Facet-based Mapping Interfaces for KOS Vocabularies

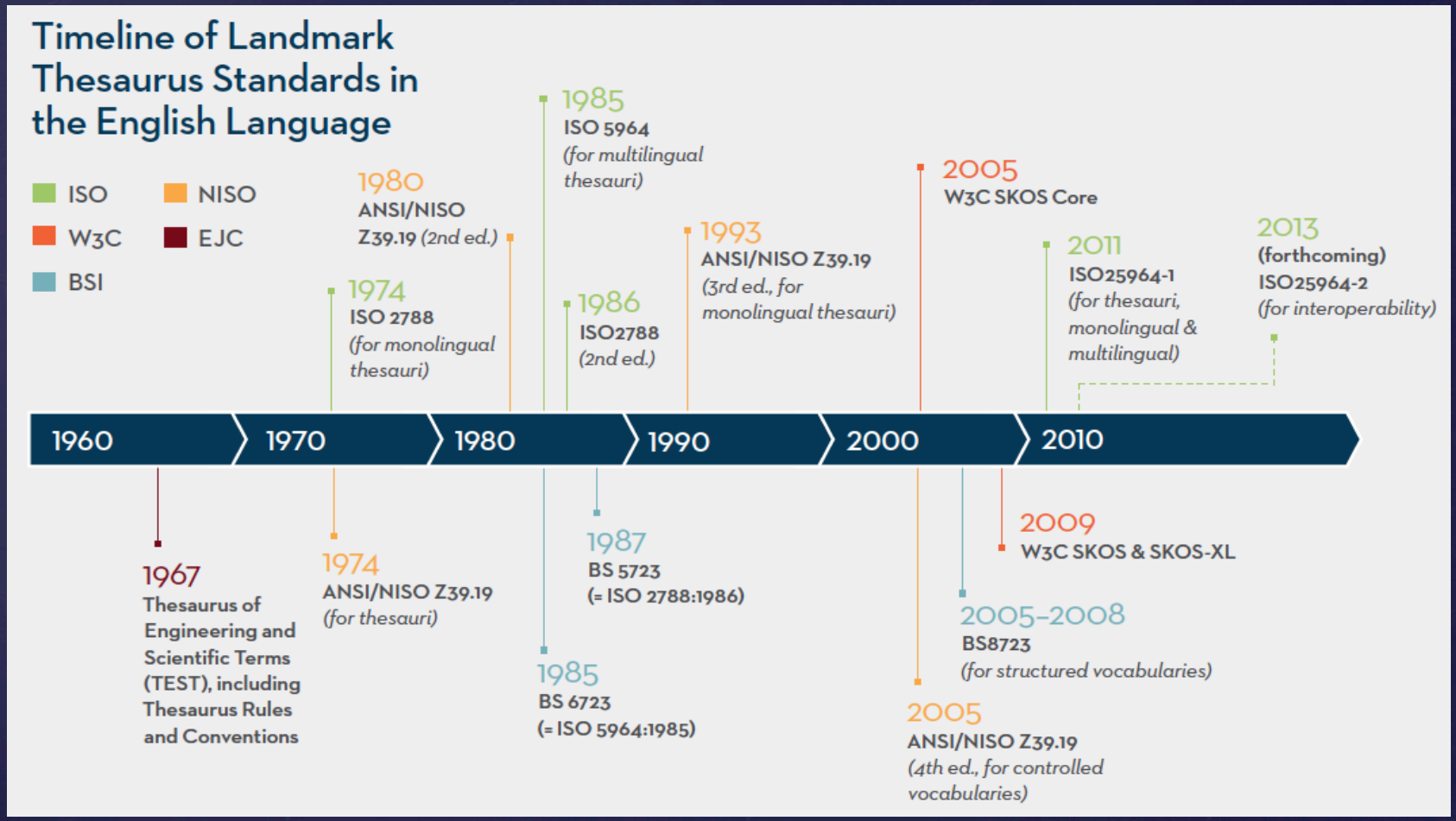
Xia Lin

Drexel University
Philadelphia,
Pennsylvania
USA

KOS Map …

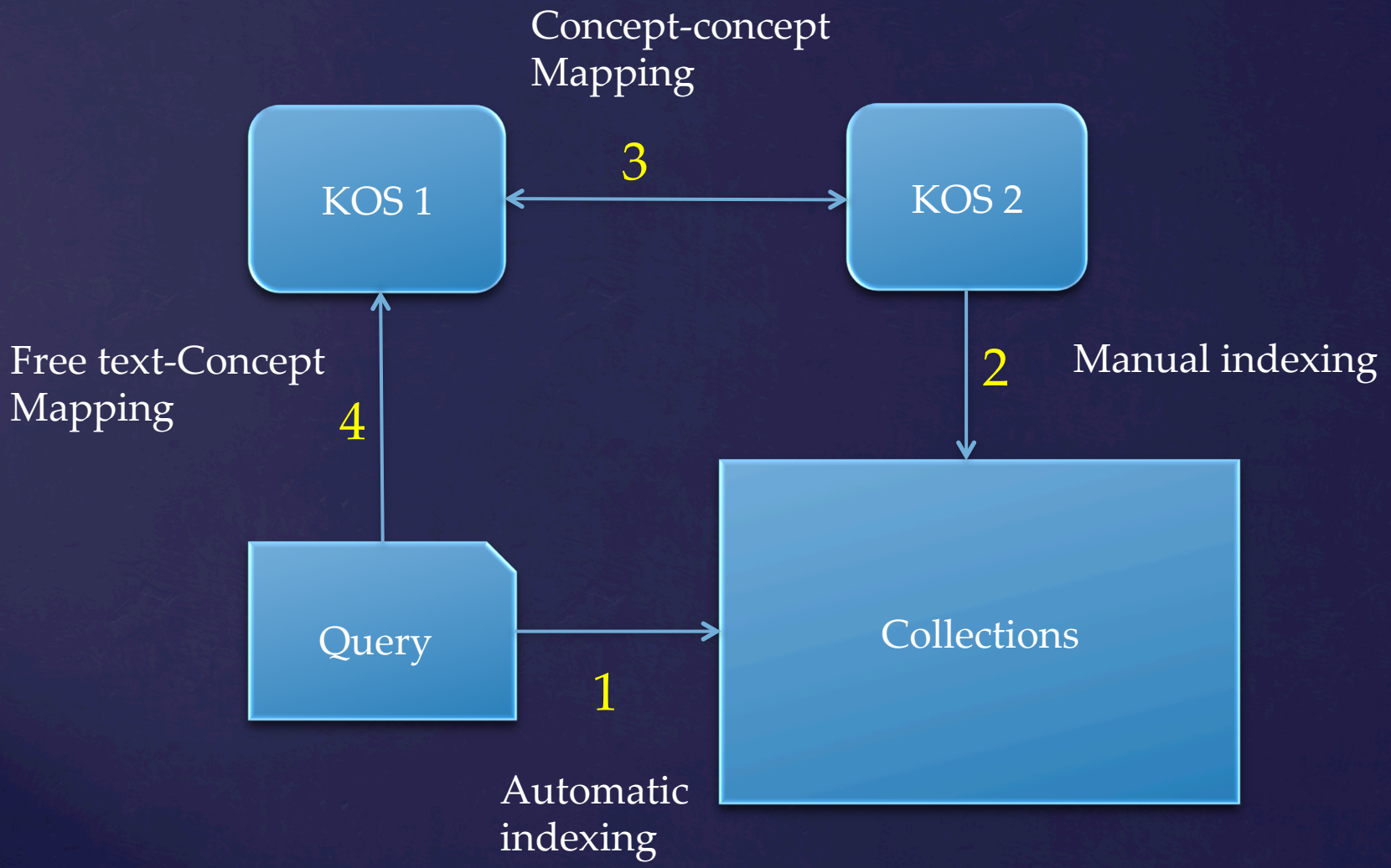
Kos is located in the south-eastern Aegean Sea south of Kalymnos and north of Nisyros at the ancient Keramiko gulf (Kerme Körfezi) or Kos gulf and it abstains just 3 naval miles from Asia Minor coasts.

KOS Mapping



Clarke & Zeng: http://www.niso.org/publications/isq/2012/v24no1/clarke/

KOS Mapping


Concept-concept
Mapping
3
KOS 1
KOS 2
Free text-Concept
Mapping
4
2
Manual indexing
Query
1
Collections
Automatic
indexing

From KOS to MCD

- The MCD Project:
  - Develop Meaningful Concept Displays (MCD) to improve user's searching, browsing, and learning experience
  - Use Knowledge organization Systems (KOS) for query expansion and searching filtering
- Collaboration with
  - Dagobert Soergel, University of Buffalo
  - Bill Ying, ARTstor
  - Murtha Baca, Getty Research

Challenges of KOS Mapping

- Knowledge Store:
  - Create a unified database for multiple KOS
  - Create mapping tables that link concepts and their subcomponents from one KOS to another
  - Build standardized APIs
- Analysis and mapping:
  - Semi-automatic mapping
    - Manually identify string patterns and facets:
      - "oil on canvas" → "coating" on "surface"
    - Use regular expression to match and group matching patterns
  - Apply the analysis to both query analysis and search process.

Mapping Strategy 1

- Using NLP to identify patterns
  - Decompose queries (or terms) to elementary concepts whatever possible
  - "tempera on cardboard 19th century Germany"
    - Pattern: "painting on surface"
      - Painting: Tempera → AAT terms:
        - Tempera, egg tempera, gom tempera, wax tempera, …
      - Surface: cardboard →AAT terms:
        - Cardboard, Corrugated cardboard, Bristol board,
      - Date and time: "19th century"
      - Location: "Germany"
- Challenges:
