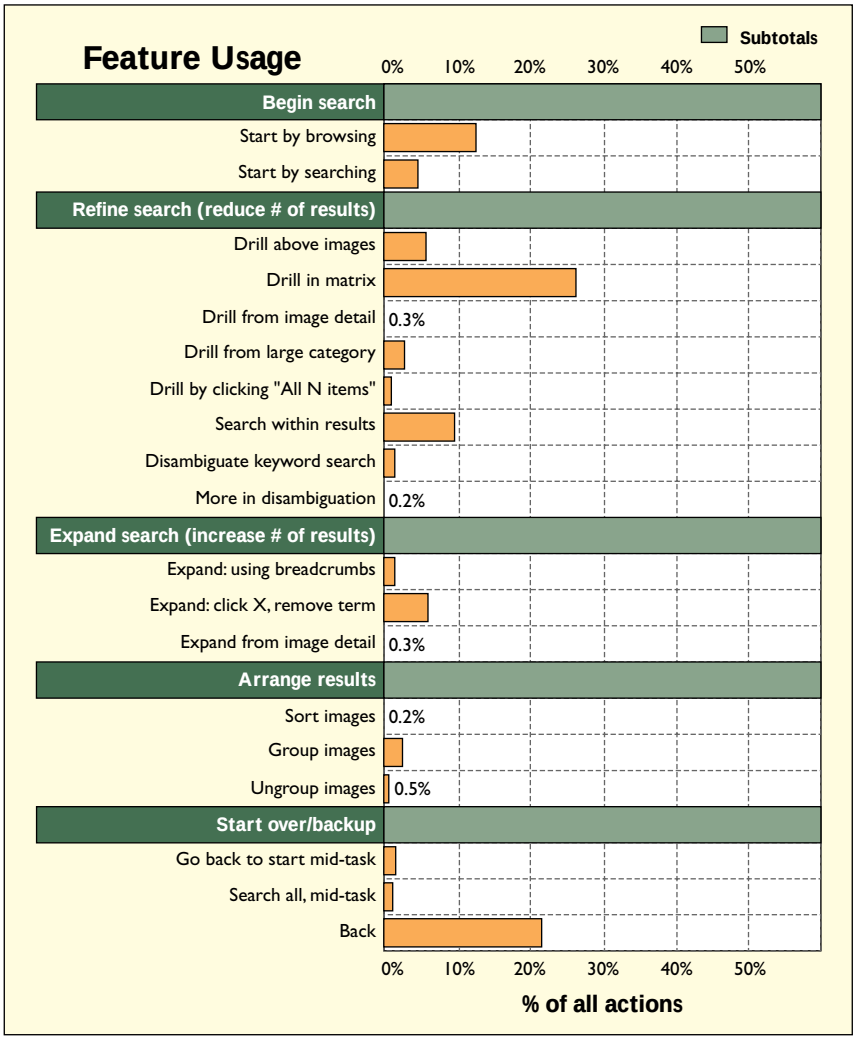


Figure 2. Percentage of time features were used. "Drill" means refine by descending a subhierarchy.

uses metadata in a manner that allows users to both refine and expand the current query, while maintaining a consistent representation of the collection's structure. This use of metadata is integrated with free-text search, allowing the user to follow links, then add search terms, then follow more links, without interrupting the interaction flow. This system builds on earlier work that shows the importance of query previews [8] for indicating next choices. Query previews allow users to recognize terms rather than remember them, and eliminate the occurrence of empty result sets.

Architects and city planners were the target user population for the studies described here. The collection consisted of images from an architecture slide library. However, we have applied the framework to other datasets, including a collection of biomedical articles and a collection of consumer products.

We approached the problem of developing the search interface framework by following user-centered design practices from the field of human-computer interaction [2]. We first performed a needs assessment of the target population, including an ethnographic analysis of how architects use and look for images as inspiration for their design work. We then built a simple prototype and evaluated it with an informal usability test. Next, we conducted two rounds of development and two formal usability studies, revising the interface based on the results of each study. By the final round, the study participants were very enthusiastic about the design. Several expressed a strong desire to use the new system in the future, despite the fact it differs significantly from conventional search interfaces.

Hierarchical-Faceted Metadata

Content-oriented category metadata has become more widespread in the last few years, and there is much activity in the creation of standards for describing content in various fields (for example, Dublin Core and the Semantic Web; dublincore.org; www.w3.org/2001/sw). Web directories such as Yahoo and the Open Directory Project (www.yahoo.com; dmoz.org) are familiar examples of the use of metadata for navigation structures. Web search engines have begun to provide search hits on category labels together with other search results.

Many individual collections already have rich metadata assigned to their contents; for example, biomedical journal articles have on average more than a dozen content attributes attached to them. Metadata for organizing content collections can be classified along several dimensions:

The metadata may be faceted, that is, composed of orthogonal sets of categories. For example, in the domain of architectural images, some facets are Materials (concrete, brick, wood, among others), Styles (Baroque, Gothic, Ming), Locations, and so on. The metadata may be hierarchical ("located in

Berkeley, California, United States") or flat ("by Ansel Adams"). The metadata may be single-valued or multivalued. That is, the data may be constrained so that one value at most can be assigned to an item ("measures 36 cm tall") or it may allow multiple values to be assigned to an item ("uses oil paint, ink, and watercolor").

There are a number of issues associated with the creation of metadata itself that are not addressed in this article. The most pressing problem is how to decide which descriptors are correct or at least the most appropriate for a collection of information. Another problem relates to how to assign metadata descriptors to items that currently do not have metadata assigned. Many researchers are addressing these issues, and the field of automated text categorization is making great strides. Additionally, many important collections with hand-assigned hierarchical metadata already exist.

We illustrate the interface using an architectural image database containing about 40,000 photographs of landscapes and buildings from a wide variety of historical periods, styles, and geographic regions (see Figure 1). The images are classified under about 16,000 hierarchical metadata terms, which we manually reorganized into nine facets: people, locations, structure types, materials, periods, styles, view types, concepts, and building names.

We use a brief scenario to demonstrate how the interface works. Imagine a user named Claire who has a beach house she plans to renovate, with the goal of bringing more natural light into the living room. Before she meets with the architect, she browses through the architectural image collection to gather a few ideas.

She begins at the starting page (see A in Figure 1). This page shows an overview of available topics, each hyperlinked to the equivalent of a query on the corresponding metadata term, and each link showing how many items have been assigned that topic label. To help her in this search, the starting page also includes three sample images from each facet. To begin her search, Claire may either click one of these links or issue a keyword search.

Claire opts to begin by clicking the hyperlink "interior views" in the View Types facet and has arrived at what we call the "matrix view" (see B). There is a column of metadata on the left and the images in the current result set on the right. The matrix shows query previews for all of the metadata terms assigned to the images in the current result set. These previews are updated as constraints are added or removed. The caption under each image gives the name of the building, the location, and the architect.

Claire's eye is drawn to one image showing an interior flooded with daylight. She clicks on this image to see a more detailed view (C). After reading the metadata categories assigned to the image, Claire clicks on the term "windows" found under the Structure Types facet. This refines her query because it conjoins the metadata term "windows" with the current query. Doing this creates a new matrix page (see D).

Now the query, consisting of metadata from the two selected facets (View Types and Structure Types), is shown at the top of the screen in the form of hyperlinked history trails (or "breadcrumbs"). The images are grouped according to subcategories of the "windows" metadata category; up to four sample items are shown in each subcategory. Note that the interface allows the user to navigate multiple hierarchies simultaneously.

To further refine her search, Claire can select terms from other facets by clicking in the matrix on the left or by selecting a subcategory on the right. The results set can be broadened (expanded) to include more items by selecting a general category within the breadcrumb or by clicking the X to remove a category constraint. Assume that Claire clicks on the "openings" category, just above "windows" in the breadcrumb, to relax the Structure Types constraint. This brings her to E.

Clicking on an image within the "skylights" subcategory brings her to the image detail (F) where she sees several other helpful terms: "daylight," "beams," and "beach houses." This page allows Claire to make lateral moves, shifting to associated categories that were not part of the original query. We have found this facility is important for promoting shifts to areas of the collection that users had not considered previously.

The interface makes a keyword search facility available at all points in the interface. The scope of the search is by default the current result set, although users can also choose to search through the entire collection.

