Faceted Metadata for Information Architecture and Search

CHI Course - April 24, 2006
Session I

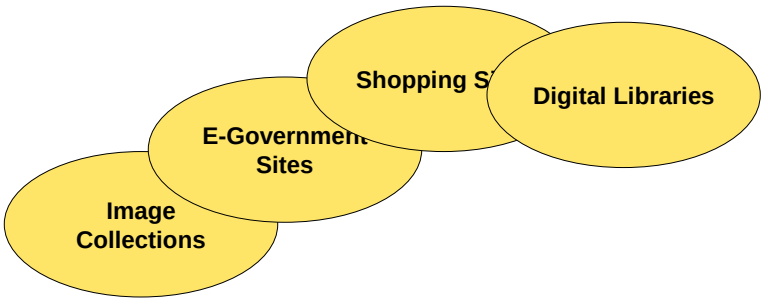
Marti Hearst, School of Information, UC Berkeley
Preston Smalley & Corey Chandler, eBay User Experience & Design

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

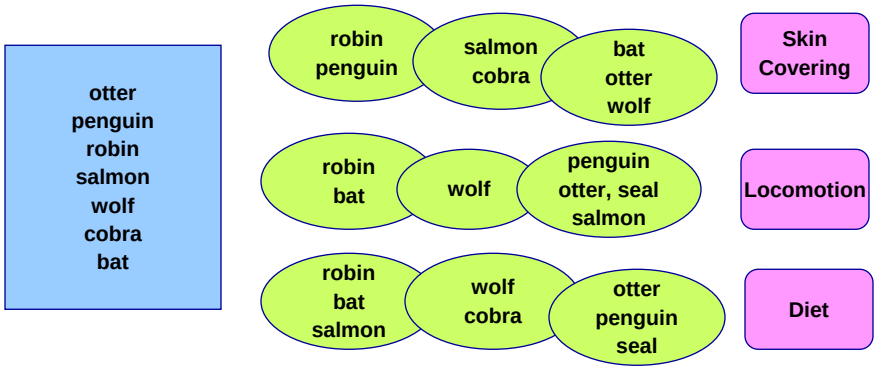
Main Idea

- 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



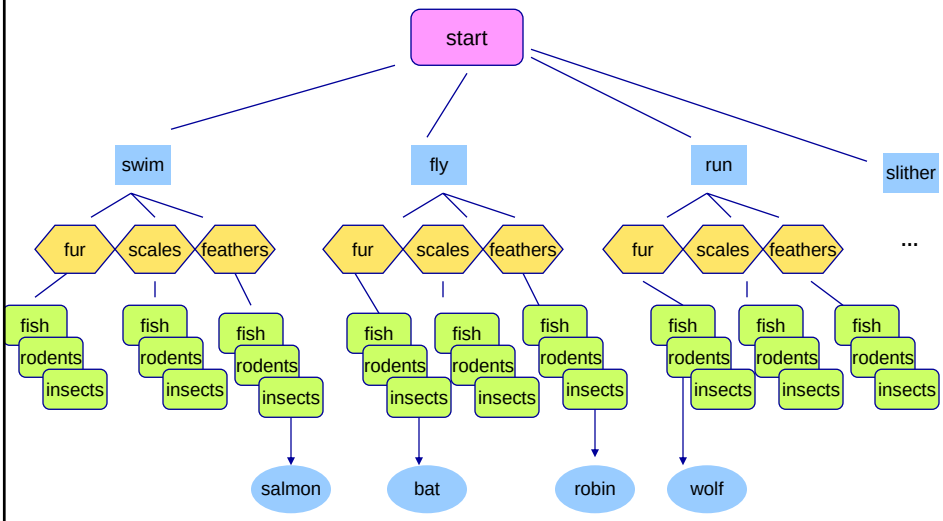
7

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 interface makes me start with that category?
- Wasteful
  - Have to repeat combinations of categories
  - Makes for extra clicking and extra coding
- Difficult to modify
  - To add a new category type, must duplicate it everywhere or change things everywhere

8

The Problem With Hierarchy



9

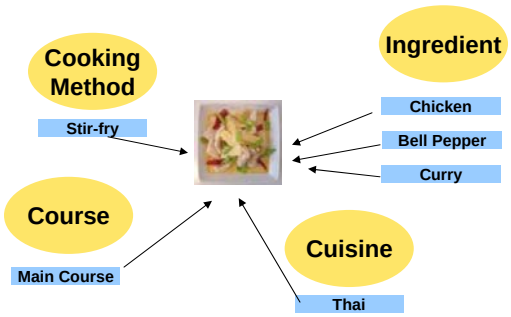
The Idea of Facets

- 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
  - Sherbet
  - Flan

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

Fruit > Pineapple
Dessert > Cake
Preparation > Bake

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

13

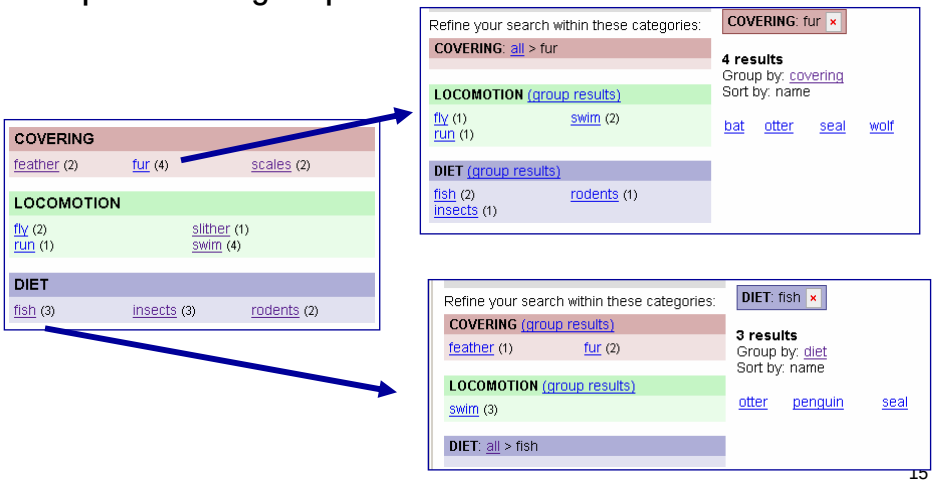
Using Facets

- 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

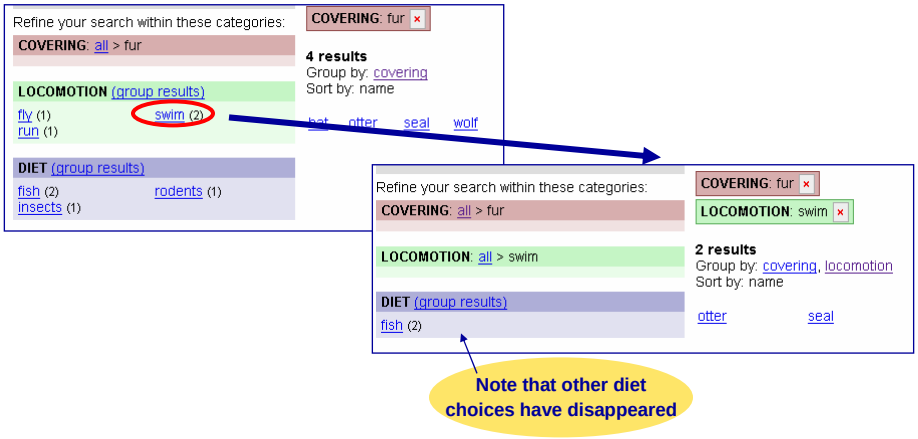
The Advantage of Facets

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



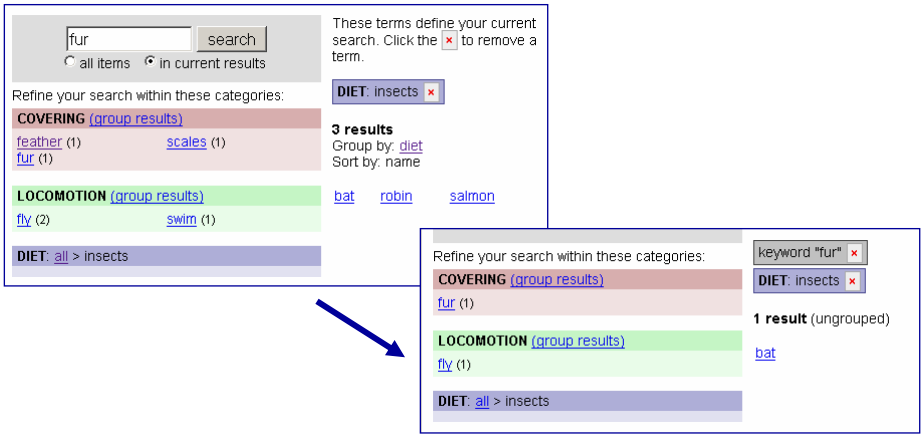
The Advantage of Facets

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



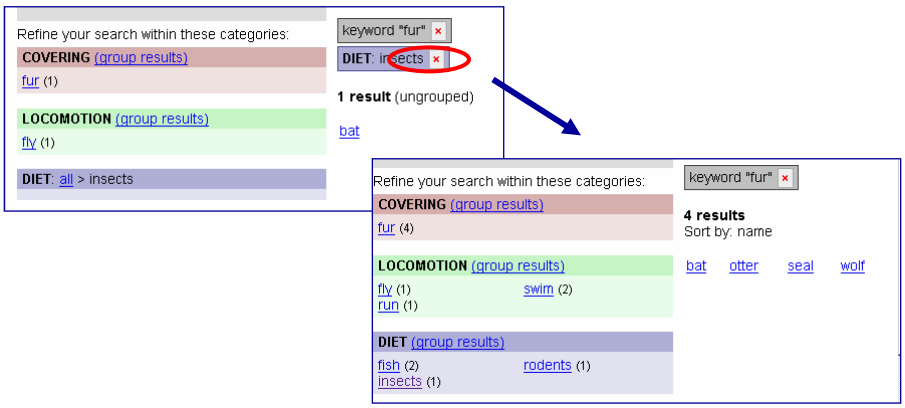
The Advantage of Facets

- 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

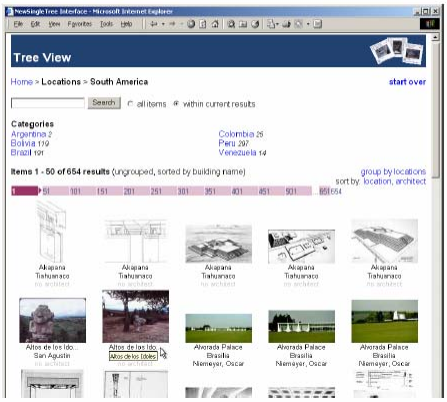
Information previews

- 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?

Photo by J. Hearst, jhearst.typepad.com

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”, “fawn”, “doe”

24

The Chess Analogy

25

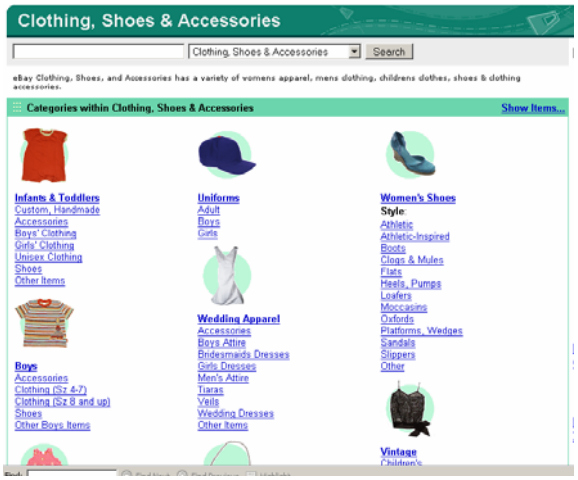
Analogy: Chess

- 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.