  - Patterns are currently defined manually
  - Regular expression were used to match and group matching patterns

Mapping Strategy 2

- Using KOS structures to expand concept mapping
  - Map the patterns to KOS terms in facets
    - Identify facets of the elementary concepts
    - Map the elementary concepts to KOS terms
  - Display synonyms, one level broader terms, and all the narrower terms of matching concepts

- Challenges
  - Same terms appearing in different facets may have different meanings
  - Concept-to-concept mappings between two KOSs are not exact string mapping.

Mapping Strategy 3

- Making user's selection a part of mapping
  - The user can browse KOS hierarchies and facets for the matching terms
  - The user can choose from terms from multiple facets to build the queries
  - The user can use use the terms from multiple facets to limit the search and narrow down search results.

- Challenges:
  - Interactive and visualization design is challenging.

Demo

- Demo:
  - Searching a ARTstor collection using AAT concepts to improve precision/recall
    - The collection is not indexed by AAT
    - User can start with a free-text query

- Assumption:
    - The user is familiar with AAT concepts
    - The user is NOT familiar with targeting KOS (ARTStor indexing terms)
    - The User chooses ARTstor terms to find images

DEMO: Search & Explore

Welcome to *MCD* Art Explorer

chalk sticks watercolor paint

Go!

Example search

Your Query: *chalk sticks watercolor paint*

Refine by Facet Mapping

AAT

Facet: Materials Facet

watercolor paint
chalk
chalk, red
chalk, precipitated
chalk, black
chalk, French

Target: ARTSTOR

paper
pencil over traces of red chalk on two pieces of attached paper
chalk, pen, wash on paper
chalk on pper
graphite, charcoal and white

Select All | Deselect All

Search

ULAN

TGN

Search Results

Current query: *charcoal and black chalk|chalk and pastel on paper|chalk, pen and ink on paper|chalk, pen, wash on paper|chalk on pper*
13 results found.