Thus Claire might have begun her exploration by running a keyword search on the word "light." In this case, a list of all matching metadata terms appears above the result set, with search terms highlighted as shown in G. Selecting the metadata term "skylights" in the list converts the keyword constraint into a category constraint as shown in H.

This, in turn, could lead her to the image detail in F.

In some situations, there are too many subcategories or keyword matches to fit on the page. When this occurs, an alphabetized list is presented on a page of its own so the user can make a selection. The links labeled "more"—visible in the matrix view—can also take the user to listing pages of this type.

Returning now to the discussion of usability guidelines for search interfaces, notice this interface supports six out of Shneiderman et al.'s eight design desiderata. It is consistent and it constantly gives feedback on the query state. Reversal of actions is supported by various methods for going back, canceling query terms, and starting over. The system maintains the query state entirely in the URL so the browser's back button and bookmark features work correctly. These browser features help to keep the user in control and also serve as important memory aids. The use of query previews emphasizes recognition over recall, which also reduces short-term memory load, and helps provide an information scent of where to go next. The ever-present search box provides a shortcut for the user who has a specific goal in mind.

A Usability Study

We conducted a usability study in which 19 architects and city planners (practitioners and students) participated. About half stated they looked for images "all the time;" the other half said they searched for images on a monthly or yearly basis. Data was recorded with multiple methods: server logs, behavioral logs (time-stamped observations), online post-task questionnaires, and paper surveys at the end of the session. Two experienced usability analysts conducted each session. A within-subjects design was used in which the interface presented here was compared to a similar one with slightly less functionality, and participants performed several different types of search and browsing tasks. Space restrictions prevent detailed reporting of the results, so only some highlights are presented here; see [3] for more information.

Mean ratings for feature usefulness and understanding were high (ranging between 5.6 and 6.9 on a 7-point Likert scale). This was in contrast to previous iterations where participants did not notice, did not understand, or did not like some of the most powerful features.

One concern was that with so many varied options participants might find the interface too browsable, and feel lost. However, the results were that participants felt a strong sense of control (average 5.65 on a 7-point Likert scale).

A more direct measure of usefulness is how often the features are actually used. Figure 2 summarizes these results, and shows that participants chose to begin more frequently by browsing (12.7% of all operations) than by searching (5%). For refining actions, participants refined by using "Drill in matrix" 26.6% of the time, while the "Search Within" facility was used only 9% of the time. We think this shows the power of the faceted hierarchies, which allow participants to flexibly modify their query rather than forcing them to choose appropriate keywords for searching.

The option to expand on a facet is not available in most search interfaces, so this feature was unfamiliar to most participants. Nevertheless, about 7% of the participants' actions were related to expanding a search. We suspect this feature will have heavier use once users become more experienced with the interface.

Participants chose to start over in the middle of a task only 0.02% of the time, which suggests they did not get stuck or lost while using the system.

The majority (16 out of 19) of the participants said they preferred the power and flexibility of the matrix-based interface to a simpler interface. This is especially significant given it is fairly uncommon for users to prefer more complex and unconventional interfaces. Participants found it easy to refine and expand their searches using the various features; they liked having the choices for refining the search displayed on the left side of the screen along with the images. Participants referred to the metadata display as a "map," an "index," a "table of contents," and a "menu." Some participants were initially put off by the text-heavy appearance of the matrix, but grew to like it after they had completed one or two tasks.

Search usability studies show that non-expert searchers have difficulty with Boolean queries beyond simple conjunction. [5]. An advantage of our approach is it allows users to easily compose queries consisting of ANDs of ORs: selecting a category term is effectively an OR of all of its subcategories, and selecting more than one facet produces an AND across facets. Research in the biomedical literature tells us that forming ANDs of ORs of related terms is one of the more effective ways to search [7].

This interface is not without problems. It does appear to have more functionality than is needed for direct search; if users know exactly which item they want, a simpler interface seems to be more efficient. Furthermore, it is hampered by a fundamental problem with the use of metadata: the terminology pro-

vided may not match the set of words wanted by users. To address these and other problems, we plan to augment the system in several ways, including incorporating thesaurus term matching into the search, and using techniques from adaptive user interface research. For instance, a relevance feedback mechanism could take into account which metadata facets are most often used together, and could show the most popular facets before less popular ones.

Implementation

An added advantage of this framework is it can be built using off-the-shelf database technology. (However, special-purpose software may be required in order to scale to millions of items.) The system allows content creators to add new items, and can be applied to entirely new collections without requiring changes to the application logic or the interface. The system is implemented using Python, MySQL, and the WebWare toolkit (www.python.org; www.mysql.com; Webware.sourceforge.net). Collections are stored according to a generic database schema that accommodates a wide range of metadata: facets can be hierarchical or flat, single-valued or multivalued. All components of the interface are dynamically generated, based on the facets and metadata terms defined in the database. A clean abstraction layer translates queries composed of metadata terms into standard SQL queries over the schema. Query previews are generated using the SQL group by operator to count the number of items that fall into each subcategory.

The interface design we've described reflects bits and pieces of what can be found in existing Web interfaces, especially on e-commerce sites. Until recently, however, most of these interfaces were confusing and cluttered, or did not allow expansion, or did not successfully integrate search within the navigation metadata. However, some recent commercial systems have begun to incorporate the ideas presented here.[1]

Conclusion

This article has discussed the importance of usability results and user-centered design practices in the development of better user interfaces for different types of search tasks. We have illustrated the results of this approach when applied to an interface that allows for browsing and searching through the use of faceted hierarchies of metadata and hyperlinked query previews, and verified the promise of the approach through usability studies. For more information and a demonstration, see flamenco.berkeley.edu. C

[1]In particular, a company called Endeca (www.endeca.com) has begun creating Web sites with many of the interface ideas found useful in our research.

References

1. Bates, M.J. The berry-picking search: User interface design. *User Interface Design.* H. Thimbleby, Ed. Addison Wesley, Reading, PA, 1990.
2. Brinck, T., Gergle, D., and Wood, S.D. *Usability for the Web: Designing Web Sites that Work.* Morgan Kaufmann, New York, 2002.
3. English, J., Hearst, M.A., Sinha, R., Swearingen, K., and Yee, P. Flexible search and browsing using faceted metadata, 2002. In preparation.
4. Hearst, M.A. Untangling text data mining. In *Proceedings of the 37th Meeting of the Association for Computational Linguistics.* (College Park, MD, June 1999), 3–10.
5. Hearst, M.A. User interfaces and visualization. *Modern Information Retrieval.* Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Eds. ACM Press, New York, 1999, 257–323.
6. Johnson, K., Manning, H., Hagen, P.R., and Dorsey, M. Specialize Your Site's Search. Forrester Research, (Dec. 2001), Cambridge, MA; www.forrester.com/ER/Research/Report/Summary/0,1338,13322,00.html,
7. Lowe, H.J., and Barnett, G.O. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA* 271, 4 (1994), 1103–1108.
8. Plaisant, C., Shneiderman, B., Doan, K., and Bruns, T. Interface and data architecture for query preview in networked information systems. *ACM Trans. Information Systems 17*, 3 (1999). ACM, New York, 320–341.
9. Pratt, W., Hearst, M.A., and Fagan, L. A knowledge-based approach to organizing retrieved documents. In *Proceedings of 16th Annual Conference on Artificial Intelligence*. (Orlando, FL, 1999).
10. Shneiderman, B., Byrd, D., and Croft, W.B. Clarifying search: A user-interface framework for text searches. *DLIB Mag.* (Jan. 1997).
11. Sutcliffe, A.G., Ennis, M., and Hu, J. Evaluating the effectiveness of visual user interfaces for information retrieval. *Intern. J. Human Computer Studies 53,* 5 (Nov. 2000) 741–763.
12. Sutcliffe, A.G., Ennis, M., and Watkinson, S.J. Empirical studies of end-user information searching. *J. American Society for Info. Science 51*, 13 (2000).

This research was supported by a NSF Career grant number IIS-9984741.