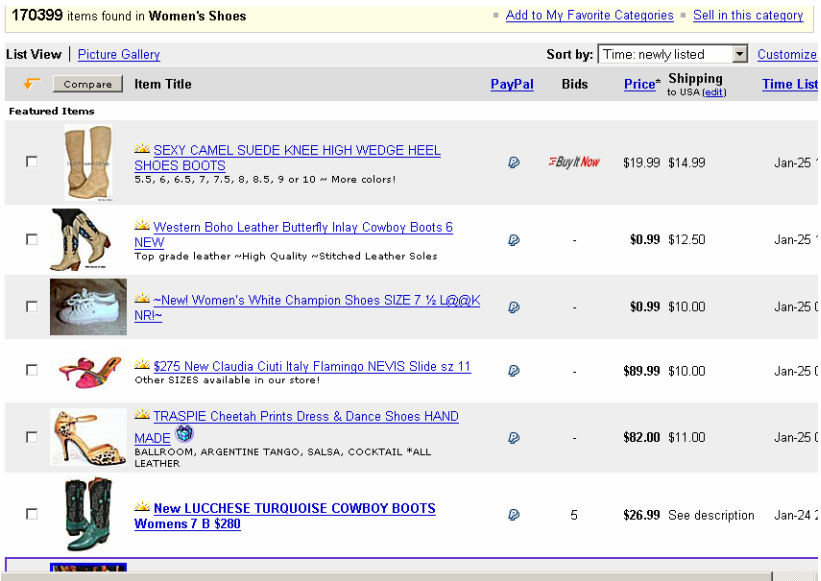
26

The Opening

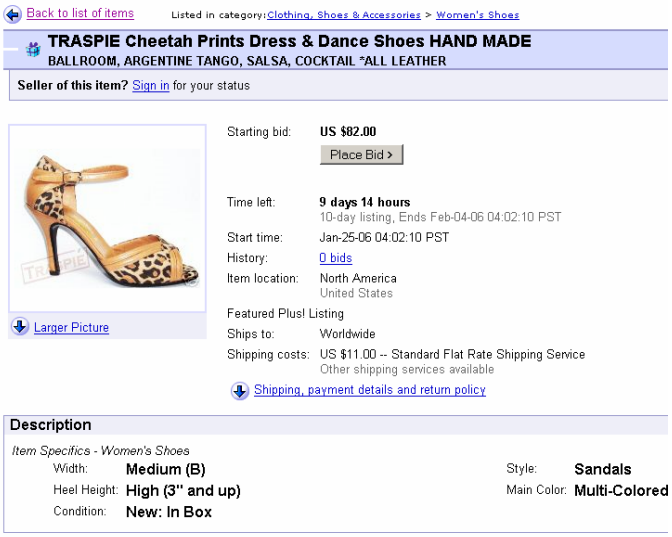
- Usually exposes top-level hierarchy or top-level facets
- Usually also has a search component



The Endgame – Penultimate Pages



The Endgame – Content Pages



29

The Middlegame

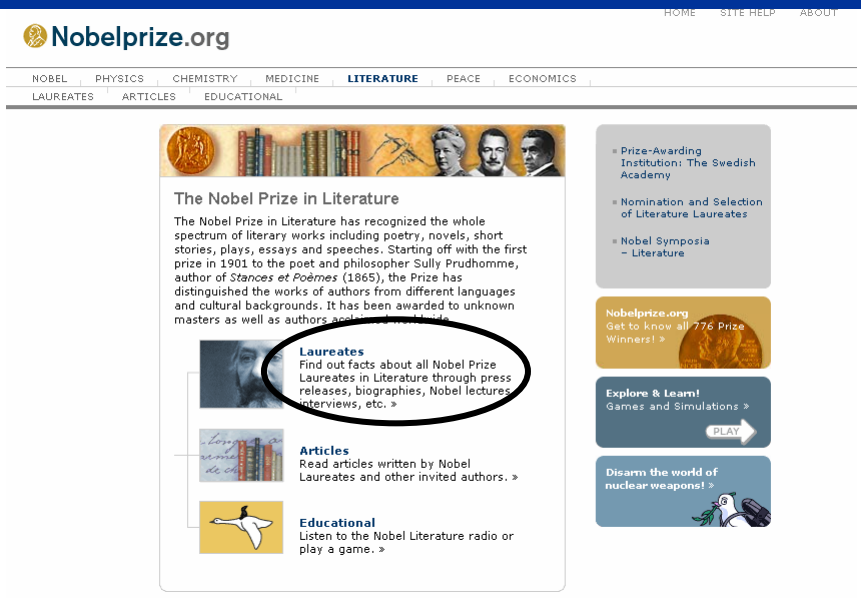
- The heart of the navigation experience
- There is a strategic advantage to having a good middlegame
  - Standard Web search doesn't handle this well
- This is where the flexible faceted metadata approach can work best.

30

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

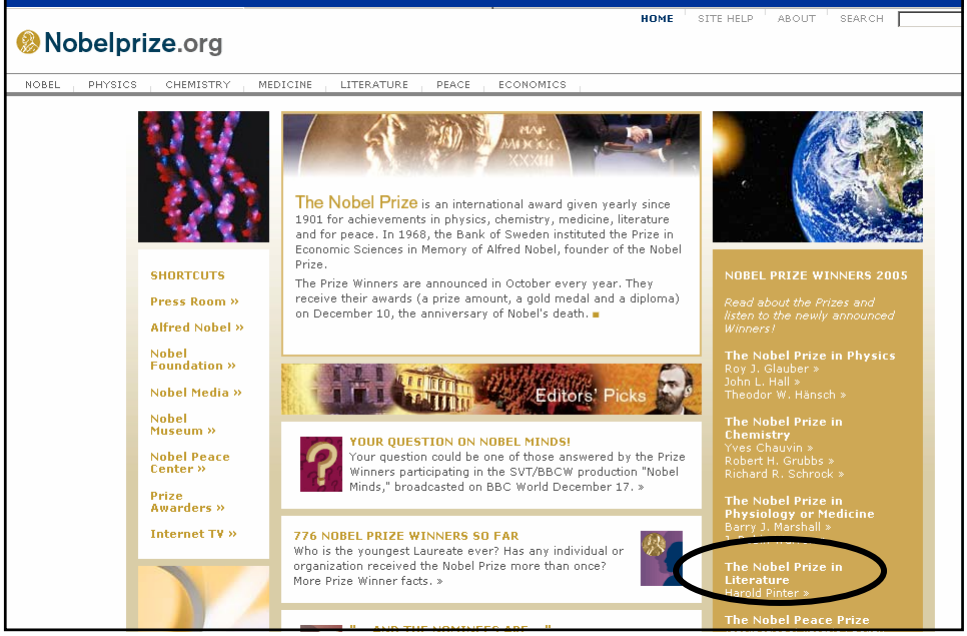
31

Only One Way to View Laureates

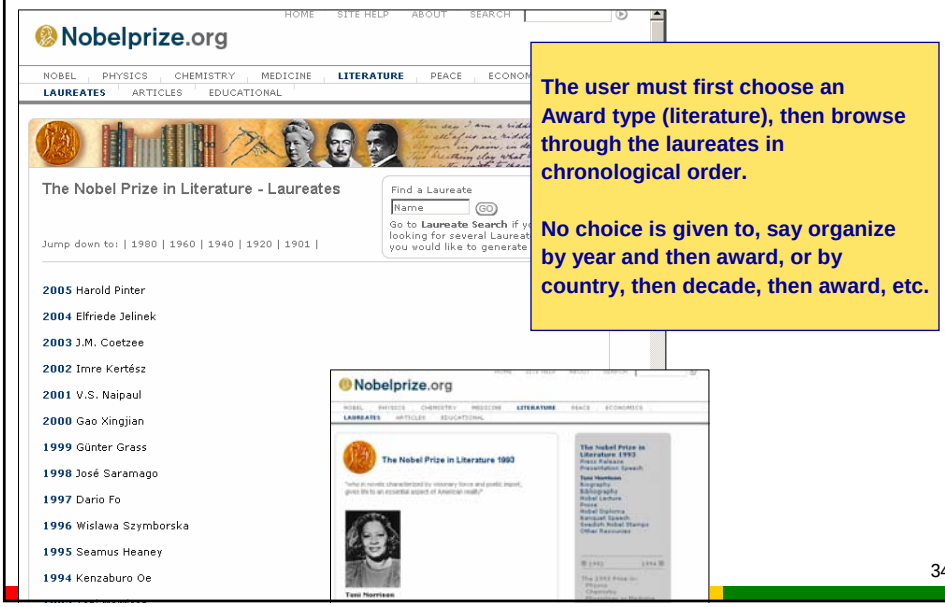


32

First, Choose Prize Type



Next, view the list!



The user must first choose an Award type (literature), then browse through the laureates in chronological order.

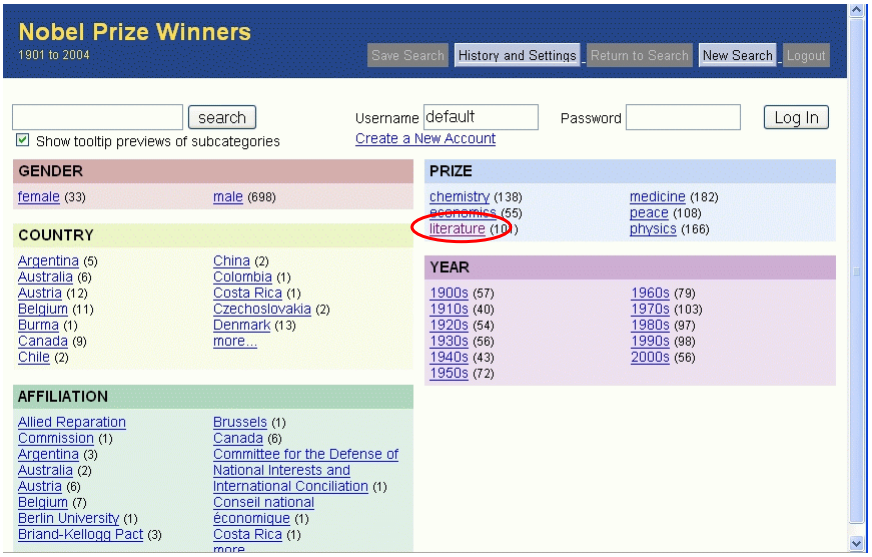
No choice is given to, say organize by year and then award, or by country, then decade, then award, etc.