←Previous Page Page 1/1 Next Page→

Sketch for top of the Baldacchino

Creator: Bernini, Gian Lorenzo, 1598-1680
Material: chalk on pper
Subject: Drawing--Italy--17th C. A.D

Holy Family

Creator: Barocci, Federigo, 1528-1612
Material: chalk, pen, wash on paper
Subject: Drawing--Italy--16th C. A.D | Holy Family

Women Fishing

Creator: Beckmann, Max, 1884-1950|Max Beckmann|Beckmann, Max|Beckman, Max
Material: charcoal and black chalk
Subject: Anthropomorphism | Drawing--Germany--20th C. A.D | Nude | corsets | fishing | fishing poles

Sketches for Stairway, Profile of Column Bases, Laurentian Library

Creator: Michelangelo Buonarroti, 1475-1564

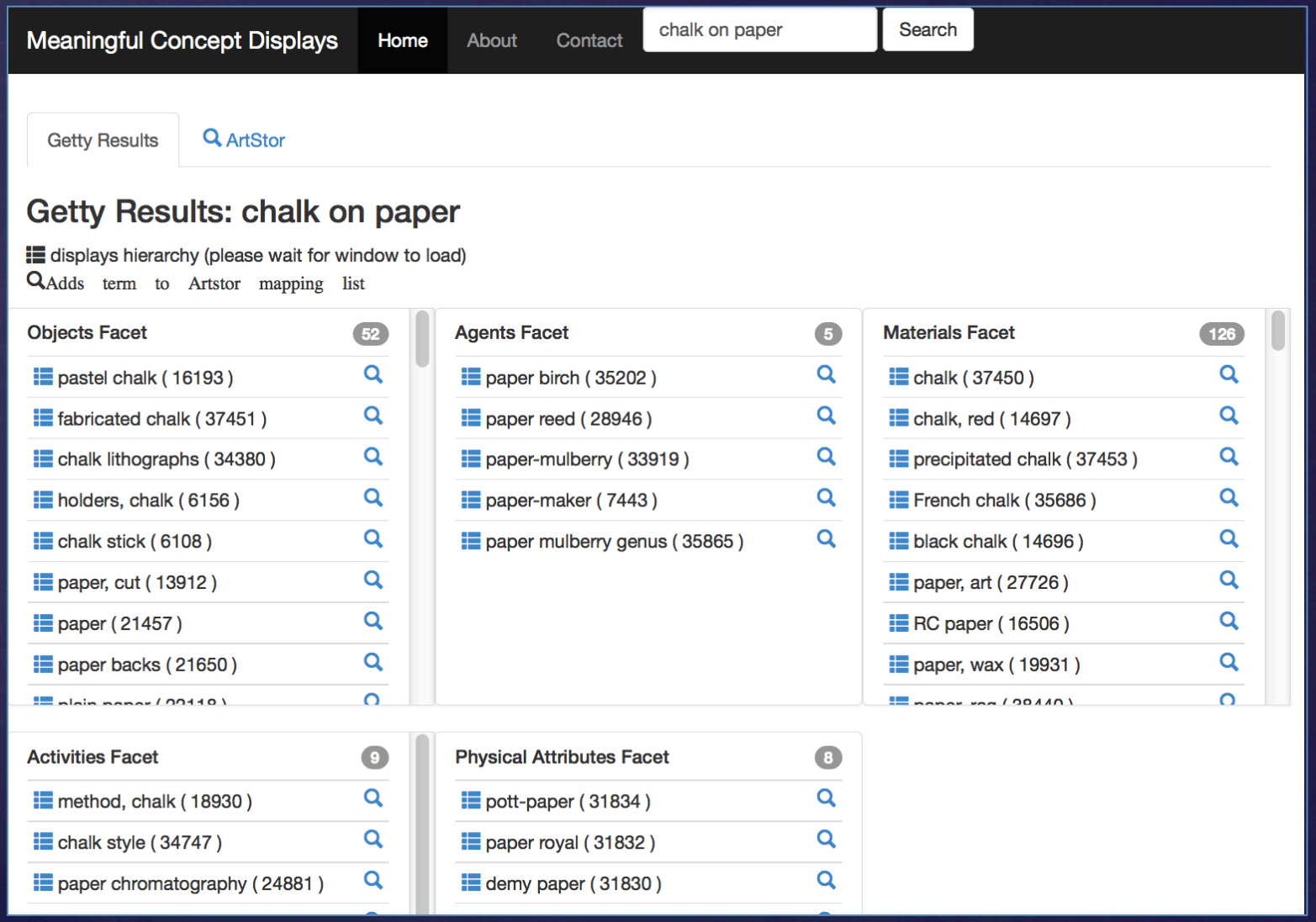
Sketches for the top of the Baldacchino

Creator: Bernini, Gian Lorenzo, 1598-1680
Material: chalk, pen and ink on paper
Subject: Drawing--Italy--17th C. A.D

Kneeling Figure of St. Luke

Creator: Goes, Hugo van der, 1440-1482
Material: black chalk, pen and ink on paper