Marti Hearst (hearst@sims.berkeley.edu) is an associate professor in the School of Information Management and Systems (SIMS) at University of California, Berkeley.
Ame Elliott (ame@rii.ricoh.com) is a postdoctoral research scientist at Ricoh Innovations, Inc., Menlo Park, CA.
Jennifer English (jenglish@sims.berkeley.edu) is a graduate of the School of Information Management and Systems (SIMS) Masters program at University of California, Berkeley.
Rashmi Sinha (sinha@sims.berkeley.edu) is a researcher at the School of Information Management and Systems (SIMS) at University of California, Berkeley.
Kirsten Swearingen (kirstens@sims.berkeley.edu) is a graduate of the School of Information Management and Systems (SIMS) masters program at University of California, Berkeley.
Ka-Ping Yee (ping@zesty.ca) is a Ph.D. student in computer science at the University of California, Berkeley.

Faceted Metadata for Image Search and Browsing

Ka-Ping Yee[1] ping@zesty.ca
Kirsten Swearingen[2] kirstens@sims.berkeley.edu
Kevin Li[1] kevinli@cs.berkeley.edu
Marti Hearst[2] hearst@sims.berkeley.edu
[1]Computer Science Division
University of California, Berkeley
[2]School of Information Management and Systems
University of California, Berkeley

ABSTRACT

There are currently two dominant interface types for searching and browsing large image collections: keyword-based search, and searching by overall similarity to sample images. We present an alternative based on enabling users to navigate along conceptual dimensions that describe the images. The interface makes use of hierarchical faceted metadata and dynamically generated query previews. A usability study, in which 32 art history students explored a collection of 35,000 fine arts images, compares this approach to a standard image search interface. Despite the unfamiliarity and power of the interface (attributes that often lead to rejection of new search interfaces), the study results show that 90% of the participants preferred the metadata approach overall, 97% said that it helped them learn more about the collection, 75% found it more flexible, and 72% found it easier to use than a standard baseline system. These results indicate that a category-based approach is a successful way to provide access to image collections.

Keywords: Image Search Interfaces, Faceted Metadata

INTRODUCTION

Image collections are rapidly coming online, and many researchers have developed user interfaces for browsing and searching such collections. Probably the most familiar image search interface today is that used by Web image search engines, in which users enter keyword terms, and images are shown in a table ordered by some measure of relevance. These systems can be effective for searching for very specific items, but do not support browsing and exploratory tasks well [7, 9, 10]. Many research systems approach image retrieval by analyzing images in terms of visual properties such as color and texture. However, results of usability studies call into question the usefulness of image searching according to low-level visual properties [10, 15].

In contrast, and perhaps counter-intuitively, ethnographic studies indicate that professionals who look for images on a regular basis (e.g., journalists, designers, and art directors) want to browse and search images using textual category labels [1, 5, 7, 10]. Despite this, few image search engines provide the ability to navigate images by rich category sets, and those that do often have unwieldy interfaces [10].

We have developed an interface for large image collections that allows users to navigate explicitly along conceptual dimensions that describe the images [8]. The interface uses hierarchical faceted metadata (described below) and dynamically generated query previews [14], seamlessly integrating category browsing with keyword searching. To arrive at the current design, we conducted several rounds of usability studies and interface redesign. This paper presents the results of a new usability study whose goal is to directly compare the faceted category design to the current most popular approach to image search. Conducted with 32 art history students using a fine arts image collection, the study found strong preference results for the faceted category interface over that of the baseline, suggesting this to be a promising direction for image search interfaces.

We now describe related work, the faceted metadata, the category-based interface design, the baseline interface, and the study design and results, concluding with a discussion of the larger lessons that can be drawn from this effort.

RELATED WORK

The bulk of image retrieval research falls under the rubric of "content-based" image retrieval; this term refers to systems that perform image analysis in order to extract low-level visual properties, such as color and texture [12, 13] or object segmentation [4]. Some systems also incorporate information extracted from associated text [17]. A good summary of content-based image retrieval can be found in [18].

There has been a great deal of research on these systems, but only a small subset of the past work has included usability studies. Rodden et al. [15] performed a series of experiments whose goal was to determine if and how organization by visual similarity is useful, using as features global image properties (colors and textures) and the spatial layout of image regions. Their results suggested that images organized by category labels were more understandable than those grouped by visual features.

Ethnographic studies of image search needs have indicated that there is a great need for more conceptually rich image search. In a study of art directors, art buyers, and stock photo

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.

researchers [7], Garber & Grunes found that the search for appropriate images is an iterative process: after specifying and weighting criteria, searchers view retrieved images, then add criteria, add restrictions, change criteria, or redefine the search. The concept often starts out loosely defined and becomes more refined as the process continues.

Markkula and Sormunen [10] reported on a field study of journalists and newspaper editors choosing photos from a digital archive in order to illustrate newspaper articles. Journalists stressed the need for browsing, and considered searching for photos of specific objects to be a "trivial task". Selection of search keys for general topics was considered difficult; journalists emphasized the need for photos dealing with places, types of objects, and themes. The journalists had access to an "advanced search" interface that allowed them to search on many different features at once, but its format, which consisted of about 40 entry forms and drop-down boxes, was seen as too complex, and was rarely used. Thus, although they had the desire to do searches on multiple categories, the interface discouraged them from doing so.

A query study also supports the notion that users want to search for images according to combinations of topical categories. Armitage and Enser [1] analyzed a set of 1,749 queries submitted to 7 image and film libraries. They classified the queries into a 3-by-4 facet matrix; for example, *Rio Carnivals* fell under *Geographic Location* and *Kind of Event*. They did not summarize how many queries contain multiple facets, but showed a set of 45 selected queries, to which they assigned an average of 1.9 facets per query.

The system proposed by Garber & Grunes [7] is the interface most similar to our approach. The interface operated in two modes: (i) showing metadata associated with a target image, and presenting images in an order reflecting the number of categories they had in common with the target image; and (ii) allowing the user to select a set of category labels, and showing sample images for similar categories (e.g., showing images labeled *New England*, *Africa*, and *Egypt* when the category label *Florida* is selected). Hierarchy information was not shown, and no information was provided about how many images are available in each category. Focus groups observing the demonstration were very enthusiastic about it, but no followup work appears to have been done.

METADATA

Here we define and illustrate the notion of faceted metadata.

Faceted Metadata

Content-oriented category metadata has become more prevalent in the last few years. Many individual collections already have rich metadata assigned to their contents; for example, biomedical journal articles typically have a dozen or more content attributes attached to them. Metadata for organizing collections can be classified along several dimensions:

- The metadata may be *faceted*, that is, composed of orthogonal sets of categories. For example, in the domain of fine arts images, possible facets might be themes (military, religious, etc.), artist names, time periods, media (etching, woodblock, ceramic, etc.), geographical locations, and so on.
- The metadata (or an individual facet) may be *flat* ("by Pablo Picasso") or *hierarchical* ("located in Vienna $>$ Austria $>$ Europe").
- The metadata (or an individual facet) may be *single-valued* or *multi-valued*. That is, the data may allow at most one value to be assigned to an item ("measures 36 cm tall") or it may allow multiple values to be assigned to an item ("uses oil paint, ink, and watercolor").

There are a number of challenges associated with metadata, including choosing the most appropriate descriptors for a given collection and assigning metadata descriptors to items that do not have any metadata assigned. Researchers are investigating these problems (e.g., [17]), but there are in fact many existing, important collections whose contents already have hierarchical metadata assigned.