34

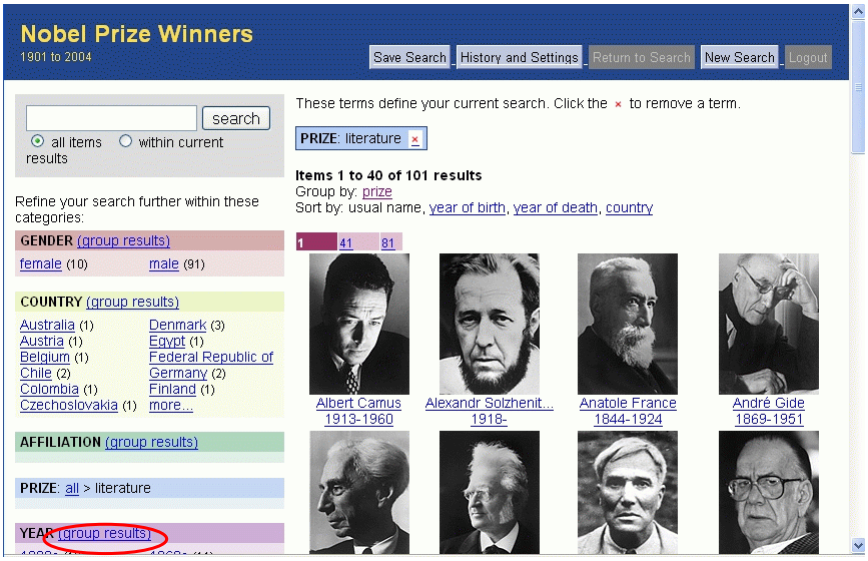
Using Hierarchical Faceted Metadata

35

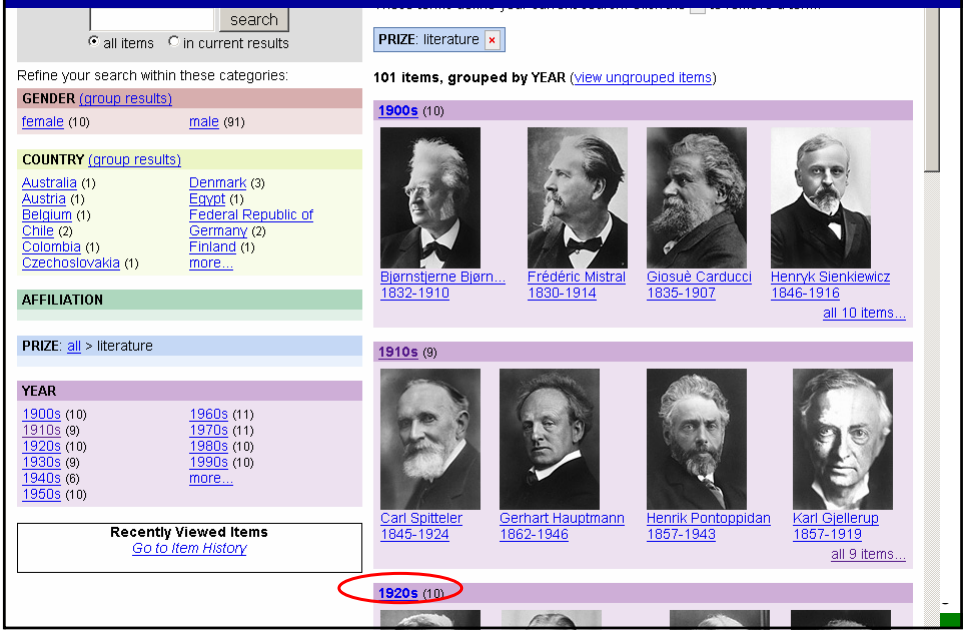
Opening View
Select literature from PRIZE facet



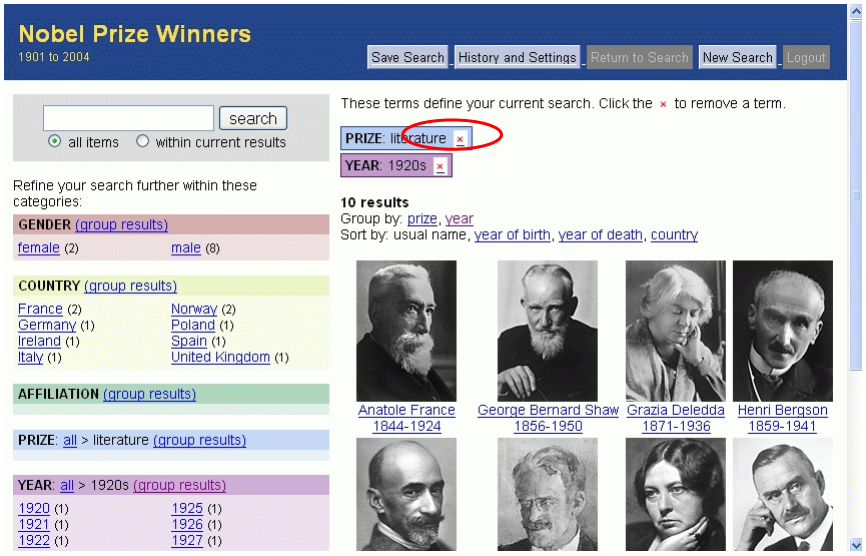
Group results by YEAR facet



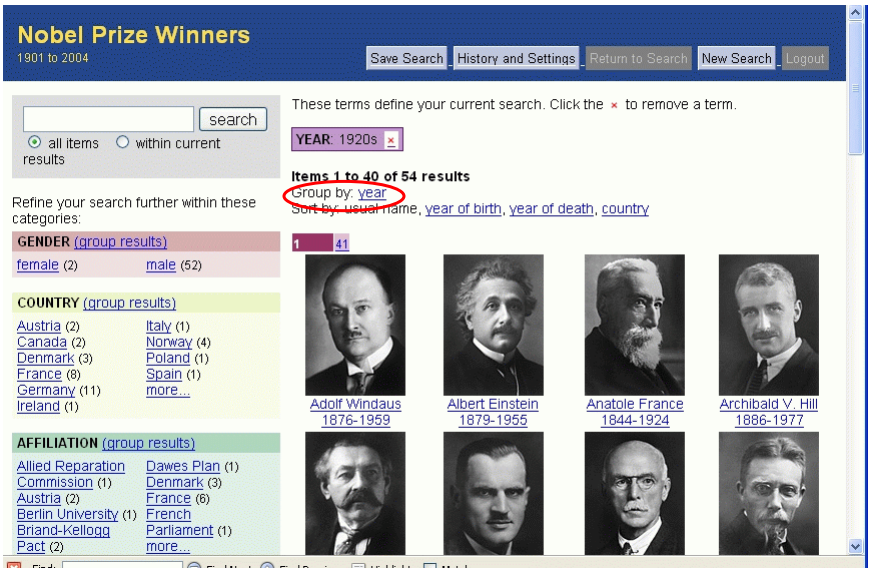
Select 1920's from YEAR facet



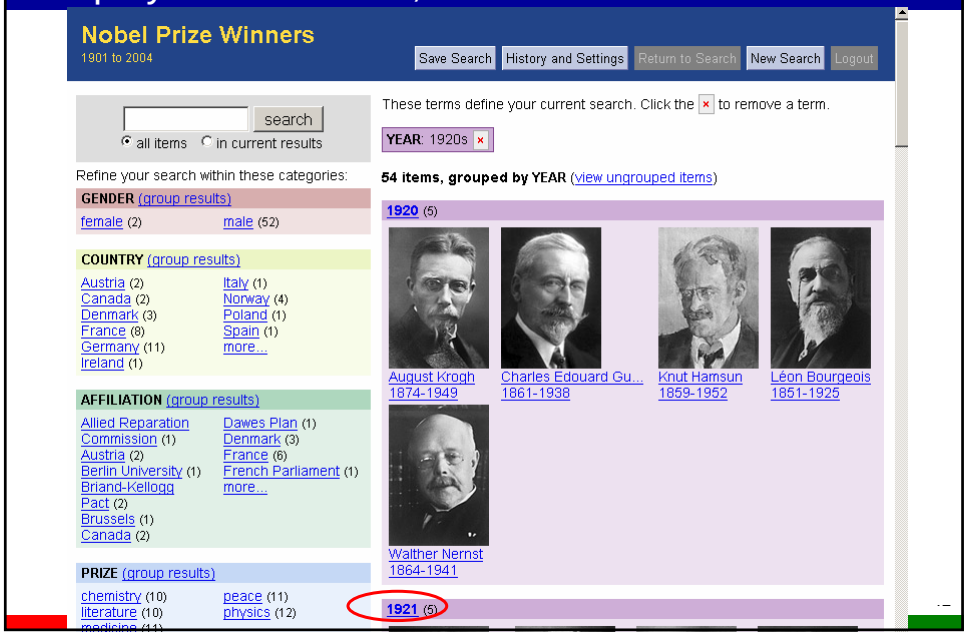
Current query is PRIZE > literature AND YEAR: 1920's. Now remove PRIZE > literature



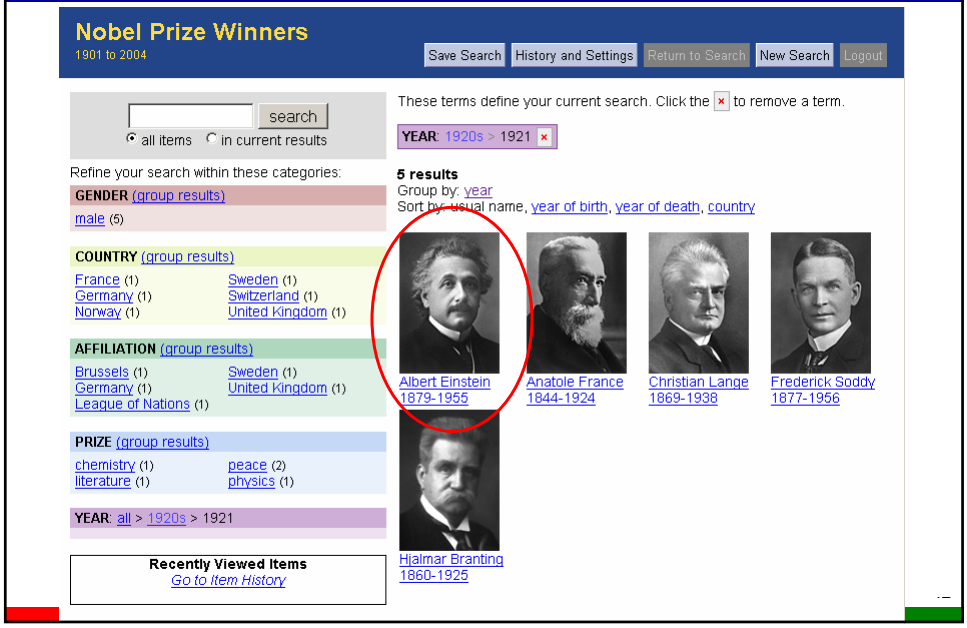
Now Group By YEAR > 1920's



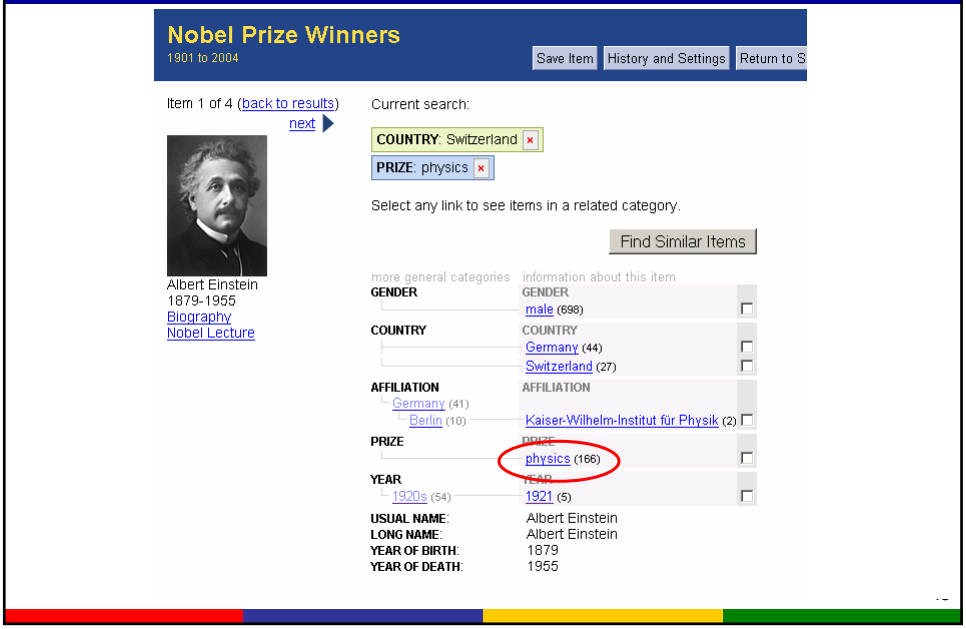
Hierarchy Traversal:
Group By YEAR > 1920's, and drill down to 1921



