A Framework of Faceted Navigation for XML Data

Takahiro Komamizu, Toshiyuki Amagasa, and Hiroyuki Kitagawa
University of Tsukuba, Japan

Outline

1. Background and Motivation
2. Preliminaries
3. Faceted Navigation over XML data
4. Framework
5. Experimental Results
6. Conclusion

Background

- XML has become a de fact standard for representing semi-structured documents or data.
  - Scientific field: Swiss-Prot, KEGG*, etc.
  - Business applications: ebXML†, XBRL‡, etc.
  - Download format: Wikipedia, DBLP, etc.
- As the XML data keep growing, searching desired (part of) XML data out of a huge XML repository is becoming considerably difficult.
  - Efficient methods to retrieve XML data are expected.

*: Kyoto Encyclopedia of Genes and Genome
†: Electronic Business using eXtensible Markup Language
‡: eXtensible Business Reporting Language

Search over XML data

- Path-based search
  - Use path to access to XML elements.
    - /root/to/element
  - e.g. XPath, XSLT, and XQuery
- Keyword-based search
  - Input keywords and search most likely sub-trees.
  - e.g. LCA-based approach

Problem and Strategy

- Problem
  - Cases where users have ambiguous information needs for XML data search.
    - Concrete paths or keywords are difficult to be provided.
- Strategy
  - Apply faceted navigation for XML data search.
  - Faceted navigation helps users to find the ambiguous information needs from fractions of information that objective data have.

Outline

1. Background and Motivation
2. Preliminaries
   - Faceted Navigation
3. Faceted Navigation over XML data
4. Framework
5. Experimental Results
6. Conclusion

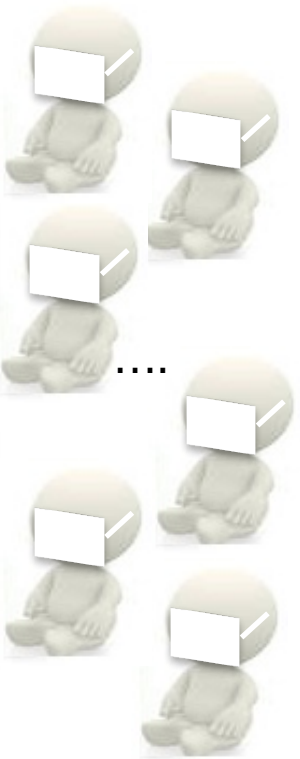
Faceted Navigation

- One of exploratory searches which enables users to explore objects through attributes, called facets.
- General faceted navigation is used for searching objects (or records) which contain multiple attributes.
- Benefits:
  - Users can observe what kinds of attributes are contained in the objects through facets.
  - Users do not need to input terms by themselves but select suggested values of facets.

Example (Medical data)

Facets

PID	Gender	Age	Weight	Disease
1234	male	18	50kg	Bronchial Asthma
1235	female	25	78kg	Pollinosis
1236	male	32	90kg	Apoplexy
⋮				
2587	female	59	66kg	Bronchial Asthma
2588	female	30	102kg	Pollinosis
2589	male	88	60kg	Bronchial Asthma
⋮				