Collection Preparation

The collection under study consisted of approximately 35,000 images out of the more than 82,000 images in the Thinker collection of the Fine Arts Museum of San Francisco (metadata was available only for a subset of images). This collection contained standard arts metadata facets, including artist names, types of media, and dates, but had little in the way of content-based metadata. That is, it lacked metadata categories that describe the appearance of items or the images depicted in them, as in the case of paintings. However, many of the images did have sentential or phrasal descriptions of their contents. For example:

- *A man riding in cart drawn by two horses.*
- *Soup can, not in traditional colors: i.e. green lid, purple and orange lettering, etc.; Campbell's condensed tomato soup in purple, aqua and orange on purple background.*

We developed an algorithm to semi-automatically convert these descriptions into a set of metadata categories assumed to be useful for students and scholars of art history. This was done by comparing the words in the descriptions to their higher-level category labels in WordNet [6], and retaining a subset of the most frequently occuring categories. Some categories tended to correspond to highly ambiguous terms (e.g., "arm", "head" and other body part terms) and so were discarded. Other ambiguous words (such as "punt") only had one sense in the collection and so could be retained. Although some labels were incorrectly assigned, this algorithm worked surprisingly well. We did not directly ask our usability study participants about whether they trusted the metadata categories, but most independently volunteered comments. The majority of participants expressed pleasure at seeing content descriptors in addition to the traditional descriptors of who, what, and where. However, about one-quarter of the participants commented on a confusing or

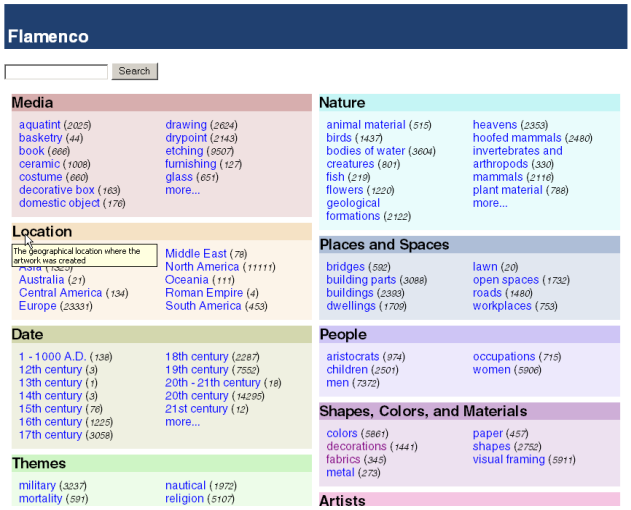


Figure 1: The opening page shows a text search box and the first level of metadata terms. Hovering over a facet name yields a tooltip (here shown below "Location") explaining the meaning of the facet.

misfiled classification; these issues did not appear to disrupt the flow of the participants' searches nor did they negatively affect their evaluation of the system. The leaf-level category labels were manually organized into hierarchical facets, using breadth and depth guidelines similar to those in [2].

INTERFACE DESIGN

The Faceted Category Interface

Unifying Goals

Our design goals are to support search usability guidelines [16], while avoiding negative consequences like empty result sets or feelings of being lost. Because searching and browsing are useful for different types of tasks, our design strives to seamlessly integrate both searching and browsing functionality throughout. Results can be selected by keyword search, by pre-assigned metadata terms, or by a combination of both. Each facet is associated with a particular hue throughout the interface. Categories, query terms, and item groups in each facet are shown in lightly shaded boxes, whose colors are computed by adjusting value and saturation but maintaining the appropriate hue.

In working with a large collection of items and a large number of metadata terms, it is essential to avoid overwhelming the user with complexity. We do this by keeping results organized, by sticking to simple point-and-click interactions instead of imposing any special query syntax on the user, and by not showing any links that would lead to zero results. Every hyperlink that selects a new result set is displayed with a query preview (an indicator of the number of results to expect).

The design can be thought of as having three stages, by rough analogy to a game of chess: the opening, middle game, and endgame. The most natural progression is to proceed through the stages in order, but users are not forced to do so.

Figure 2: Middle game (items grouped by location).

Opening

The primary aims of the opening are to present a broad overview of the entire collection and to allow many starting paths for exploration. The opening page (Figure 1) displays each metadata facet along with its top-level categories. This provides many navigation possibilities, while immediately familiarizing the user with the high-level information structure of the collection. The opening also provides a text box for entering keyword searches, giving the user the freedom to choose between starting by searching or browsing.

Selecting a category or entering a keyword gathers an initial result set of matching items for further refinement, and brings the user into the middle game.

Middle Game

In the middle game (Figure 2) the result set is evaluated and manipulated, usually to narrow it down. There are three main parts of this display: the result set, which occupies most of the page; the category terms that apply to the items in the result set, which are listed along the left by facet (we refer to this category listing as The Matrix); and the current query, which is shown at the top. A search box remains available (for searching within the current result set or within the entire collection), and a link provides a way to return to the opening.

The key aim here is organization, so the design offers flexible methods of organizing the results. The items in the result set can be sorted on various fields, or they can be grouped into categories by any facet. Selecting a category both narrows the result set and organizes the result set in terms of the newly selected facet. For instance, suppose a user is currently looking at the results of selecting the category *Bridges* from the *Places* facet. If the user then selects *Europe* from the *Locations* facet, not only is the category *Europe* added to the query, but the results are organized by the subcategories of *Europe*, namely *France*, *Italy*, and so on. Generalizing or removing a category term broadens the result set. Selecting an individual item takes the user to the endgame.

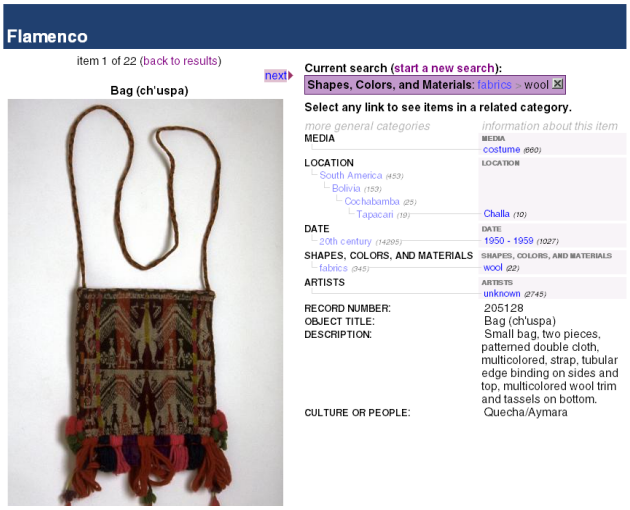


Figure 3: Endgame view of an individual item, with contextualized links for expanding the query in several conceptual directions.

Endgame

The endgame (Figure 3) shows a single selected item in the context of the current query. Next to the item, the query terms are displayed, together with an innovative hybrid-tree layout that shows all of the metadata terms assigned to the item and their locations within their hierarchies. This layout combines a simple attribute list in the right-hand column, where the most specific assigned terms can be quickly read off, with an outline tree view in the left-hand column, where each term is situated in its context within the metadata hierarchy. Selecting a metadata term switches to a new query showing all the items associated with just that term.

This view exposes metadata terms of interest, while also making it easy to navigate *laterally* through the collection. After refining a query in the middle game, a user can head in a totally new direction by choosing an image and then expanding the search from a related category in the endgame.

Keyword Matching

Each item is associated with the text of all its metadata, as well as any additional collection-specific text. The result set formed by a keyword search then contains all items whose text contains the keyword. Keyword search terms can be freely intersected with metadata query terms. In response to a keyword search, an additional panel appears at the top of the middle game display. This disambiguation panel lists all the metadata terms that contain the search key, with the search key highlighted in color wherever it appears. The user can select one of these terms to replace the keyword query term with a particular metadata term, or ignore the panel and continue to browse, leaving the keyword term in their query.

Intermediate Listings

When a query yields too many items or subcategories to show at once, an intermediate page is shown, listing all the subcategories and suggesting that the user choose one. Subcategories are listed in columns and grouped alphabetically.

System	Collection	Results Per Page	Show Cats?	Used Before
Google	Web images	20	No	27
AltaVista	Web images	15	No	8
Corbis	Photos	9–36	No	8
Getty	Photos, art	12–90	Yes	6
MS Office	Clip art, photos	6–100	Yes	NA
Thinker	Fine arts images	10	Yes	4
Baseline	Fine arts images	40	Yes	NA

Table 1: Comparison of features in popular existing image search interfaces. Show Cats? indicates whether hyperlinked categories are shown when images are selected; Used Before indicates how many study participants had previously used an interface.

Implementation

The system is built using Python, MySQL, and the WebWare toolkit[1]. All components of the interface are dynamically generated, based on the facets and facet values defined in a relational database. Query previews are generated using the SQL `COUNT(*)` and `GROUP BY` operators to count the number of items that fall into each subcategory.

The Baseline Interface

Today many users are familiar with keyword-based image search, as embodied by Web image search engines. Table 1 compares some of the features of 5 image search engines: Google Image Search, AltaVista Image Search, Corbis, GettyImages, and MS Office Clipart, in addition to The Thinker, the search engine currently available for the art history collection used in our study.