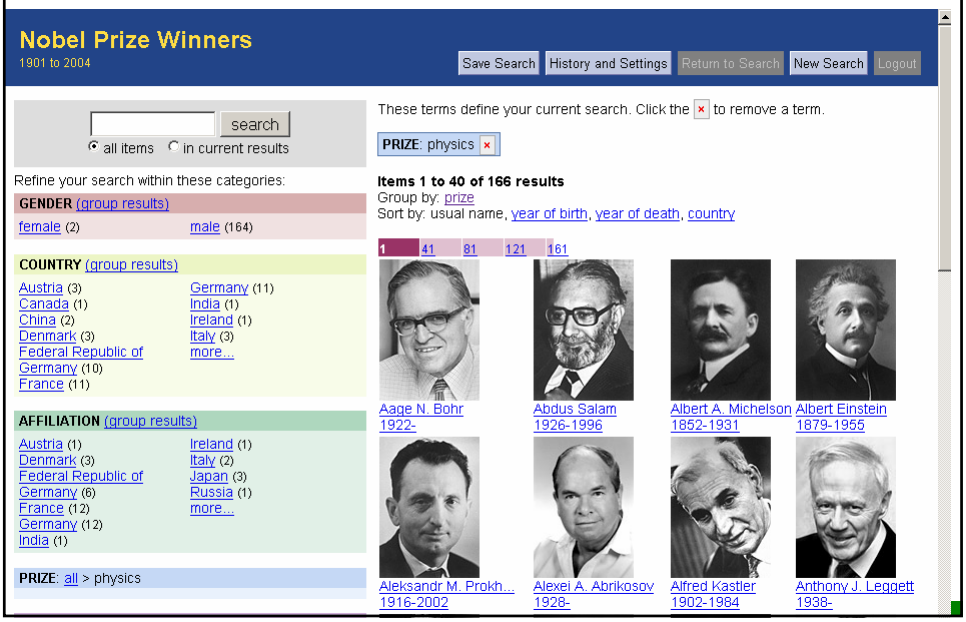
Select an individual item



Use Endgame to expand out

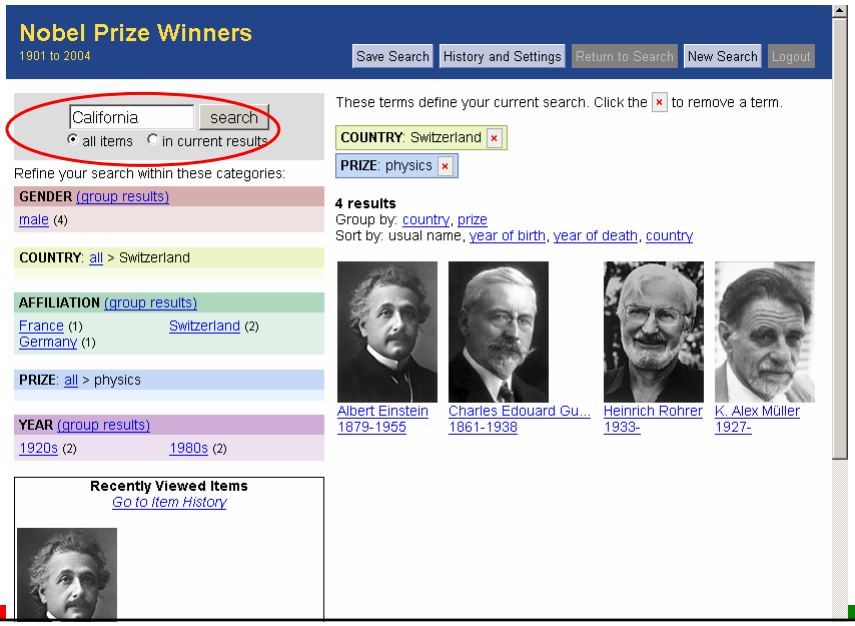


Use Endgame to expand out

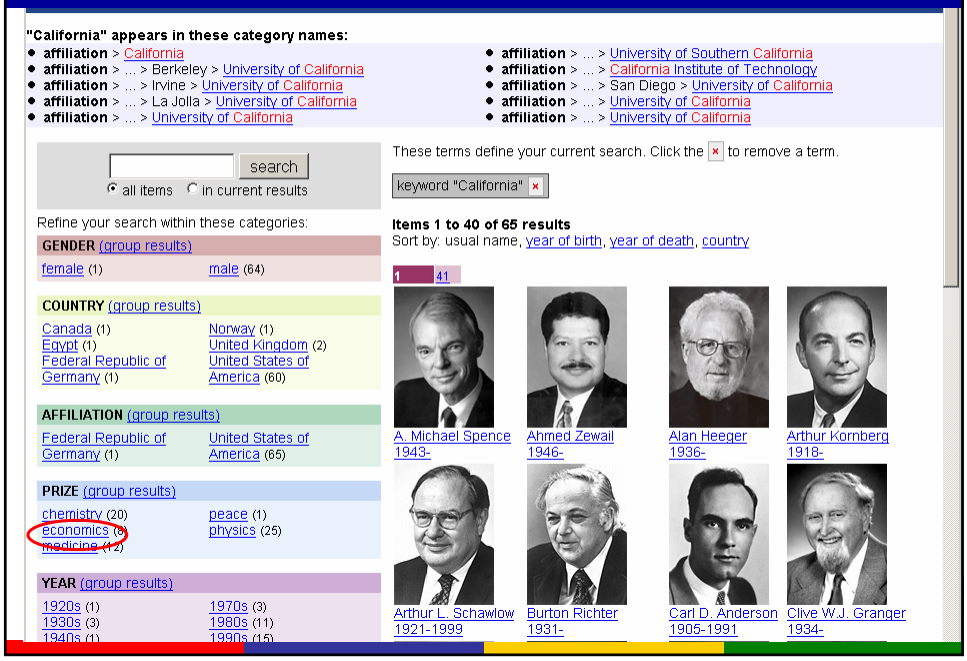


Or use “More like this” to find similar items

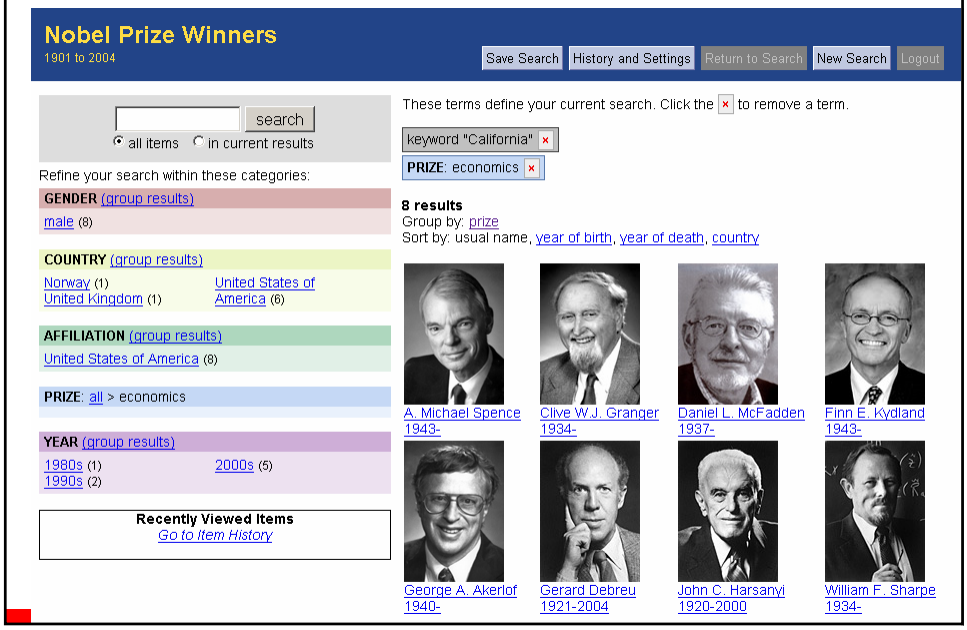
Start a new search using keyword “California”



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

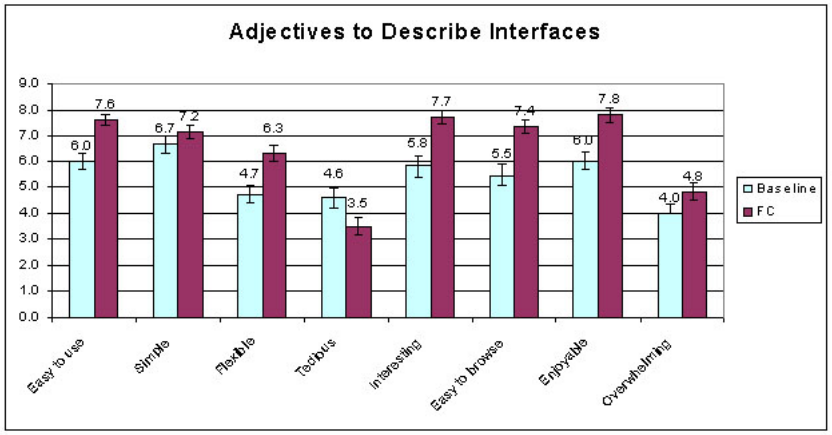
- Floogle (takes the best of the existing keyword-based image search systems)

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

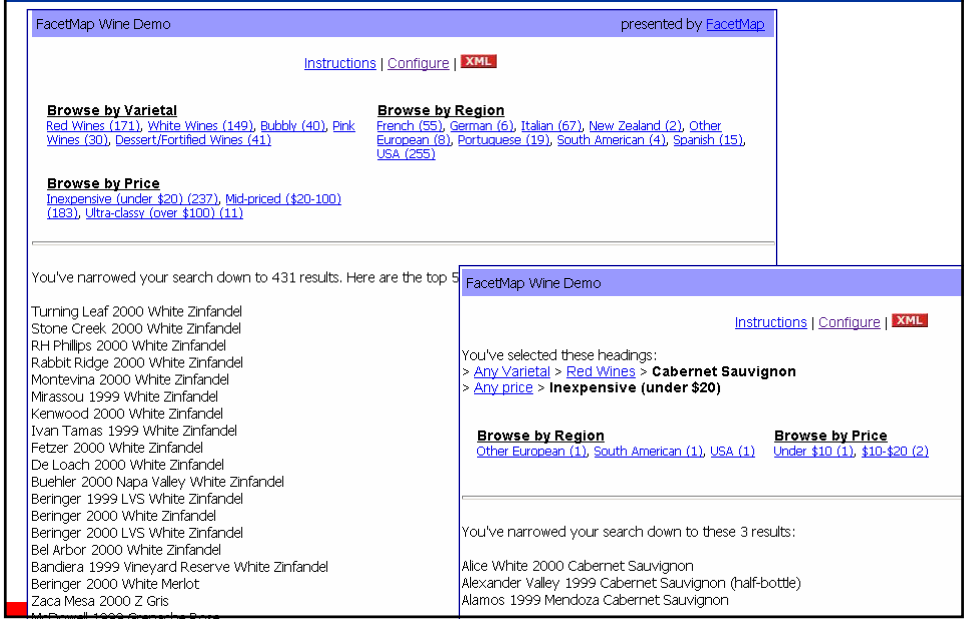
59

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
  - www.rawsugar.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

