Exploring highly interconnected humanities data: are faceted browsers always the answer?

Michele Pasin
Information Architect
Nature Publishing Group
michele.pasin@nature.com
Outline

1. **Background.** Interaction models in search interfaces: retrieval model vs explorational model

2. **Use-case.** DJFacet, an app targeting multi-result, highly structured DH datasets

3. **Evaluation & Conclusions.** Are faceted browsers always the answer?
1 - Background
Two models of interaction

Retrieval Model

VS

Exploration model

DYNAMIC TAXONOMIES AND FACETED SEARCH, Theory, Practice and Experience. Giovanni Maria Sacco, Yannis Tzitzikas, Springer-Verlag, 2009 (Chapter 1, ‘The Model’).
Retrieval model: structured search
Retrieval model: guided search

Query

Yahoo! Directory

Astronomy Pictures

Directory > Science > Astronomy > Pictures

Categories
- Astrophotography@
- Comet Hale-Bopp@
- Companies@
- Galaxies@
- Hubble Space Telescope@
- Messier Objects
- NASA@
- Planets@
- Sun@

Result

Astronomy Picture of the Day (APOD)
Discover the cosmos! Each day there is a different image or photograph of our fascinating universe, along with explanation written by a professional astronomer.
apollo.gsfc.nasa.gov/apod/astropix.html

NASA Solar System Simulator
Simulated view of any body in the solar system from any location and time in full-color graphics. A NASA/JPL space.jpl.nasa.gov

NASA's Planetary Photojournal
Images of the planets, the Sun, asteroids, and comets. NASA's Photojournal provides easy access to images from NASA Solar System exploration missions.
photojournal.jpl.nasa.gov

Great Images in NASA (GRIN) [read review]
Collection of the photographs capturing the history of NASA's space program: human exploration, robotic missions, experimental aircrafts, planetary investigations, and deep space studies.
grin.hq.nasa.gov

NASA Image eXchange (NIX)
Collection of NASA aircraft, astronaut, spacecraft, space center, nix.nasa.gov

NASA Images
See images of the Universe, the Solar System, and the Earth at Images. Images is a service of Internet Archive, a nonprofit library, to offer collections.
Exploration model: faceted search

Query

Result
Exploration model: faceted search

EARLY MODERN LONDON THEATRES

Choose more filters in order to narrow down your query further.

Matching *Events* Items (16)

Listing items 1 to 16, page 1 of 1

<table>
<thead>
<tr>
<th>SUMMARY</th>
<th>VENUE(S)</th>
<th>TROUPE(S)</th>
<th>PEOPLE</th>
<th>DATE HAPPENED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dekker places the player John Bentley among a company of playwrights</td>
<td>Bel Savage; Bell; Bull; Theatre</td>
<td>Queen's Men</td>
<td>Atchelow, Thomas; Bentley, John; Dekker, Thomas; Greene, Robert; Kyd, Thomas; Marlowe, Christopher; Nashe, Thomas; Peele, George; Watson, Thomas</td>
<td>undated</td>
</tr>
<tr>
<td>A play about Sophy of Persia is performed at the Red Bull</td>
<td>Blackfriars (II); Curtain; Red Bull</td>
<td>Children of the Queen's Revels (1604-c1614); Queen Anne's Men</td>
<td>Beaumont, Francis</td>
<td>From: c.1604</td>
</tr>
<tr>
<td>The interior of the Curtain is round</td>
<td>Curtain</td>
<td>...</td>
<td>Day, John; Rowley, William; Wilkins, George</td>
<td>From: 1604</td>
</tr>
<tr>
<td>Queen Anne's Men perform 'Sir Thomas</td>
<td>Boar's Head; Curtain; Red</td>
<td>Queen Anne's Men</td>
<td>Dekker, Thomas; Webster, John</td>
<td>From: before</td>
</tr>
</tbody>
</table>

Query
The Retrieval model explained

Facet 1
value 1-1
value 1-2
value 1-3
value 1-4
............

Facet 2
value 2-1
value 2-2
value 2-3
value 2-4
............

Facet 3
value 3-1
............
............

Information Space
The Exploration model explained

Facet 1
value 1-1
value 1-2
value 1-3
value 1-4
............

Facet 2
value 2-1
value 2-2
value 2-3
value 2-4
............

Facet 3
value 3-1
............
............

Information Space

Self Adapting
Exploration Structures
‘Dynamic Taxonomies Search Systems’
Faceted Browsers: key features

- Widely **tested**, many implementations
  - /facet (2006); Exhibit (2007); Humboldt (2008); Collex (2007);

- Implement a ‘**schema-less**’ approach
  - Nowivskie: allow to “explore lateral relationships” and “possibilities for algorithmic serendipity in research”

- **Easy** to use, **user**-centered
  - support both experts and non-experts
  - expose domain features

- Highly **scalable** / **convergent**
  - bottom-up classification / prevent inconclusive searches

- Allow for a ‘**relaxed**’ faceted classification
  - faceted classifications vs dynamic taxonomies search systems
  - available structured data are often enough to bootstrap a FB
2 - New Directions
Extending the model: multiple result types

- Facet 1
  - value 1-1
  - value 1-2
  - value 1-3
  - value 1-4
  - ...........

- Facet 2
  - value 2-1
  - value 2-2
  - value 2-3
  - value 2-4
  - ...........

- Facet 3
  - value 3-1
  - ...........
  - ...........