When the user selects an image for detailed viewing, three systems (Getty, MS Office, and The Thinker) show related topical category labels, hyperlinked to act as queries (e.g., showing the categories *Flowers* and *Nature* next to an image of poppies). The categories are not explicitly faceted or hierarchical, and are usually not shown in any meaningful order.

To create a fair comparison of search interfaces, we built an image search system that is representative of the best aspects of the six image search engines in Table 1. When in doubt we usually opted to make the baseline resemble Google Image Search, due to its familiarity to the user population.

The starting page for the baseline interface provides an entry form for typing in search terms, an illustrative image, a two-sentence description of the collection (mimicking the starting page of The Thinker), and some information on how to search the collection. If multiple search terms are entered in the query, they are implicitly ANDed, as this practice has become widely adopted due to Google's use of it. Only one participant (in the pre-test) asked about doing advanced Boolean queries. Adjacent words enclosed in quotation marks are treated as phrases. Stemming is not done, both because of the confusion it can cause [11], and because Google does not use it.

[1] www.python.org,www.mysql.com,webware.sourceforge.net

After the user enters search terms, a linked list of pages of search results is shown, along with a description of how many images were found as a result of the query. The images are displayed in a table of 10 rows of 4 images each, in alphabetical order according to image title[2]. The user can click through a page at a time, enter a new query in the search form that appears at the top of each page (the default is to search the entire collection), or click on a particular image to see more detail.

In the detailed view, a larger version of the image is shown along with a listing of its associated metadata. In addition, the baseline has a feature that makes it more powerful than the other keyword search systems. It shows a hyperlinked list of category labels that translate into queries on the corresponding category label in the faceted category interface. For example, if an image has been assigned the category label *Bridge* in the faceted category interface, the detailed view of that image in the baseline interface also includes a hyperlink to a query that retrieves all items in the *Bridge* category. The categories are shown in alphabetical order, but no preview is shown of the number of items in the category. Thus, here the baseline interface departs from the Google design in order to incorporate functionality roughly equivalent to the category views provided by the other systems in Table 1.

Since the baseline interface does not need to compute query previews, it is much faster than the faceted category interface. Using our records of actual queries performed during the studies described below, we measured the average processing time for the category interface to be an order of magnitude longer than that of the baseline interface.

Prior Work

To develop the target interface, we followed standard interface design practice. Beginning with the domain of architectural design, we did an ethnographic study of how architects search for and use images as inspiration for design [5]. This was followed by a cycle of low-fidelity prototyping, informal usability testing, and redesign. After this, we conducted two rounds of development and two usability studies. These studies were useful for answering questions about various design features, and determining whether users would respond well to navigation of multiple simultaneous hierarchical facets. However, up to this point we had not compared the design to a more standard baseline, to determine if this richer method of search would be preferred and more effective over a more standard interface. Hence this paper presents the results of a new study to answer the question: is this design better than the current state of the art in image search interfaces?

[2]It is difficult to determine the ranking algorithm used by the Web search engines; presumably it is a function of the match of the query terms to the words near the images where they are found. The other systems do not seem to have a ranking function; three systems allow grouping according to broad categorical features such as color vs. black-and-white or media type.

USABILITY STUDY

Participants

Working with participants who are interested in the collection in question has been found to be especially important in search usability studies [3]; this has been our experience as well. We chose to use a fine arts collection for this study because it was possible to recruit art history students and people who have recently taken art courses as the study participants. Data from 32 participants was used in the analysis. (A pre-test was conducted on three participants and data for two outliers was discarded.) The participants were all regular users of the Internet, searching for information either every day or a few times a week. They searched for images online less frequently, with the majority searching for images less than once per week. Table 1 summarizes their familiarity with various image search systems; four people had previously used the Fine Arts image collection with its official Web interface, The Thinker.

Apparatus

Participants received a $15 gift certificate for participating in a session that lasted about 1.5 hours. We tested all the participants in a lab setting, using Internet Explorer 6 on Windows 2000 workstations with 21-inch monitors set at 1280 by 1024 pixels in 24-bit color. Data was recorded with multiple methods: (a) server logs, (b) behavioral logs (time-stamped observations), and (c) paper surveys after each task, each interface, and at the end of the session. One or two experienced usability analysts conducted the sessions; when two were available, one analyst took written notes while the other facilitated the session. We collated data from all the sources to create a complete record of each test session.

Design and Procedure

The study used a within-subjects design. Each participant used both the faceted category interface (henceforth FC) and the baseline interface; each interface was the starting view for half the participants. The interfaces were assigned neutral names (“Denali” for FC and “Shasta” for Baseline).

In earlier studies we walked participants through the features of the experimental interfaces. By contrast, and to better mimic the situation that occurs in practice, in this study participants were not introduced to the features nor told anything in advance about the systems other than that they both accessed the same collection of 35,000 fine arts images. We did inform participants that keyword searching was available in both interfaces and briefly explained the text search syntax (the use of quotation marks to delimit phrases).

Throughout the study, subjective ratings were reported on a 9-point Likert scale, with 1 meaning “strongly disagree”, 9 meaning “strongly agree”, and 5 meaning “neutral”. Because we have found that participants tend to be generally positive about the current interface, we adopted a wide range in order to have a more sensitive testing instrument.

Tasks

The tasks were designed to reflect the contents of the collection and the art history background of the students. Participants completed four tasks on each interface, two structured and two unstructured:

1. (3 min, unstructured). Search for images of interest.
2. (11-14 min, structured). Gather materials for an art history essay on a given topic. Complete 4 subtasks, ranging from very specific to more open ended, e.g.: (i) Find all woodcuts created in the United States; (ii) choose the decade for which the collection seems to have the most images of U.S. woodcuts; (iii) select one of the artists who worked during this period and show all of his or her woodcuts; (iv) choose one of the subjects depicted in these works and find another U.S. woodcut artist who has treated the same subject in a different way.
3. (10 min, structured). Compare related images in order to write an essay, e.g.: Find images by artists from two different countries that depict conflict between peoples.
4. (5 min, unstructured). Search for images of interest.

Task 2 used metadata categories clearly visible in the start page and matrix of FC. However, we carefully framed the wording of Task 3 so as not to reflect the wording of a particular facet. Each of Tasks 2 and 3 had two versions; study design was balanced in terms of which queries were assigned to each interface. At the end of the session, we asked participants whether they felt the structured queries were equally difficult; 30 out of 32 stated that they were equivalent. As a double-check, we looked at the difficulty ratings in the post-task questionnaires for the different tasks; we found no significant differences between the two task sets (both t's < 1.7, both p's > 0.05).

Results

It is difficult to evaluate browsing tasks, since there are no correct answers and since the goal is not necessarily to minimize time used. Thus the tasks and measures were designed to test the following hypotheses about FC:

1. Participants will experience greater search satisfaction and success in FC than in the Baseline, feel greater confidence in the results, produce higher levels of recall, and encounter fewer dead ends.
2. Overall, FC will be perceived to be more useful and flexible than the Baseline.
3. Using FC, participants will feel more familiar with the contents of a collection.
4. Participants will use FC to perform multiple-facet queries during their self-directed searches.

Task Satisfaction and Success

After each structured task, participants completed a short questionnaire. Using FC, participants felt significantly more confident that they had found all of the relevant images in the collection (Task 2: $t(62) = 2.18$, $p < .05$, Task 3: $t(62) = 2.03$, $p < .05$) and significantly more satisfied with the results (Task 2: $t(62) = 3.78$, $p < .001$, Task 3: $t(62) = 2.03$, $p < .05$) than when they used Baseline (thus supporting hypothesis 1).

We evaluated participant success in retrieving all the relevant images for part (a) of Task 2, which was to find all woodcuts created in the United States or all aquatints created in France. In Baseline, 57% of the participants conducting the aquatints task retrieved all the relevant results; in FC, 81% of the participants were successful. For the woodcuts task, 21% of those using Baseline and 77% using FC managed to retrieve all the relevant images (thus supporting hypothesis 1). The differences were caused in part by the Baseline users not querying both singular and plural forms of words.

Participants indicated that they more often found themselves at a dead-end or empty results when using Baseline; this difference was not significant (Task 2: $t(62) = 1.41$, $p = .163$, Task 3: $t(62) = .499$, $p = .619$). However, during the structured tasks participants actually did receive empty results in Baseline 82 times, while in FC, they received empty results only 26 times (thus supporting hypothesis 1).