MCD Librarian's Interface

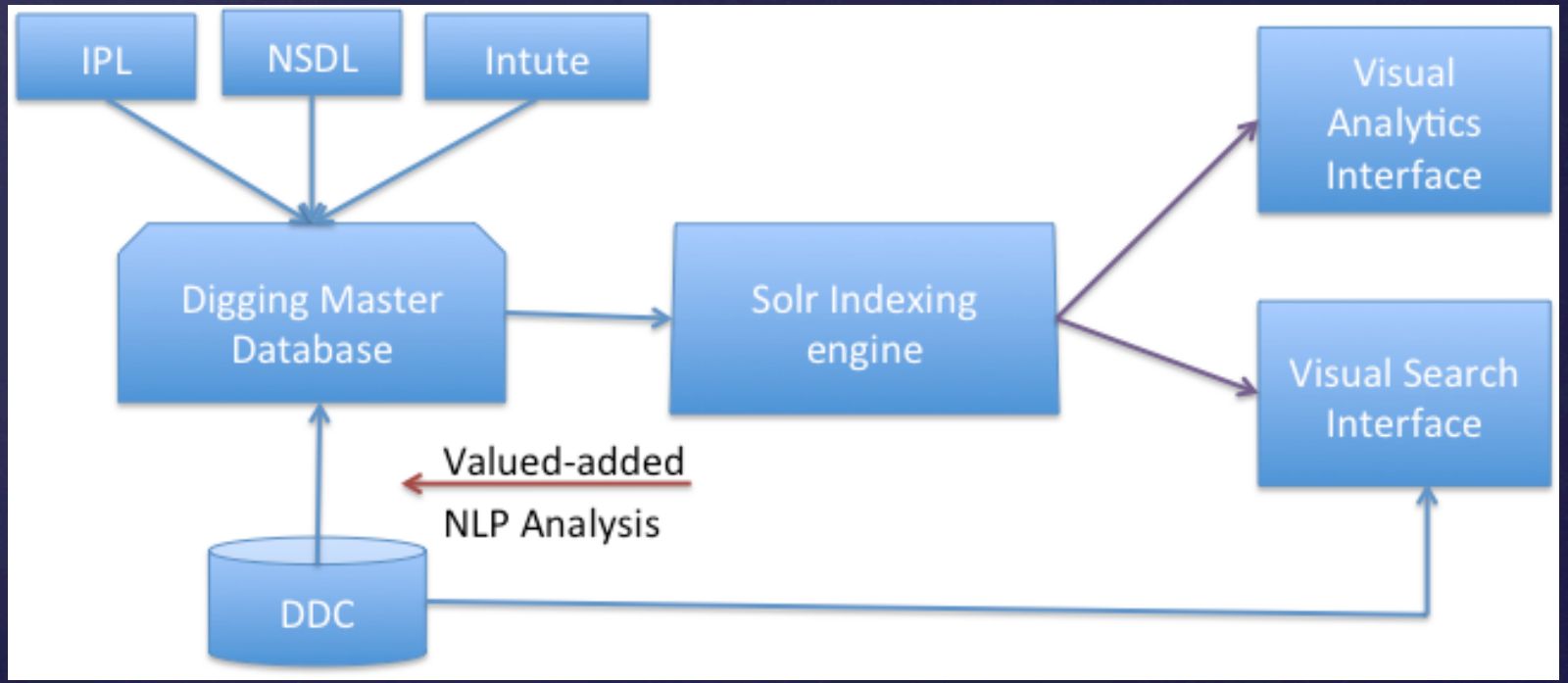


Visual Exploratory Interface

- Support visual interactions for metadata exploration
  - Using DDC classification structures
  - Using the semantic network structures of
    - document-document associative relationships
    - term-document associative relationships
    - DDC class-document associative relationships
  - Using various visualization techniques
- The user must be in control.