Information Space

Result-type
(normal ly unique and stable)

E.g.: cars, documents, people
Extending the model: multiple result types

Information Space

Events
- Facets: date; transaction-type; possessions exchanged; place; etc...

People
- Facets: gender; surname; forename; title; etc...

Documents
- Facets: language; date; category; place; etc...

have participants
evidence for
Data Model and facets in POMS
DJango Facet: a Python multi-result FBS

- Python/Django based
- Easy to install / integrate
- Back-end agnostic
- Minimal look and feel
- REST architecture
- Supports pivoting

And more: caching system, custom facets, hierarchical facets etc..

http://www.michelepasin.org/software/
facetslist = [
    {'appearance': {
        'label': 'Person name',
        'uniquename': 'personname',
        'model': Person,
        'dbfield': 'name',
        'displayfield': 'name',
        'grouping': ['personinfo'],
    }},
    'behaviour': [{
        'resulttype': 'persons',
        'querypath': 'name',
    },
    { 'resulttype': 'events',
        'querypath': 'associatedpeople__name',
    },
    { 'resulttype': 'documents',
        'querypath': 'associatedfactoids__associatedpeople__name',
    }],
]
Case studies: EMLOT <www.emlot.kcl.ac.uk>
EMLOT: complex queries made simple

Matching *Transcription Records* Items (2)

<table>
<thead>
<tr>
<th>SECONDARY SOURCE</th>
<th>AUTHOR(S)</th>
<th>PAGE(S) (TRANSCRIPTION RECORD)</th>
<th>PRIMARY SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Professional Theatre, 1530-1660</td>
<td>Berry, Herbert; Ingram, William; Wickham, Glynne William Gladstone</td>
<td>569</td>
<td>LMA, M1/SR/0519</td>
</tr>
</tbody>
</table>
EMLOT: complex queries made simple

Active Filters (4)
- Event & Person
- Event type: Arrest
- Troupe & Venue
- Troupe type: Adult players
- Citation
- Repositories: London Metropolitan Archives, London, UK
- Document: description
- Primary source date: 1613

Matching *People* Items (3)
- SURNAME: Collyns
- FORENAME: Nicholas
- TITLE: Esquire
- SURNAME: Fulsis
- FORENAME: Alexander
- TITLE: 

Matching *Events* Items (1)
- SUMMARY: A purse is snatched at the Red Bull
- VENUE(S): Red Bull
- TROUPE(S): Queen Anne's Men
- PEOPLE: Collyns, Nicholas; Fulsis, Robert

Matching *Transcription Record* Items (2)
- SECONDARY SOURCE: English Professional Theatre, 1530-1660
- AUTHOR(S): Berry, Herbert; Ingram, William; Wickham, Glynne; William Gladstone
- PAGES: 569
EMLOT: complex queries... simple?

Facet: Place of publication: London

Facet: Venue Name: "Phoenix/Cockpit"

Facet: Person Role: "Playwright"

Facet: Troupe type: "Adult players"

Facet: Troupes

Facet: Sources

Facet: Events

Facet: Tr. records

Facet: Venues
Events happened at the Phoenix that involved Adult Players and Playwrights, and were reported in Documents published in London...
Events happened at the **Phoenix** that involved **Adult Players** and **Playwrights**, and were reported in Documents published in **London**...
3 - Evaluation & Conclusions
Evaluation

- **Purpose:**
  - improving the general efficiency of DJFacet
  - testing the intuitiveness of the search and navigation facilities;
  - testing the comprehension of the specific facets we are using
  - testing the comprehension of the ‘multi-result’ approach

- **Setup:**
  - 8 people
  - face to face sessions of 30-60 minutes
  - recorded using screen-casting software
  - the performance is analysed and annotated afterwards

- **Tasks:**
  - incremental difficulty
  - level 0: warming up, exploring the interface (facets and result types)
  - level 1: queries with 1 facet
  - level 2: queries involving 2 facets
  - level 3: queries involving 3 or more facets
## Evaluation results

| Comprehension of the intended meaning of facets | - In general, quite positive  
- Document-class and document-type are very ambiguous  
- Some of the terms within the facets are not easy to interpret: eg the ‘staging context’ event-type. |
|---------------------------|--------------------------------------------------------------------------------------------------|
| Generic UI issues         | - Facets’ role in a search is more intuitive when they are open  
- Clear separation between controls and results  
- Result-type switches are not obvious, people confuse them with “other” facet controls |
| Comprehension of the significance of results | - Pivoting action is not explained properly  
- People with no familiarity with the domain don’t get the implicit relations between result-types  
- People with familiarity with the domain perform quite well |
Evaluation results: future work

- Cues that help users **understand** the **DB model**:
  - **static** section in the help menu
  - **dynamic** ‘query explanation’ mechanism
    - via graphical diagram providing a visual representation of the query
    - via a natural language rendering of the query

- Messages that help users notice the ‘**pivoting**’ action:
  - popups or messages before changing result-type
  - make this control less prominent when filters are already selected
Conclusions

- Are FB useful in the DH?
  - ..definitely YES!
  - a plethora of reasons:
    - easy to use, scalable, transparent, algorithmic serendipity etc..

- But.... careful with:
  - rich domain models (and we want to reveal this richness)
    - fb may quickly become complicated (eg lots of search facets)
  - more than one result-type
    - need to provide ways to tackle ambiguity
  - domain-specific questions:
    - other types of entry points to the datasets will be more effective (eg interactive visualizations)
... thanks!

www.michelepasin.org/software/djfacet

email me at: michele.pasin@nature.com