Looking within: Locations: Italy ✕
Periods: Renaissance ✕
Techniques: decoration ✕

REFINE
Arctic 1
Canada 167
Mexico 571
United States of America 10835

AFTER

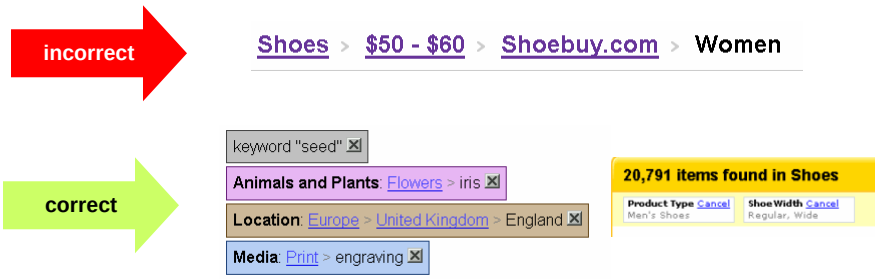
keyword "seed" ✕
Animals and Plants: Flowers > iris ✕
Location: Europe > United Kingdom > England ✕
Media: Print > engraving ✕

Objects (group results)
Clothing (40)
Containers (6)
Food and meals (23)
Lighting (2)
Musical instruments (6)
Vehicles (14)
Weapons (13)
Writing tools (5)

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.



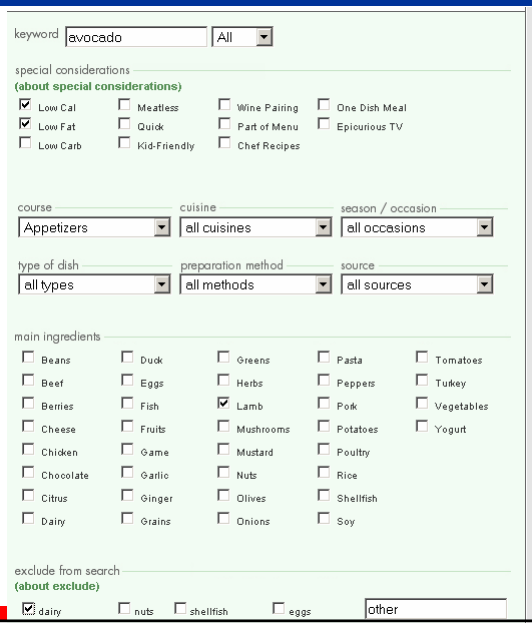
65

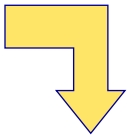
Checkboxes vs. Hyperlinks

- 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

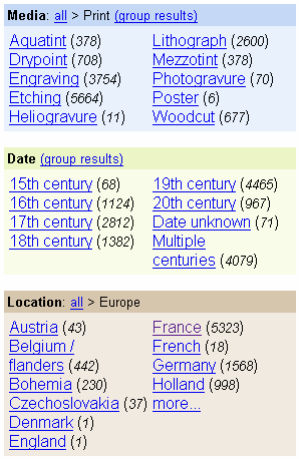
How many facets?

- 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

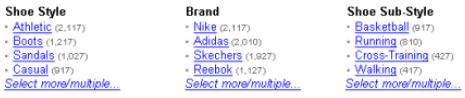
Revealing Hierarchy

- 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”, “fawn”, “doe”

82

Interactive Exercise

- Introduce yourself to 4-6 people near you that you don't already know
- Exchange business cards and note:
  - Region they live (e.g. Canada, Western Europe)
  - Role (e.g. Information Architect, User Researcher)
  - Number of years of experience

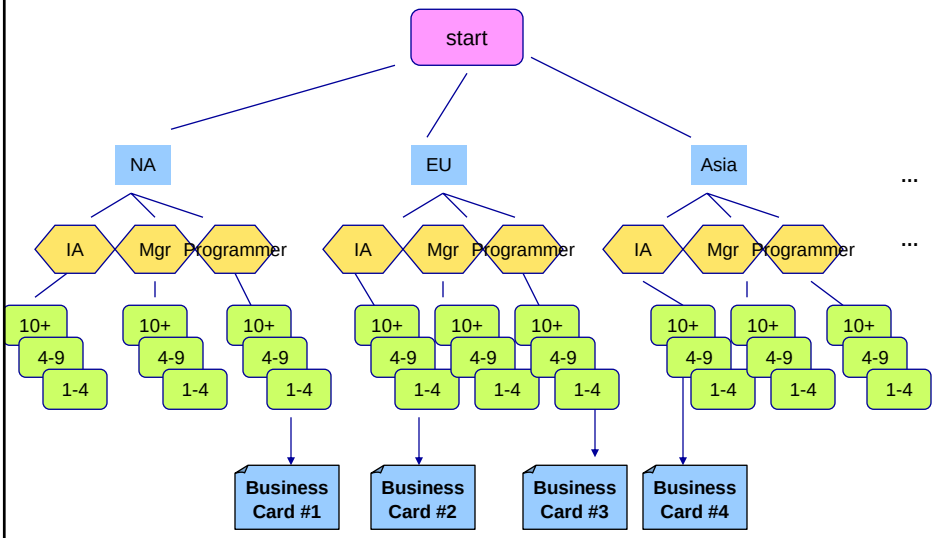
83

Interactive Exercise

- Organize the business cards using a hierarchy assuming you are a talent recruiter in Montreal

84

Interactive Exercise: One way…



85

Evolution of IA at eBay

Flat Structure
(2000 and earlier)

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

86

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

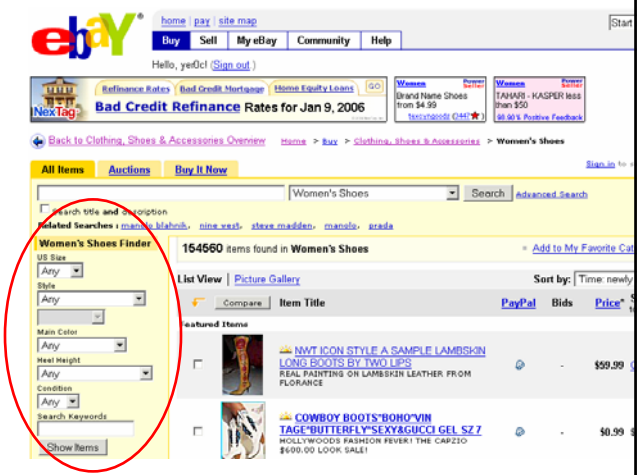
87

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)



88

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.

89

Evolution of IA at eBay

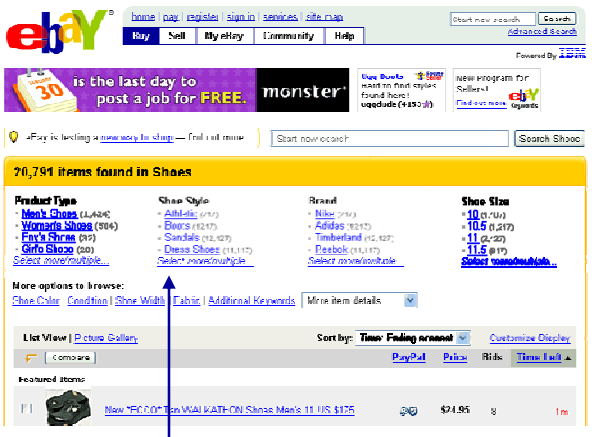
Faceted Metadata

(May 2005 Magellan Test)

Clothing, Shoes & Accessories

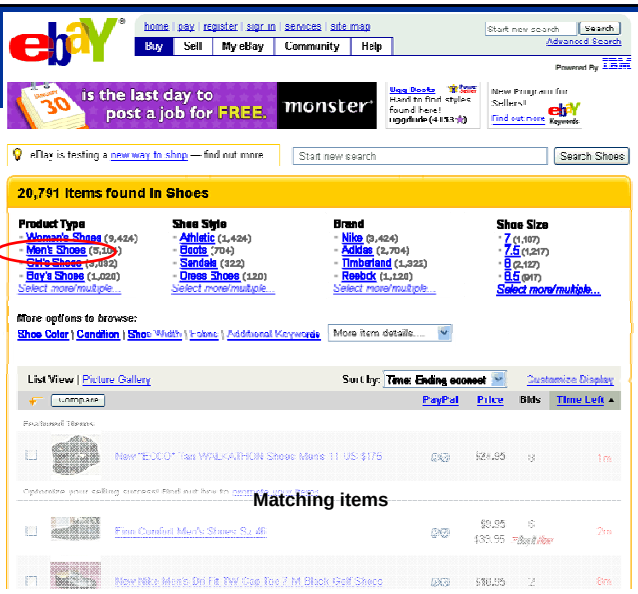
- **Shoes**
  - Women's Shoes
    - Style (Boots, Pumps, Sandals…)
    - Size (6, 6.5, 7, 7.5…)
    - Color (Black, Red, Tan…)
    - Condition (New, Used…)
    - Brand (Nine West, Coach…)
- **Brands**
  - Coach
  - Louis Vuitton
- Materials
  - Cotton
  - Leather

Added Hierarchical Facets

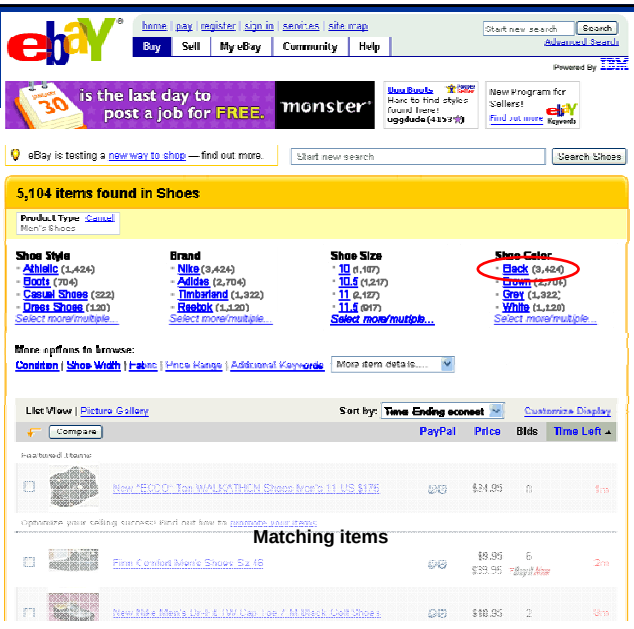


Moved to a Top Positioned Link Structure

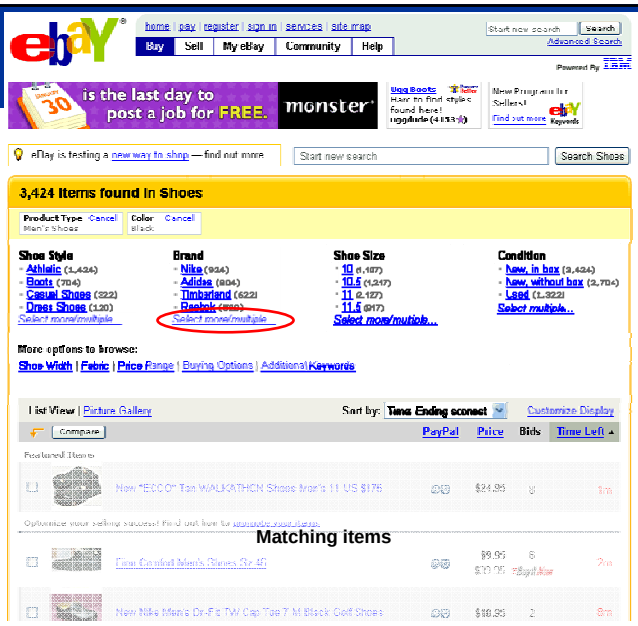
90


20,791 items found in Shoes
Product Type
Shoe Style
Brand
Shoe Size
Matching items