For search success, we also looked at how many items users opted to bookmark in each system and the usefulness ratings (on a scale from 1 to 10) for those items. In Baseline, participants rated 266 items with an average rating of 8.1; in FC, participants rated 215 items with an average rating of 7.9. In Baseline, participants may have been able to rate more items because the processing speed was so much faster than in FC. The differences in item ratings were not significant ($t(481) = 1.12$, $p = .26$).

As indicated above, all tasks were assigned time limits, but participants were allowed 3 extra minutes on Task 2 when using FC because of its slower response time.[3] Participants could complete a task before the time limit had expired. We did not encourage participants to rush through the searches; instead, we asked them to search as they normally would.

Participants spent an average of 9 min 30 s on Task 2 using Baseline; in FC, the time spent on this task averaged 12 min 6 s. For Task 3, participants spent 7 min 45 s in Baseline and about 9 min in FC. These differences were significant (both p's $< .05$), but may be caused by the slower processing time, and the fact that system errors occurred during 5 of the 32 sessions with FC; restarting the system added time to the tasks. Thus FC did not result in faster usage times; however, we had not hypothesized that it would, given that success in browsing tasks is not reflected by faster completion times.

Post-Test Interface Comparison

In the post-interface assessment, much stronger differences emerged. Immediately after completing the fourth task on

[3] For Task 2, the average processing time per step was 0.3 s for Baseline, but 3.7 s for FC. For Task 3, this was 0.37 s for Baseline, but 4.3 s for FC.

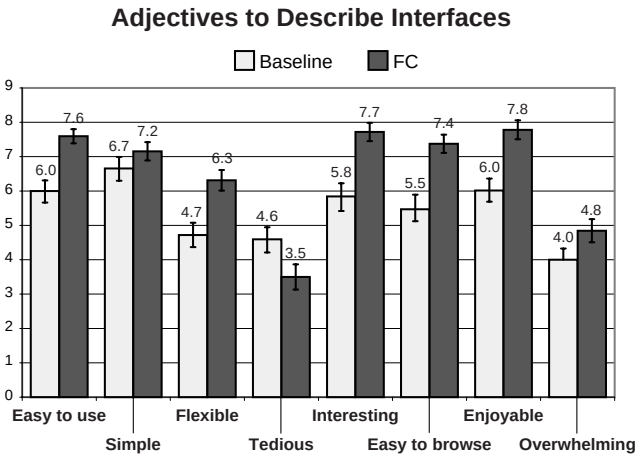


Figure 4: Post-interface assessments. All results were statistically significant at $p < .001$ except "simple" and "overwhelming"; "tedious" was significant at $p < .05$.

an interface, participants completed an interface evaluation. FC received more positive ratings than Baseline for nearly every measure, as shown in Figure 4. Noteworthy ratings are those for "easy to use" and "easy to browse." Given FC's complex screen design, it is remarkable that users assigned it an average rating of 7.6 for "simple." Similarly, the fact that FC was not rated to be significantly more "overwhelming" than Baseline ($t(62) = 1.79, p > .05$) testifies to the success of the design. Participants indicated they were more likely to use FC in the future ($t(62) = -3.75$, $p < .001$). They also felt more familiar with the collection ($t(62) = -2.17$, $p < .05$). These results support hypotheses 1, 2, and 3.

The order in which interfaces were viewed had a strong effect on these ratings. When FC was viewed first, the interface ratings for Baseline were considerably lower than when Baseline was the first interface shown ($t(26) = 2.67, p < .01$). The ratings for FC were not significantly affected by being viewed after Baseline ($t(26) = -0.27, p = .783$).

Participants were also asked to compare Baseline to FC and indicate which interface they preferred for different situations (see Table 2). For finding images of roses (a simple, single-facet task), about 50% preferred Baseline. However, for every other type of searching, FC was preferred: 88% said that FC was more useful for the types of searching they usually do and 91% said they preferred FC to Baseline overall. Those who preferred the Baseline commented on its simplicity and stated that the categories felt too restrictive.

Facet Usage

Facet usage in the structured tasks was driven largely by the task content, causing participants to focus on Date, Location, Media, Artist and Theme. However, for the unstructured searches, usage was more evenly distributed across all the facets. Artists (17%), Date (15%) and Location (15%) were the most used facets on the start page, but 111 starts occurred in the other facets with percentages ranging from 5% to 12%. For refining queries, again Artist (20%), Date (14%),

Which interface would you rather use for these tasks?	Baseline	FC
Find images of roses	15	16
Find all works from a certain time period	2	30
Find pictures by 2 artists in the same media	1	29
Overall assessment:	**Baseline**	**FC**
More useful for your usual tasks	4	28
Easiest to use	8	23
Most flexible	6	24
More likely to result in dead-ends	28	3
Helped you learn more	1	31
Overall preference	2	29

Table 2: Post-test preferences for the Baseline and Faceted Category (FC) interfaces.

and Location (19%) were most used, but the other facets were used for 6–11% of the refining actions (n=139). In the endgame, participants opted to create a new query by clicking on Artist 39%, Media 29%, and Shapes 19% of the time (n=21).

The number of facets used simultaneously was also of interest to us, since this is a unique aspect of FC. Participants constructed queries from multiple facets in the unstructured tasks 19% of the time and in the structured tasks 45% of the time, thus supporting hypothesis 4. However, when browsing only a single facet, participants frequently used "search within results" to refine their searches (15% for unstructured, 50% for structured).

Qualitative Observations

Users of the Baseline commented favorably on its simplicity and similarity to Google image search, but also noted that the category hyperlinks made it much easier to use.

Many participant reactions to FC followed a pattern. When shown the starting page, more than half explicitly remarked on it, noting that it was "well-organized" and gave them "ideas about what to search for". The query previews were a key ingredient for 9 users, who offered unsolicited comments on this feature's usefulness: "The collection seems more complete because I can tell how many are available in different categories from the front page."

Once participants tried their first queries, more than half of them commented negatively on the speed. Some wondered aloud about the cause of the slowness, a few said it was "frustrating" and "annoying", and one person commented, "At this point, I would go to a different search engine." In the middle game, more than half of the participants explicitly remarked on the matrix, saying favorable things such as that it "prompted" them about where to go next. They also generally liked seeing the images grouped into categories: "It does a lot of the work for you, the searching and the categorizing." Three were confused about how the matrix functioned: they thought it was a repetition of the first page and did not realize

they could use it to refine their existing query. All other participants did understand the matrix and stated they felt more confident in the results they obtained by browsing. Participants liked having category links in the endgame of both interfaces, but 9 out of 32 explicitly commented on the level of detail in FC, stating that the information here was "useful" and "very clear", "guiding" them through a search.

As participants continued to use the interface, they became more comfortable with it. As an example interaction sequence, one participant began Task 3 (to compare images on conflict between people) by clicking on *military* at the start page, then refining from an intermediate page to choose *war*. Since there were 824 results, he refined his search further by doing a keyword search within results for *sword*, reducing the number of images to 74. He grouped the results by *artist*, since the task called for him to contrast works by two artists. Then he began clicking on images and started formulating his thesis: "This is the Napoleonic view of war—the camera is really far away. Men look like ants and you don't see war itself, the death, just the preparations." It occurred to him that 20th-century depictions of war are more graphic. He grouped his 74 results by date and quickly found images by Goya that "zoom in on the misery and suffering" of war.

At the end of the session, participants expressed enthusiasm for the FC interface, wanting to know when it would be available for them to use. One participant said, "I wish I had this when I was writing papers." The participants found it "interesting", "enjoyable", and "easy to customize" their searches using the FC interface.

CONCLUSIONS AND FUTURE WORK

We have designed an image access interface that allows users to navigate a large collection using hierarchical faceted metadata in a flexible manner. Despite the fact that the interface was often an order of magnitude slower than a standard baseline, it was strongly preferred by most study participants. These results indicate that a category-based approach is a successful way to provide access to image collections.

We are in the process of developing algorithms to make the query preview generation faster. This is important for future attempts to make the method scale to collections that are one or two orders of magnitude larger. We also plan in the future to perform studies comparing this approach directly to similarity-based approaches, as well as studying the effects of adding personalization, history, and relevance feedback functionality to the design, and investigating the efficacy of the method on text collections.