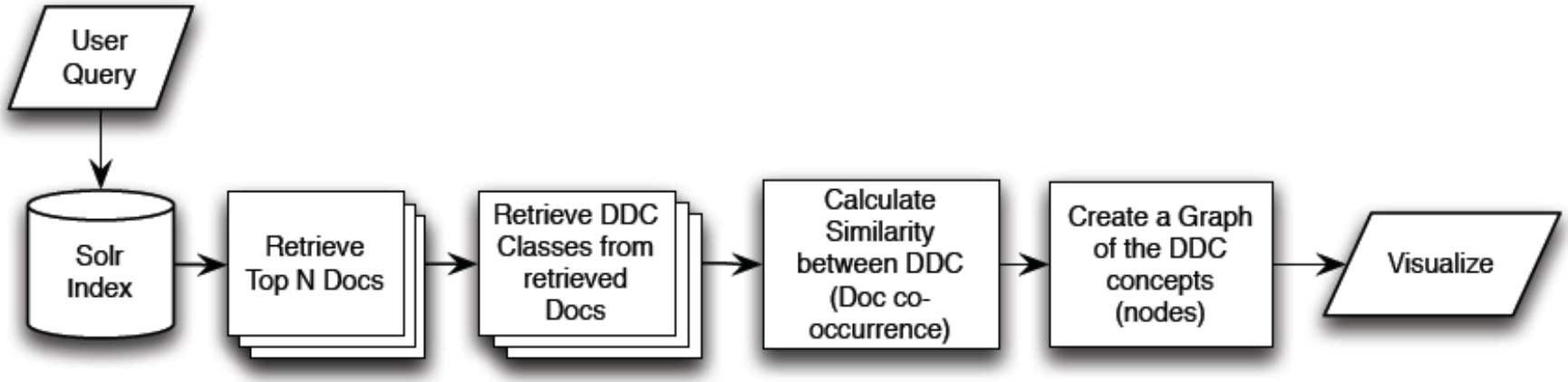
Digging into Metadata Project

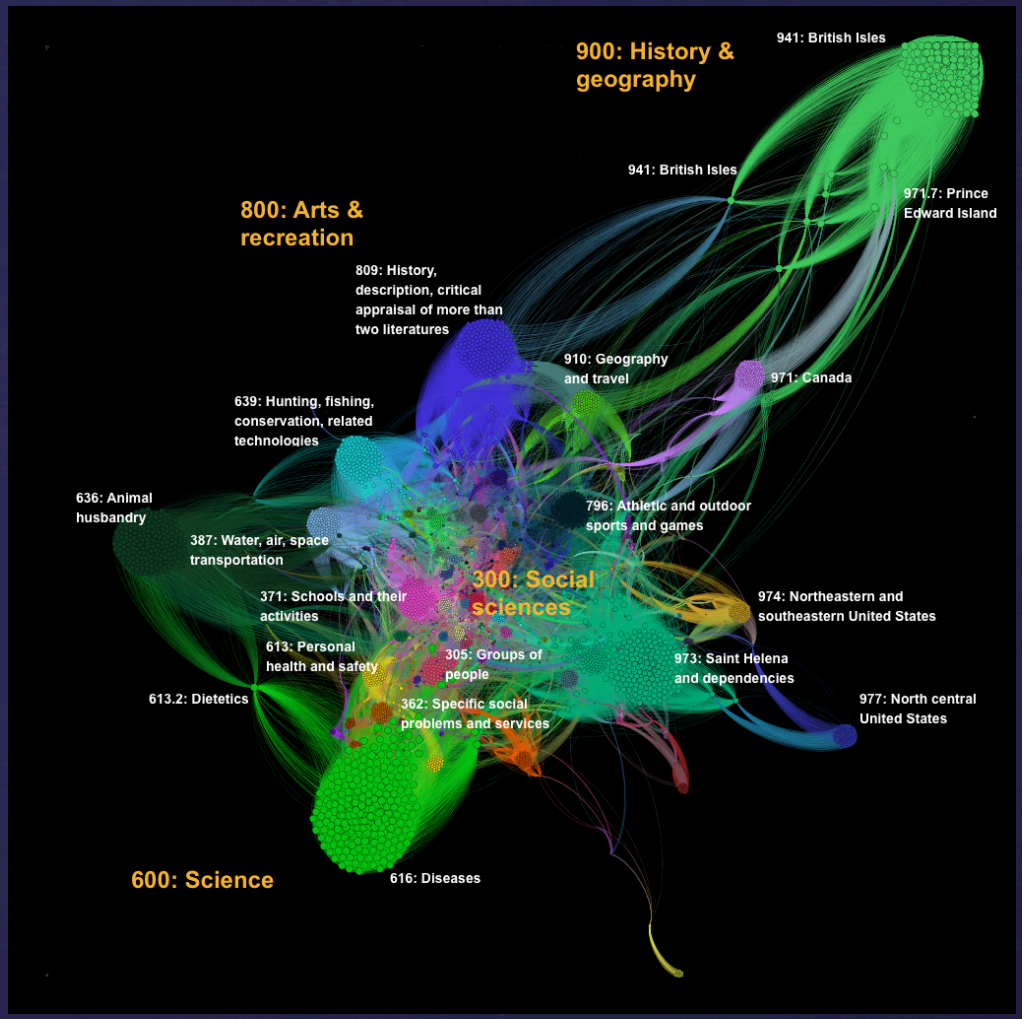
- Goal: Enhance and unify metadata of three digital libraries through Dewey Decimal Classification (DDC)



Digging into Metadata

- Analysis:
  - Porter stemming and stop word removal
  - Calculate term weights and term frequency (TF)
  - Apply Threshold : > Mean(TF) + 1SD(TF)
  - Identify and process noun phrases (Termine, Sum(TF))
  - Connect DDC nodes based on similarities of documents assigned to the DDC classes.