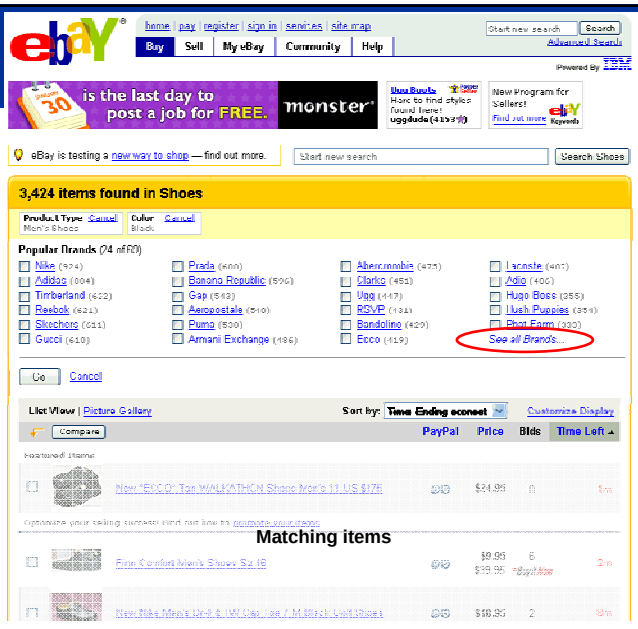
91


5,104 items found in Shoes
Shoe Style
Brand
Shoe Size
Shoe Color
Matching items

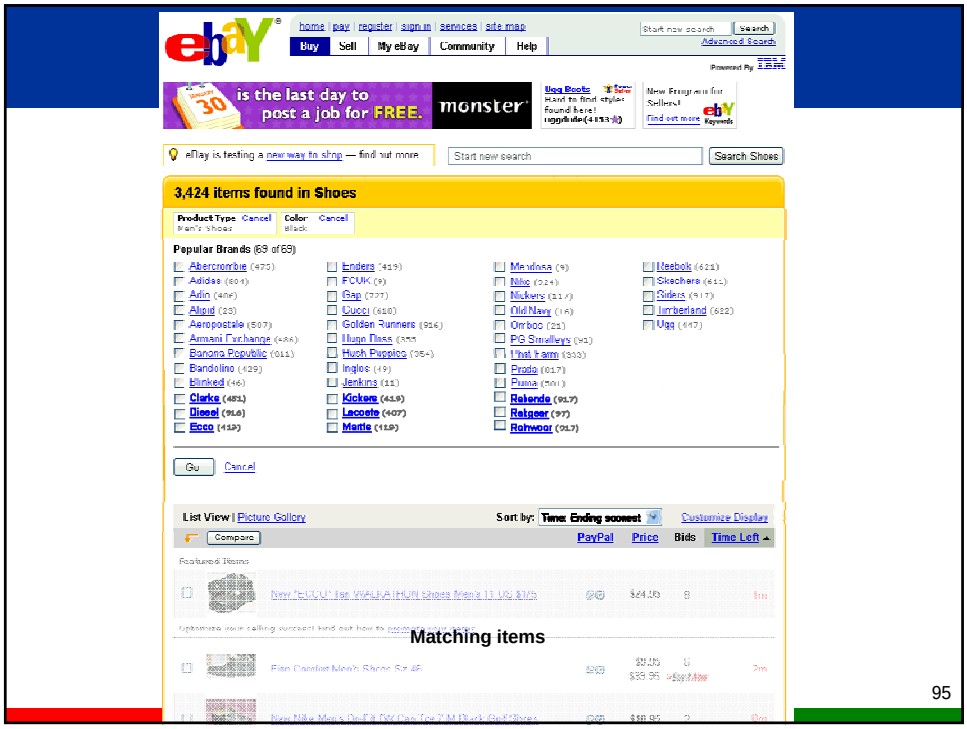
92


3,424 items found in Shoes
Shoe Style
Brand
Shoe Size
Condition
More options to browse:
Matching items