ACKNOWLEDGEMENTS

We thank Andrea Sahli for extensive help with the usability tests, Rashmi Sinha and Kevin Chen for helpful discussion, the San Francisco Fine Arts Museum for access to the dataset, and all our usability study participants. This research was funded by an NSF CAREER grant, NSF-IIS 9984741 and an IBM Faculty Fellowship.

REFERENCES

1. L. H. Armitage and P. G. B. Enser. Analysis of user need in image archives. *Journal of Information Science*, 23(4):287–299, 1997.
2. M. L. Bernard. Examining the effects of hypertext shape on user performance. *Usability News*, 4(2), 2002.
3. P. Borland and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3):225–250, 1997.
4. C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*, 1999.
5. A. M. Elliott. *Computational Support for Sketching and Image Sorting During the Early Phase of Architectural Design*. Ph.D. dissertation, University of California, Berkeley, 2002.
6. C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
7. S. R. Garber and M. B. Grunes. The art of search: A study of art directors. In *Proc. of CHI-92*, Monterey, CA, 1992.
8. M. Hearst, J. English, R. Sinha, K. Swearingen, and K.-P. Yee. Finding the flow in web site search. *Communications of the ACM*, 45(9), September 2002.
9. J. M. Jose, J. Furner, and D. J. Harper. Spatial querying for image retrieval: a user-oriented evaluation. In *Proceedings of ACM SIGIR '98*, pages 232–240, 1998.
10. M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1:259–285, 2000.
11. J. Muramatsu and W. Pratt. Transparent queries: Investigating users' mental models of search engines. In *Research and Development in Information Retrieval*, pages 217–224, 2001.
12. W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, and C. Faloutsos. The QBIC project: Querying images by content using color, texture, and shape. *SPIE: Storage and Retrieval for Image and Video Databases*, 1908, 1993.
13. M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. S. Huang. Supporting similarity queries in MARS. In *ACM Multimedia*, pages 403–413, 1997.
14. C. Plaisant, B. Shneiderman, K. Doan, and T. Bruns. Interface and data architecture for query preview in networked information systems. *ACM Transactions on Information Systems*, 17(3):320–341, 1999.
15. K. Rodden, W. Basalaj, D. Sinclair, and K. R. Wood. Does organisation by similarity assist image browsing? In *Proceedings of ACM SIGCHI 2001*, pages 190–197, 2001.
16. B. Shneiderman, D. Byrd, and W. B. Croft. Sorting out searching: A user-interface framework for text searches. *Communications of the ACM*, 41(4):95–98, 1998.
17. R. K. Srihari, Z. Zhang, and A. Rao. Intelligent indexing and semantic retrieval of multimodal documents. *Information Retrieval*, 2(2/3):245–275, 2000.
18. R. C. Veltkamp and M. Tanase. Content-Based Image Retrieval Systems: A Survey. Technical Report UU-CS-2000-34, Dept. of Computing Science, Utrecht University, 2000.

Clustering versus Faceted Categories for Information Exploration

By Marti A. Hearst

To Appear in Communications of the ACM, April 2006

Information seekers often express a desire for a user interface that organizes search results into meaningful groups, in order to help make sense of the results, and to help decide what to do next. A longitudinal study in which participants were provided with the ability to group search results found that they changed their search habits in response to having the grouping mechanism available [2] .

There are many open research questions about how to generate useful groupings and how to design interfaces to support exploration using grouping. Currently two methods are quite popular: *clustering* and *faceted categorization*. In this article I describe both approaches and summarize their advantages and disadvantages, based on the results of usability studies.

Clustering

Clustering refers to the grouping of items according to some measure of similarity. In document clustering, similarity is typically computed using associations and commonalities among features, where features are typically words and phrases [1]. One of the better implementations of clustering of web results can be found at Clusty.com.[1]

The greatest advantage of clustering is that it is fully automatable and can be easily applied to any text collection. Clustering can also reveal interesting and potentially unexpected or new trends in a group of documents. A query on “New Orleans'' run on Clusty.com on September 16, 2005 (shortly after the devastation wreaked by hurricane Katrina), revealed a top-ranked cluster titled *Hurricane*, followed by the more standard groupings of *Hotels, Louisiana, University*, and *Mardi Gras*.

Clustering can be useful for clarifying and sharpening a vague query, by showing users the dominant themes of the returned results [2]. Clustering also works well for disambiguating ambiguous queries; particularly acronyms. For example, ACL can stand for Anterior Cruciate Ligament, Association for Computational Linguistics, Atlantic Coast Line Railroad, etc. Unfortunately, because clustering algorithms are imperfect, they do not neatly group all occurrences of each acronym into one cluster, nor do they allow users to issue followup queries that only return documents from the intended sense (e.g., “ACL meeting'' will return meetings for multiple senses of the term).

An under-appreciated aspect of clusters is their utility for eliminating groups of documents from consideration. This result is supported by participant comments found in several studies [2,3]. For example, if most documents in a set are written in one language, clustering will very quickly reveal if a subset of the documents is written in another language.

[1] Some of Clusty.com's power comes from performing metasearch and showing only the top-ranked results. This alone can produce improved results, since it combines the power and judgement of several different search engines' rankings.

The disadvantages of clustering include their lack of predictability, their conflation of many dimensions simultaneously, the difficulty of labeling the groups (Clusty.com's top-level labels are among the best implementations), and the counter-intuitiveness of cluster subhierarchies. Some algorithms [2,8] build clusters around dominant phrases, which makes for understandable labels, but whose contents do not necessarily correspond to those labels.

To illustrate these weaknesses, consider a recipe example, chosen because the relevant dimensions are familiar to most people and because exploration and browsing are natural tasks for recipe collections. A search for “chicken recipes" on Clusty.com (also on Sept. 16, 2005) turns up the following motley assortment of groups:

Salad
Crockpot
Chicken Breast
Barbeque/Grilled
Soup
Recipes
Healthy
Lowfat
Easy Chicken Recipes
Italian

This list is incomplete (contains large gaps), and inconsistent. Why *Crockpot* and *Barbeque/Grilled* , but not *Baked* and *Fried* ? Why *Chicken Breast* but not *Leg* and *Wing* ? Why *Salad* and *Soup* but not *Main course*? Why *Italian* recipes but not *Indian, Thai*, or *French*? Furthermore, drilling down into the hierarchies rarely reveals intuitive results. The 29 documents listed under *Salad* are organized by the labels:

Complete selection of Trusted Chicken Recipes
Cakes
Better Homes and Gardens
Collection
Share
Boneless Chicken Breast
Pasta Salad,

and so on. Only *Pasta Salad* really belongs here as a label; it does not make sense for *Boneless Chicken Breasts* to appear in this cluster rather than in the *Chicken Breasts* cluster, and clearly *Cakes* belongs in a *Dessert* category alongside *Salad* and *Soups*.

These kinds of errors are quite typical for clustering output. Usability results show that users do not like disorderly groupings like these, prefering understandable hierarchies in which categories are presented at uniform levels of granularity [4,5].

Hierarchical Faceted Categories

A category system is a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain. They are usually created manually, although assignment of documents to categories can be automated to a certain degree of accuracy. Good category systems have the characteristics of being coherent and (relatively) complete and thus pose an advantage over the unpredictable results of clustering; the studies that compare the two find that participants prefer categories [4,5] .

A question arises as to what kind of category structure is most effective for exploration and browsing of information collections.There is increasing recognition that strictly hierarchical organization of categories is impoverished for these uses.

An alternative representation, intermediate in complexity and very rich in flexibility, has become influential over the last few years. This representation is known as hierarchical faceted categories (HFC) [7]. The main idea is quite simple. Rather than creating one large category hierarchy, build a set of category hierarchies each of which corresponds to a different facet (dimension or feature type) relevant to the collection to be navigated. In the case of chicken (and other) recipes, these category hierarchies can include Dish Type (*Main, Soup, Salad, Side, Dessert*), Ingredient Type (*Meat, Vegetables, Grains, Spices*), Cooking Method (*Bake, Fry, Grill, Easy*), Cuisine Type (*Italian, Indian, French*) , etc. Each facet has a hierarchy of terms associated with it.