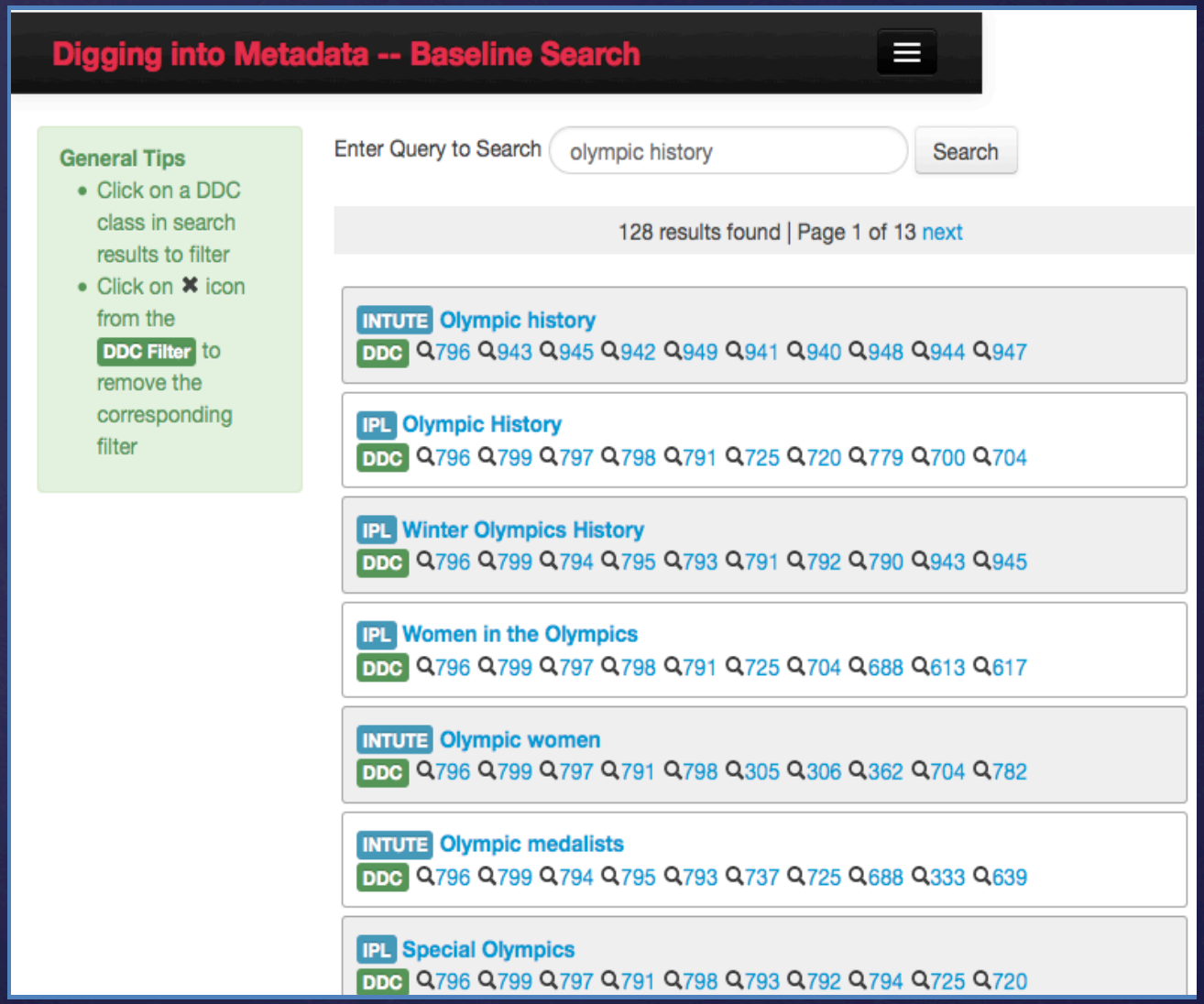



900: History & geography
941: British Isles
941: British Isles
971.7: Prince Edward Island
800: Arts & recreation
809: History, description, critical appraisal of more than two literatures
910: Geography and travel
971: Canada
639: Hunting, fishing, conservation, related technologies
636: Animal husbandry
796: Athletic and outdoor sports and games
387: Water, air, space transportation
300: Social sciences
371: Schools and their activities
974: Northeastern and southeastern United States
613: Personal health and safety
305: Groups of people
973: Saint Helena and dependencies
613.2: Dietetics
362: Specific social problems and services
977: North central United States
600: Science
616: Diseases

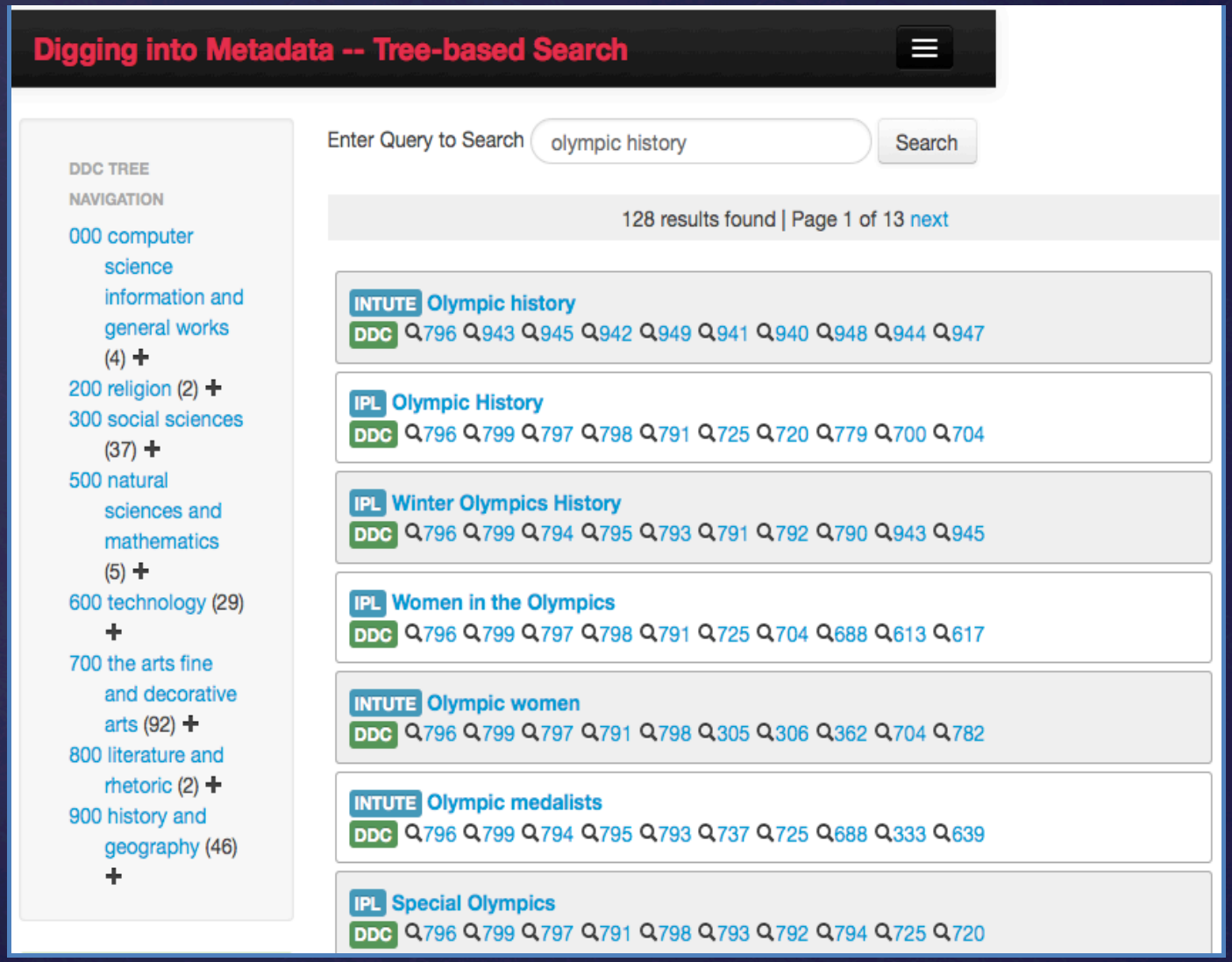
The User Study

- 30 subjects completed three tasks on the three experimental interfaces
  - Subjects completed a pre-questionnaire and a post-questionnaire
  - Watched a video on how the three interfaces worked
  - Completed one task with each interface – find best answers for the given questions.
    - The order of the interfaces and tasks were rotationally assigned to the subjects.