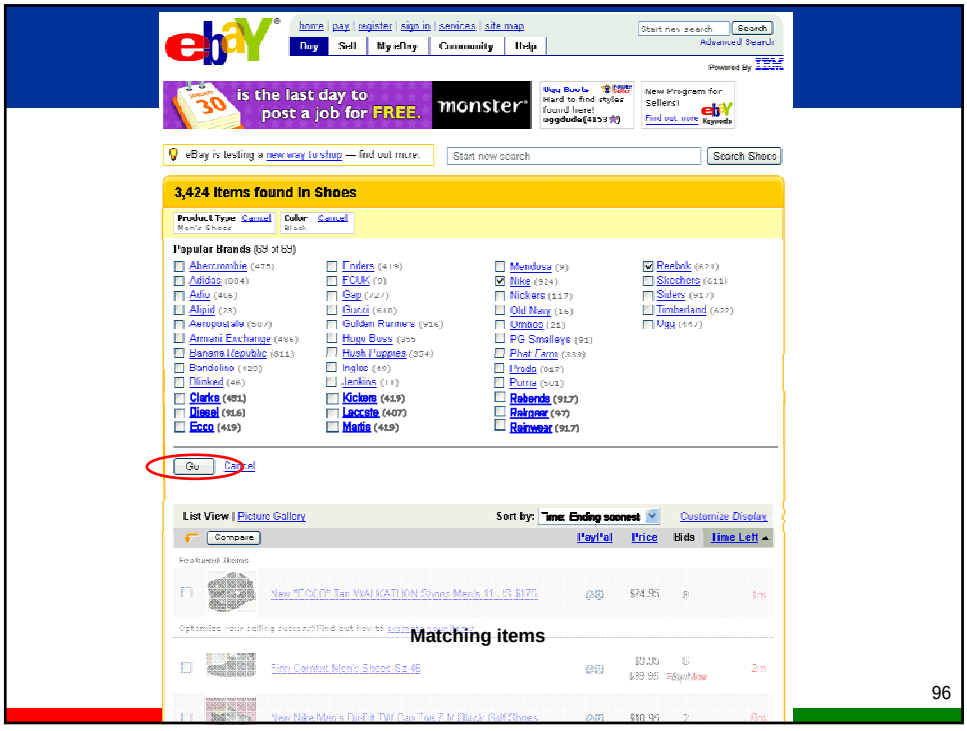
93


3,424 items found in Shoes
Popular Brands
Matching items

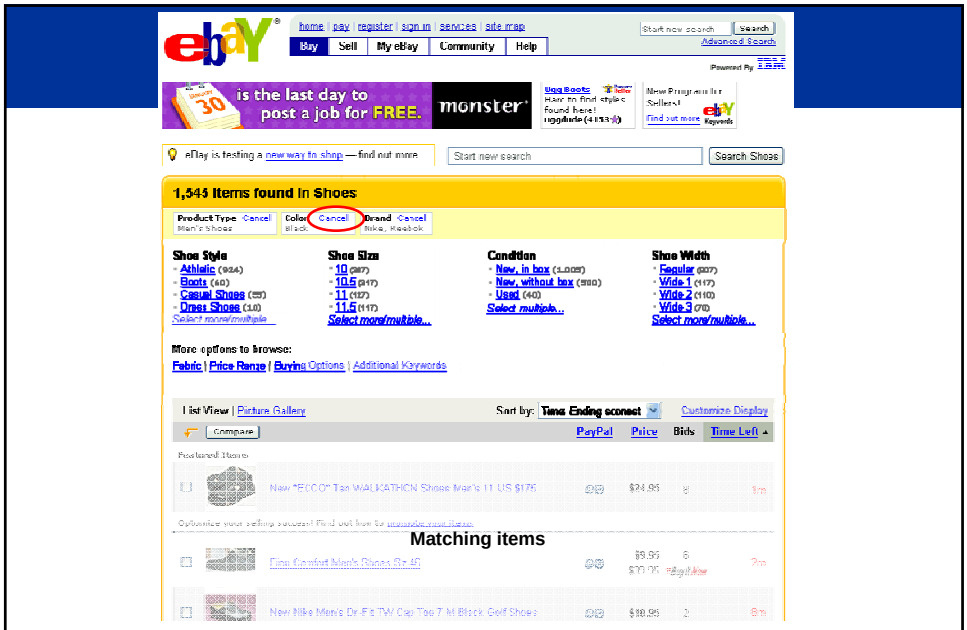
94


3,424 items found in Shoes
Matching items

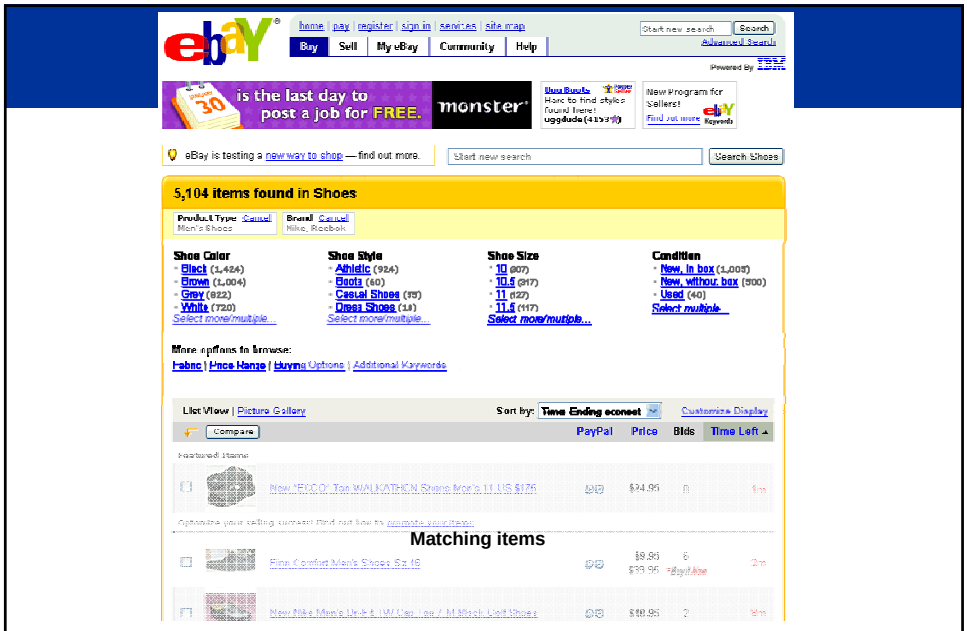
95


3,424 items found in Shoes
Matching items

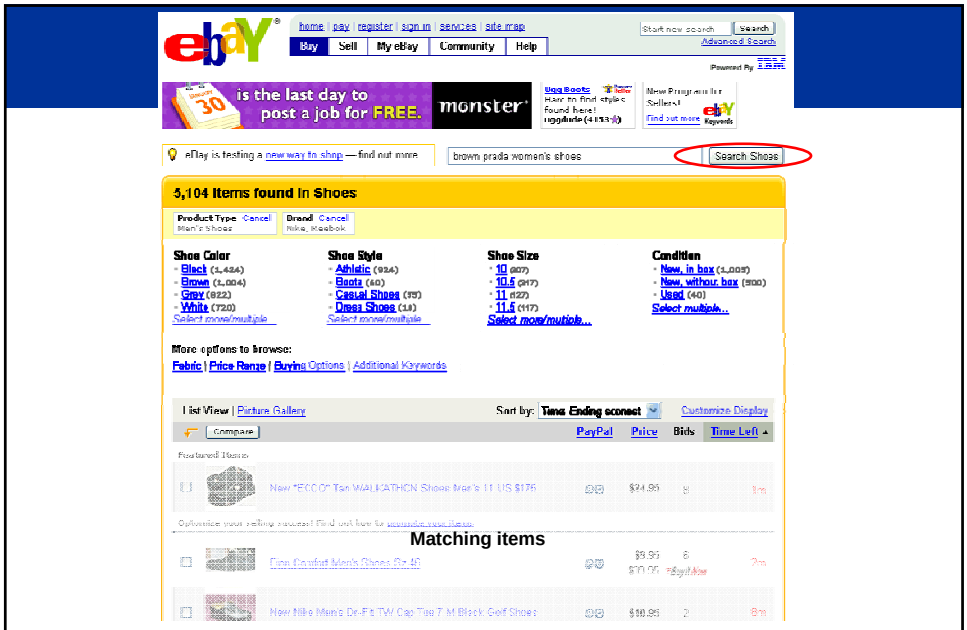
96


1,545 items found in Shoes
Matching items

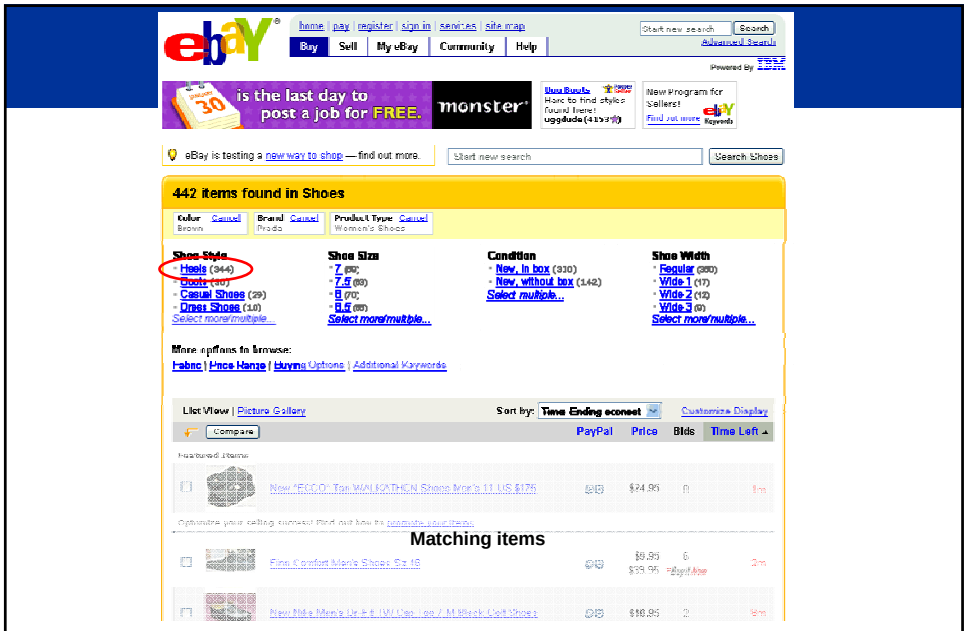
97


5,104 items found in Shoes
Matching items

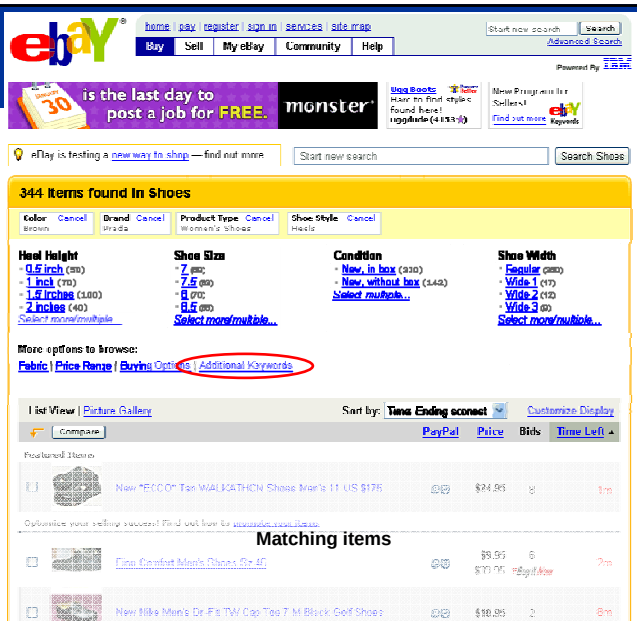
98


3,104 items found in Shoes
Matching items

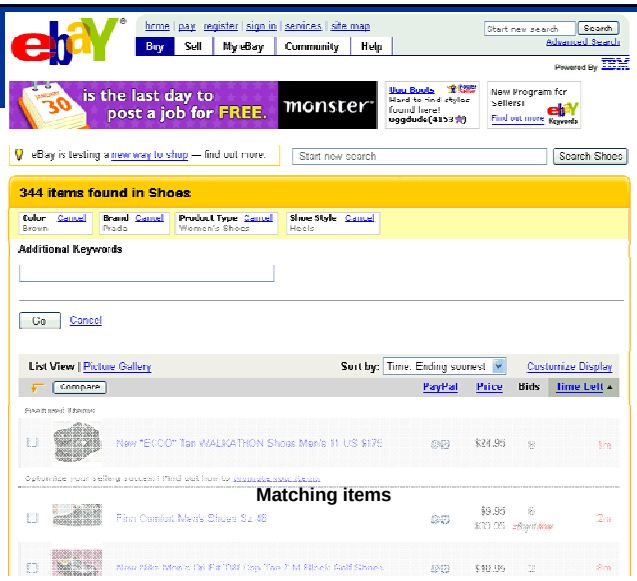
99


442 items found in Shoes
Matching items

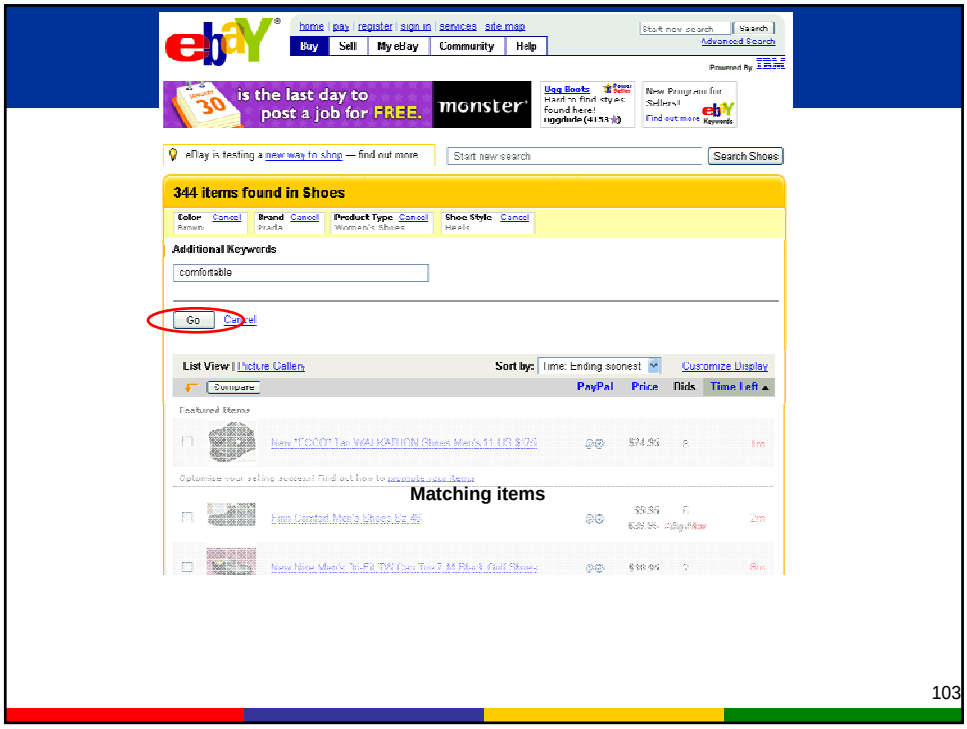
100


344 Items found in Shoes
Matching items

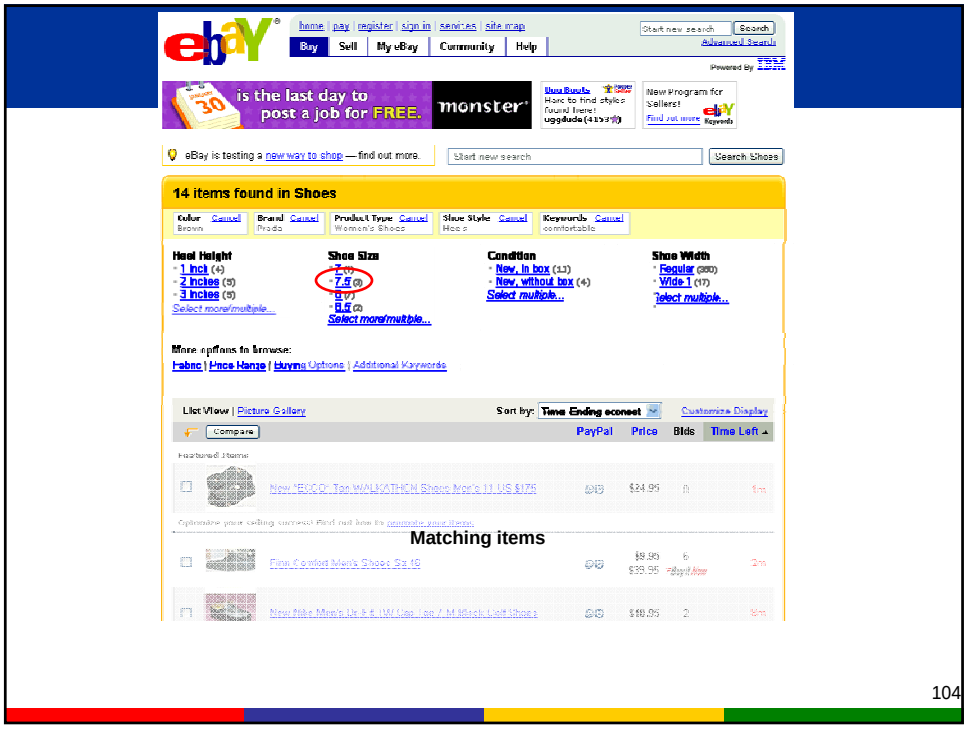
101


344 items found in Shoes
Additional Keywords
Matching items

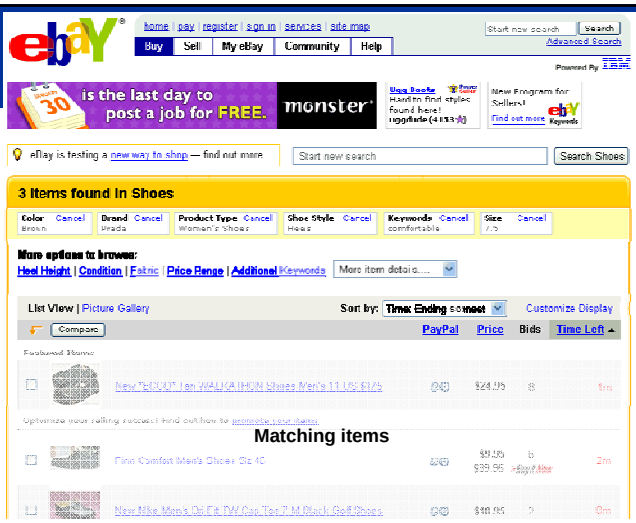
102


344 items found in Shoes
Additional Keywords
comfortable
Go
Cancel
Matching items

103


14 items found in Shoes
Heel Height
Shoe Size
Condition
Shoe Width
Matching items

104



105

Latest eBay design is now live!