After the facet hierarchies are designed, each item in the collection can be assigned many labels from the hierarchies. Thus a recipe for "Chicken Noodle Casserole" might be assigned:

Dish Type > Pasta
PreparationType > Baking
Meat > Poultry > Chicken
Vegetables > Celery
Vegetables > Carrot

and so on. Our research group has been investigating how to build an intuitive interface for exploration and discovery within information collections using hierarchical faceted categories; we call the resulting interface framework Flamenco [7] (flamenco.berkeley.edu).

This kind of interface allows flexible ways to access the contents of the underlying collection. For example, from the *Meat* facet, a user can choose to select the *Poultry* subcategory, and from this select in turn the *Chicken* subcategory. The user can choose any other facet, perhaps Dish and Courses, and from this select the *Pasta* category, and then group the resulting recipes by *Vegetables* , or *Preparation Type*, or any other facet (see Figure 1). Navigating within the hierarchy naturally builds up a complex query that is a conjunction of disjunctions over subhierarchies.

An interface using HFC simultaneously shows previews of where to go next, and how to return to previous states in the exploration, while seamlessly integrating free text search within the category structure. The approach reduces mental work by promoting recognition over recall and suggesting logical but perhaps unexpected alternatives at every turn, while at the same time avoiding empty results sets. This organizing structure for results and for subsequent queries can act as scaffolding for exploration and discovery.

We have conducted a series of usability studies that find that, for browsing tasks especially, HFC-enabled interfaces are overwhelmingly preferred over the standard keyword-and-results listing interfaces used in web search engines [7]. Study participants find the design easy to understand, flexible, and less likely to result in dead ends.

One drawback of HFC interfaces (as opposed to clusters) is that the categories of interest must be known in advance, and so important trends in the data may not be shown. But by far the largest drawback is the fact that in most cases the category hierarchies are built by hand and automated assignment of categories to items is only partly successful.

Our group has recently made some progress in the problem of nearly-automatic creation of hierarchical faceted categories [6]. A portion of the output of the system applied to the text of a recipe collection is shown in Figure 1. The algorithm, which makes use of the WordNet hierarchy, draws out detailed categories for ingredients, dishes, and (unexpectedly) cooking equipment and people, but misses facets such as cuisine. We call the algorithm nearly-automated, since the results require some

hand-editing. There is much room for improvement, and we see automatic creation of faceted hierarchies as an important area for research.

Impact and the Future

To date both HFC and clustering are boutique search interfaces; they are applied and used primarily in domain specific collections. There are many movements afoot to promote larger scale use of metadata more generally. Hierarchical faceted metadata is already common in many e-commerce interfaces; for example, eBay and Shopping.com are experimenting with different variations of the idea, and Endeca.com provides a custom solution. It is probably possible to automatically
impose a faceted structure onto grass-roots created tag collections such as those seen at Flickr. However, it is an open question whether these will eventually be widely and regularly used on the open-domain web.

Acknowledgements. This research was funded in part by NSF IIS-9984741. Thanks to Flamenco team members including Ping Yee, Kevin Li, Emilia Stoica, Ame Elliott, Rashmi Sinha, and Kirsten Swearingen.

[1] Cutting, D.R., Pedersen, J. O., Karger, D., and Tukey, J. W. Scatter/Gather: A cluster-based approach to browsing large document collections. In Proceedings of ACM SIGIR, Copenhagen, Denmark, 1992.

[2] Kaki, M. Findex: Search result categories help users when document ranking fails. In Proceedings of ACM SIGCHI, 2005.

[3] Kleiboemer, A.J., Lazear, M.B., and Pedersen, J. O. Tailoring a retrieval system for naive users. In Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval (SDAIR), Las Vegas, NV, 1996.

[4] Pratt, W., Hearst, M., and Fagan, L. A knowledge-based approach to organizing retrieved documents. In Proceedings of 16th Annual Conference on Artificial Intelligence (AAAI 99), Orlando, FL, 1999.

[5] Rodden, K., Basalaj, W., Sinclair, D., and Wood, K.R. Does organisation by similarity assist image browsing? In Proceedings of ACM SIGCHI 2001, Seattle, WA, 2001.

[6] Emilia Stoica and Marti Hearst. Nearly-automated metadata hierarchy creation. In HLT-NAACL'04, Companion Volume, Boston, MA 2004.

[7] Yee, K-P., Swearingen, K., Li, K., and Hearst, M. Faceted metadata for image search and browsing. In Proceedings of the CHI 2003, Fort Lauderdale, FL, April 2003.

[8] Zamir, O., and Etzioni, O. Grouper: A dynamic clustering interface to web search results. In Proceedings of WWW8, Toronto, Canada, 1999.

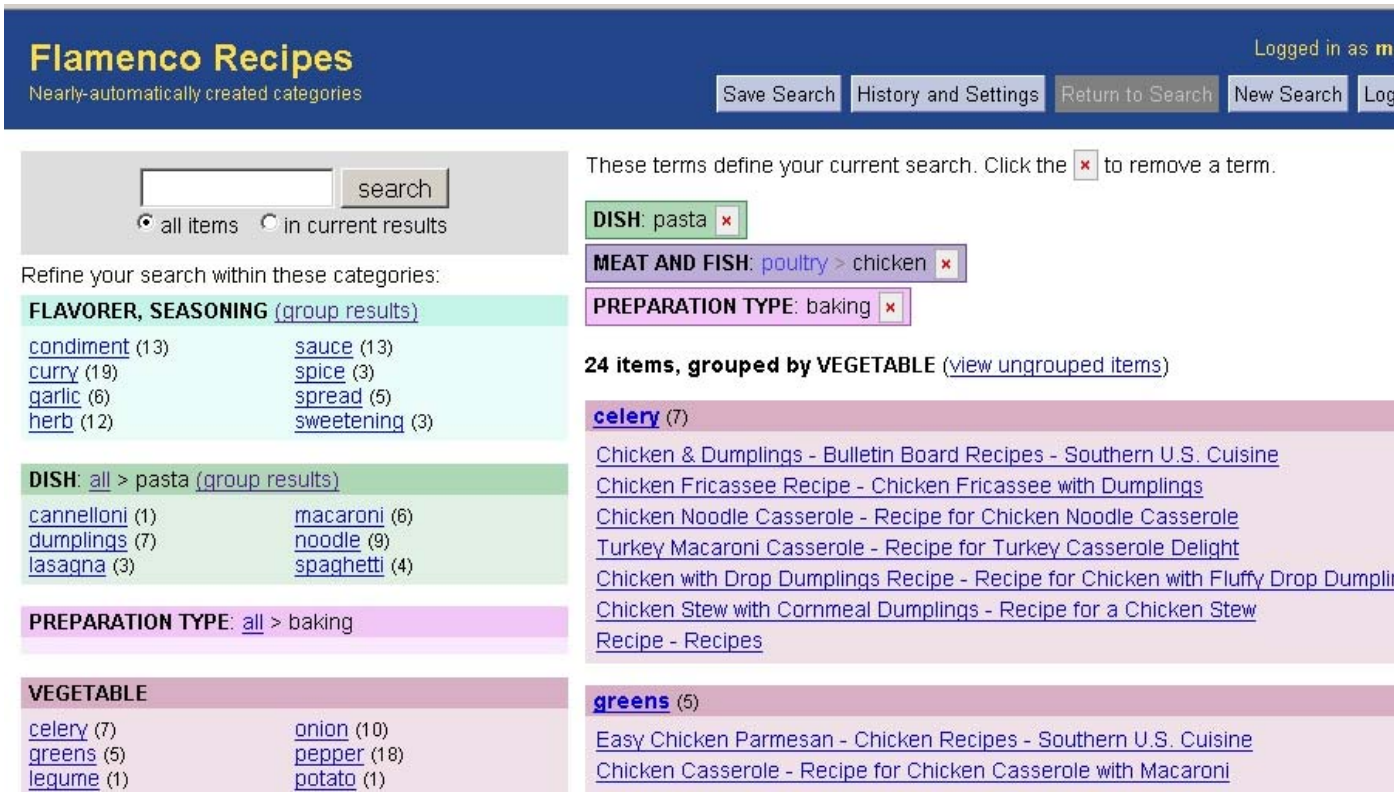


Figure 1, Caption: Navigating a Recipes Collection using Hierarchical Faceted Categories (partial results).