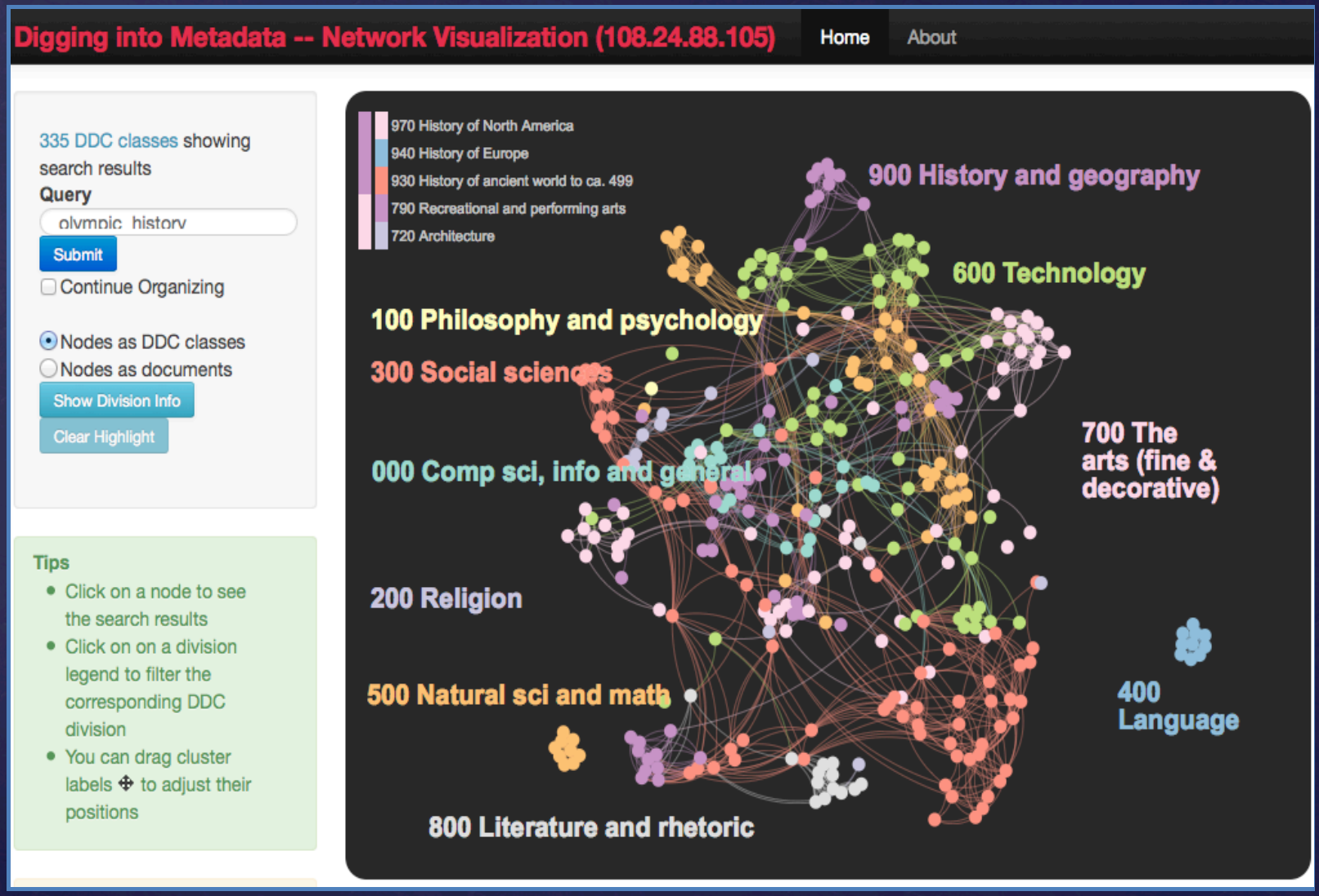
Interface 1


Digging into Metadata -- Baseline Search
General Tips
Click on a DDC class in search results to filter
Click on ✖ icon from the DDC Filter to remove the corresponding filter
Enter Query to Search
olympic history
Search
128 results found | Page 1 of 13 next
INTUTE Olympic history
DDC 796 943 945 942 949 941 940 948 944 947
IPL Olympic History
DDC 796 799 797 798 791 725 720 779 700 704
IPL Winter Olympics History
DDC 796 799 794 795 793 791 792 790 943 945
IPL Women in the Olympics
DDC 796 799 797 798 791 725 704 688 613 617
INTUTE Olympic women
DDC 796 799 797 791 798 305 306 362 704 782
INTUTE Olympic medalists
DDC 796 799 794 795 793 737 725 688 333 639
IPL Special Olympics
DDC 796 799 797 791 798 793 792 794 725 720

Interface 2


Digging into Metadata -- Tree-based Search
DDC TREE NAVIGATION
000 computer science information and general works (4) +
200 religion (2) +
300 social sciences (37) +
500 natural sciences and mathematics (5) +
600 technology (29) +
700 the arts fine and decorative arts (92) +
800 literature and rhetoric (2) +
900 history and geography (46) +
Enter Query to Search
olympic history
Search
128 results found | Page 1 of 13 next
INTUTE Olympic history
DDC 796 943 945 942 949 941 940 948 944 947
IPL Olympic History
DDC 796 799 797 798 791 725 720 779 700 704
IPL Winter Olympics History
DDC 796 799 794 795 793 791 792 790 943 945
IPL Women in the Olympics
DDC 796 799 797 798 791 725 704 688 613 617
INTUTE Olympic women
DDC 796 799 797 791 798 305 306 362 704 782
INTUTE Olympic medalists
DDC 796 799 794 795 793 737 725 688 333 639
IPL Special Olympics
DDC 796 799 797 791 798 793 792 794 725 720