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

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

106

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]

107

Lessons Learned at eBay

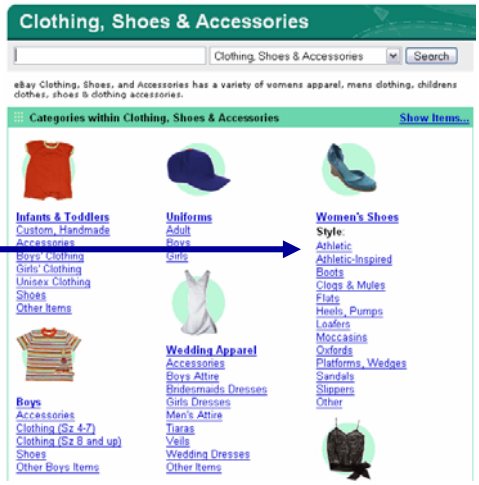
- Data Design
  - Facets
  - Flexibility of Facets vs. Hierarchy
  - Dependencies
- Presentation
  - Integrating “browse” and “search”
  - Control Placement
  - Facet Presentation
  - Breadcrumbs

108

Facets

Lesson: Users desire facets above the domain

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

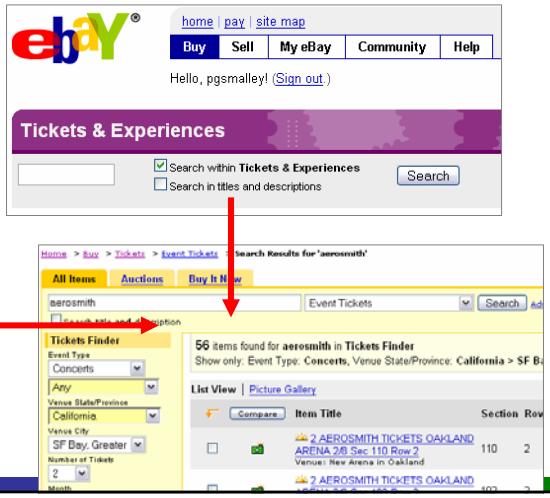


109

Flexibility of Facets vs. Hierarchy

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

Tickets?



110

Dependencies

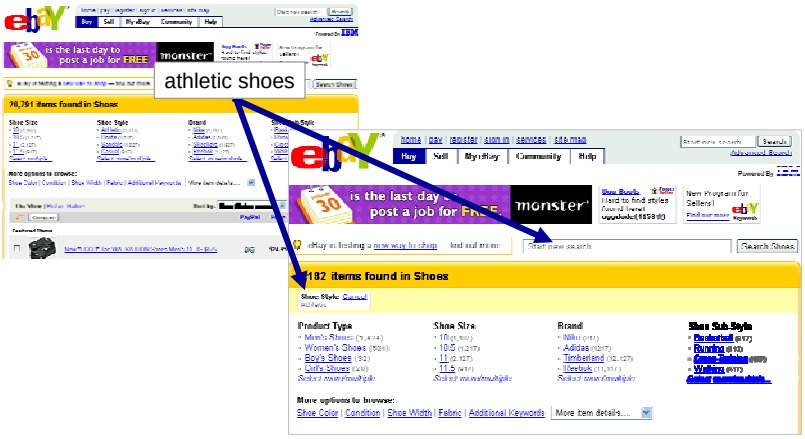
Lesson: Users understand result of removing a parent facet (dependent facets also removed)

Lessons Learned at eBay

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

Integrating “browse” and “search”

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



113

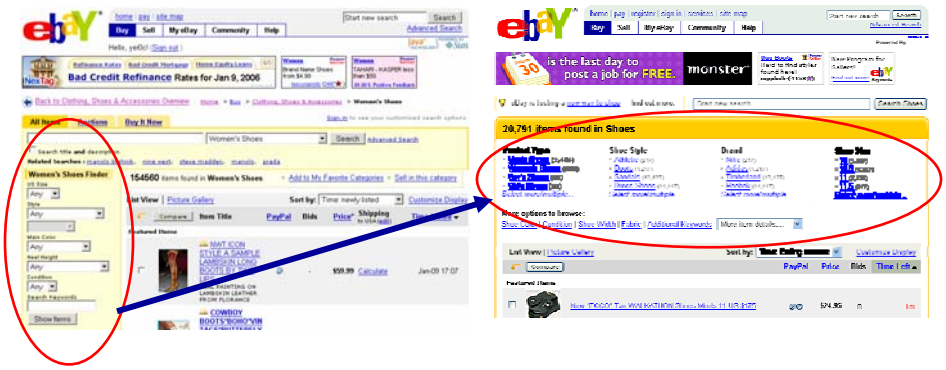
Integrating “browse” and “search”

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

114

Control Placement

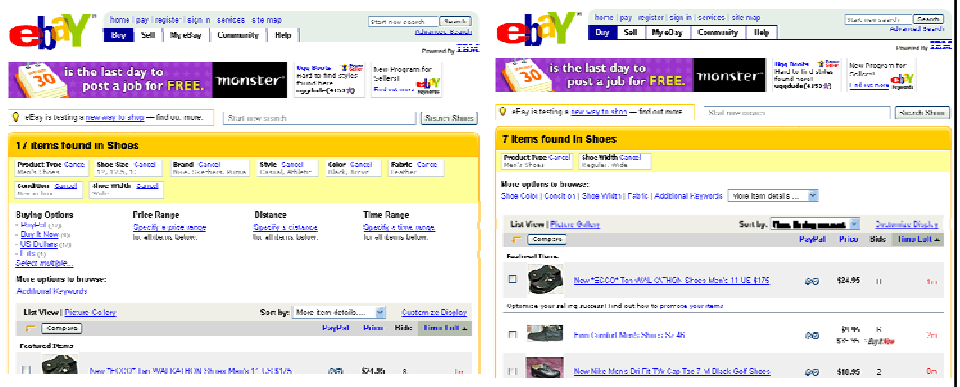
Lesson: Controls placed along the top of the page are used more than when on the left side



115

Facet Presentation

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



116

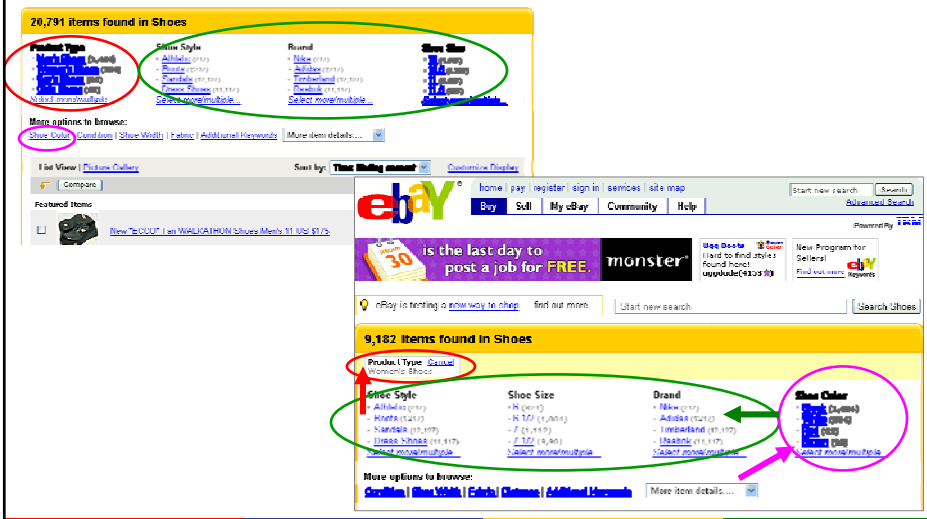
Facet Presentation

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

117

Facet Presentation

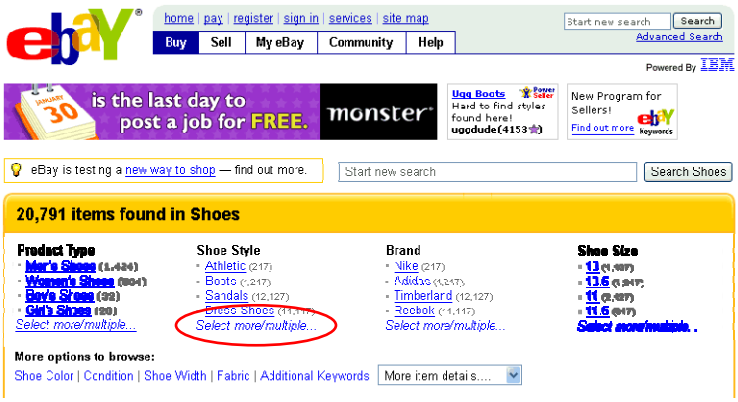
Lesson: Shifting columns doesn't disturb people



118

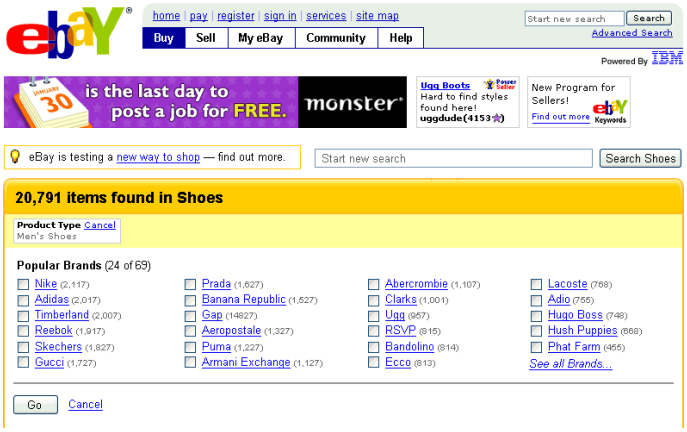
Facet Presentation

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



Facet Presentation

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

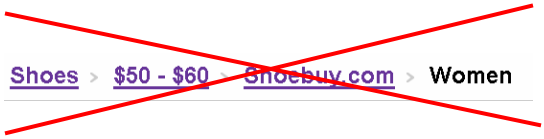


Facet Presentation

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



Breadcrumbs

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

123

Lessons Learned at eBay

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

124

Discussion and Q&A

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

Marti Hearst
School of Information
UC Berkeley

Preston Smalley & Corey Chandler
User Experience & Design
eBay Marketplaces

125

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

126