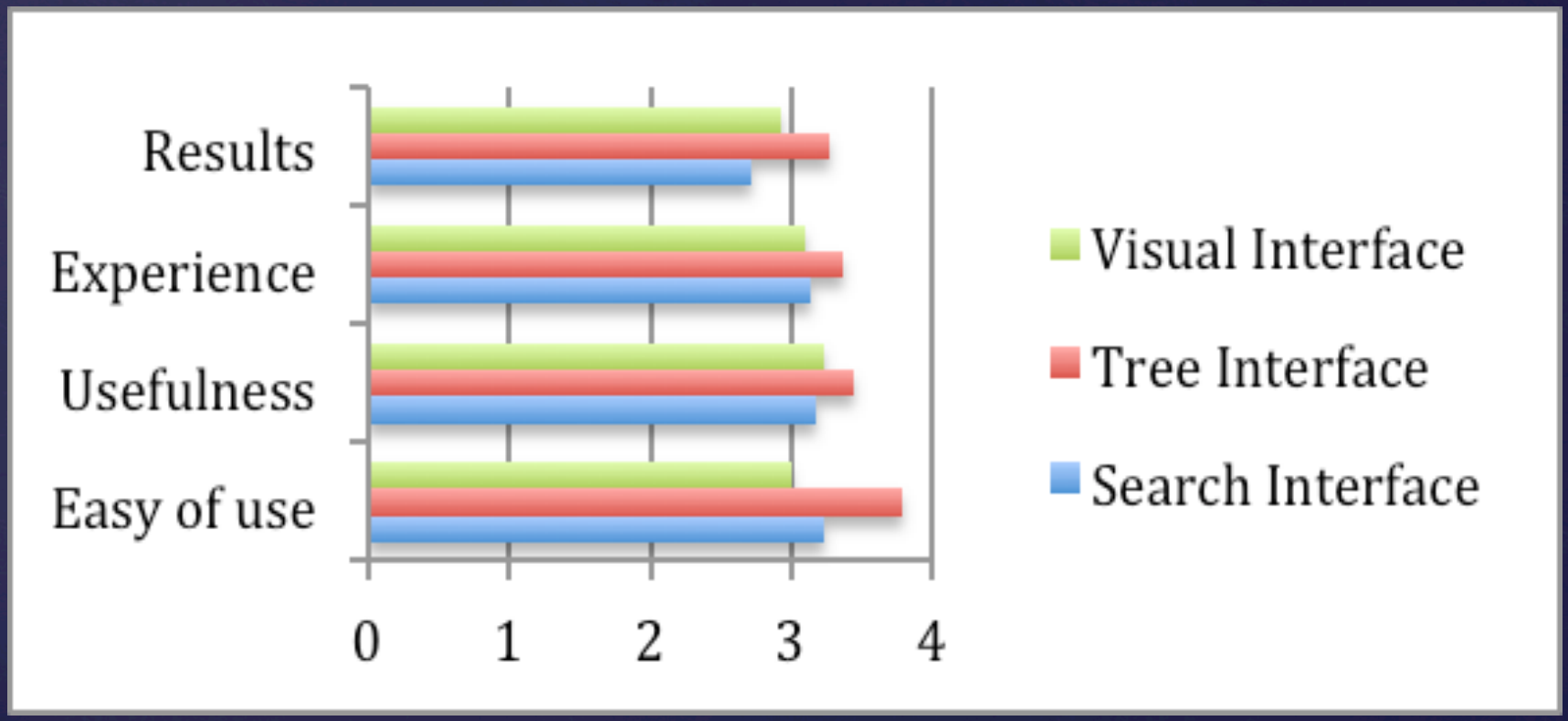
Interface 3


Digging into Metadata -- Network Visualization (108.24.88.105)
Home
About
335 DDC classes showing search results
Query
olympic history
Submit
Continue Organizing
Nodes as DDC classes
Nodes as documents
Show Division Info
Clear Highlight
Tips
Click on a node to see the search results
Click on on a division legend to filter the corresponding DDC division
You can drag cluster labels to adjust their positions
970 History of North America
940 History of Europe
930 History of ancient world to ca. 499
790 Recreational and performing arts
720 Architecture
900 History and geography
600 Technology
100 Philosophy and psychology
300 Social sciences
700 The arts (fine & decorative)
000 Comp sci, info and general
200 Religion
500 Natural sci and math
400 Language
800 Literature and rhetoric

The data Collected

- Pre/Post questionnaires
- Subjects' performance results
- Subjects' interaction logs
  - designed to log the user's every clicks
- Server's search logs
  - Include user's queries and time spent on each page

Results


Results
Experience
Usefulness
Easy of use
0
1
2
3
4
Visual Interface
Tree Interface
Search Interface

Subject's comments

- The interfaces gave the subjects new perspectives on DDC
  - "I did not know such classes exist"
  - "all these classes are useful"; "it is interesting to explore, but it does not help much for the topic"
  - "Once I was in a detailed topic I found it hard to return to look through broad DDC codes"
- The visual interfaces are both interesting and confusing
  - The visual interface "was most interesting but I feel like it was a bit hard to find the information that I Wanted"
  - the graph "was easy to understand, but a lot of things were unrelated to the search"

Results

- Successful
  - Use DDC class as a filter for searching
  - Let the user find DDC classes to explore
  - Let the user interact with the exploratory interfaces
- Not so successful
  - Identify which interfaces help the user the most
  - Identify what kind of tasks that the interfaces are the most useful
  - Detail analysis of how the user interact with the interface

Lessons learned

- Users are interested in exploratory interfaces
  - Enriched metadata are potentially useful for exploration.
  - Visualization is potentially useful for exploration.
  - DDC are useful as filters in searching and exploration, but it still lacks of precision.
  - Putting together, we need more research to understand how each of these will help the user for what kind of tasks.
- Testing the interfaces is very challenging
  - Still searching for the right testing tasks
  - Difficult to compare the user's performance
  - Difficult to evaluate user's interactions.

Thank you
And
Questions?