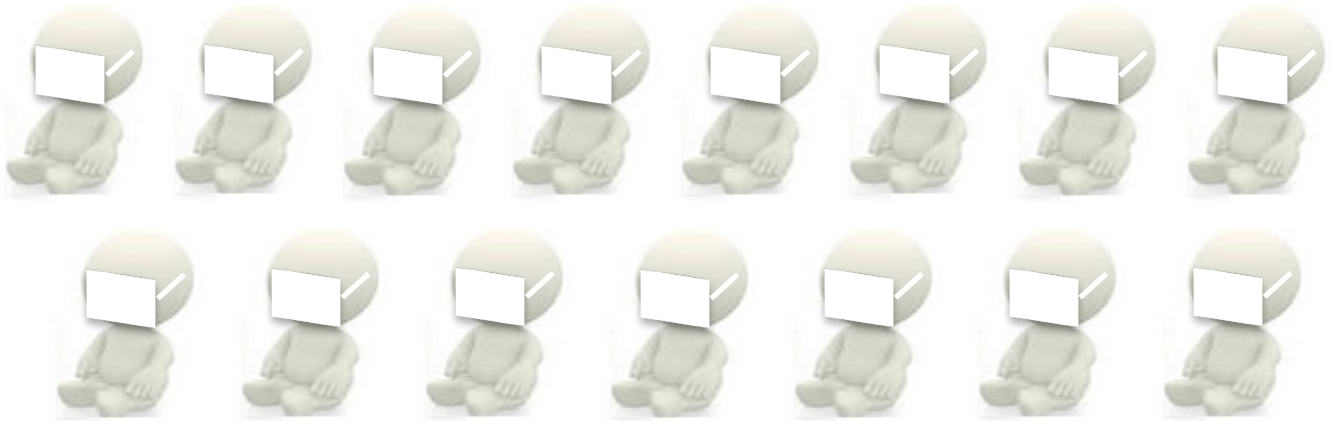
Example (cont.)

Age	count
32	52
47	48
51	47
…	

Gender	count
male	523
female	437

Disease	count
Bronchial Asthma	234
Pollinosis	176
Apoplexy	123
…	

Weight	count
50kg	10
78kg	8
90kg	4
…	



What kind of patients have Bronchial Asthma ?

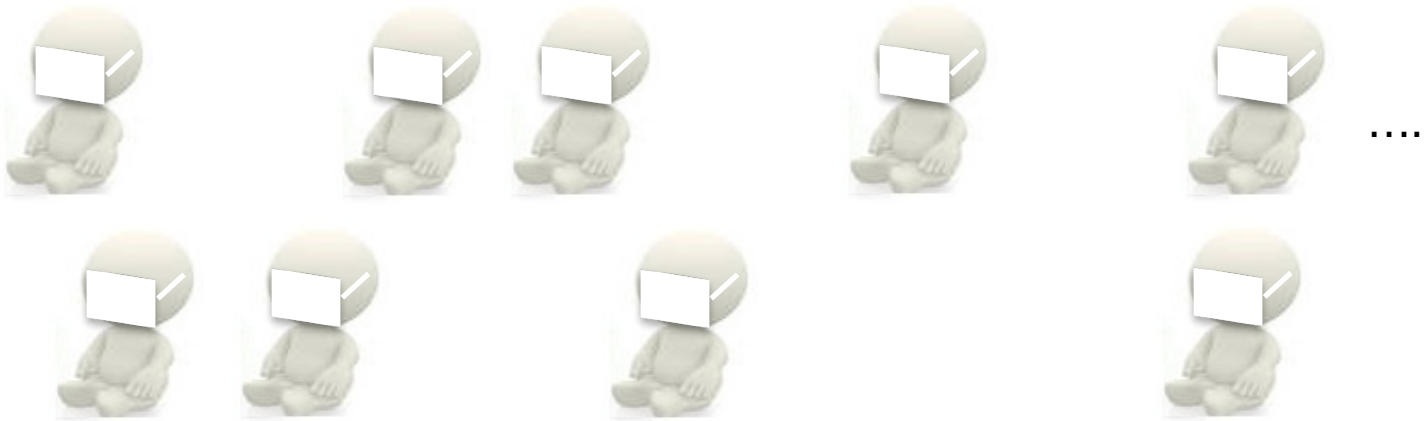
Example (cont.)

Age	count
32	30
27	26
41	25
…	

Gender	count
male	123
female	111

Disease	count
Bronchial Asthma	234

Weight	count
69kg	4
82kg	4
72kg	3
…	



What about male patients?

Example (cont.)

Age	count
32	28
41	20
27	17
…	

Gender	count
male	123

Disease	count
Bronchial Asthma	123

Weight	count
69kg	4
82kg	3
91kg	3
…	




....


Outline

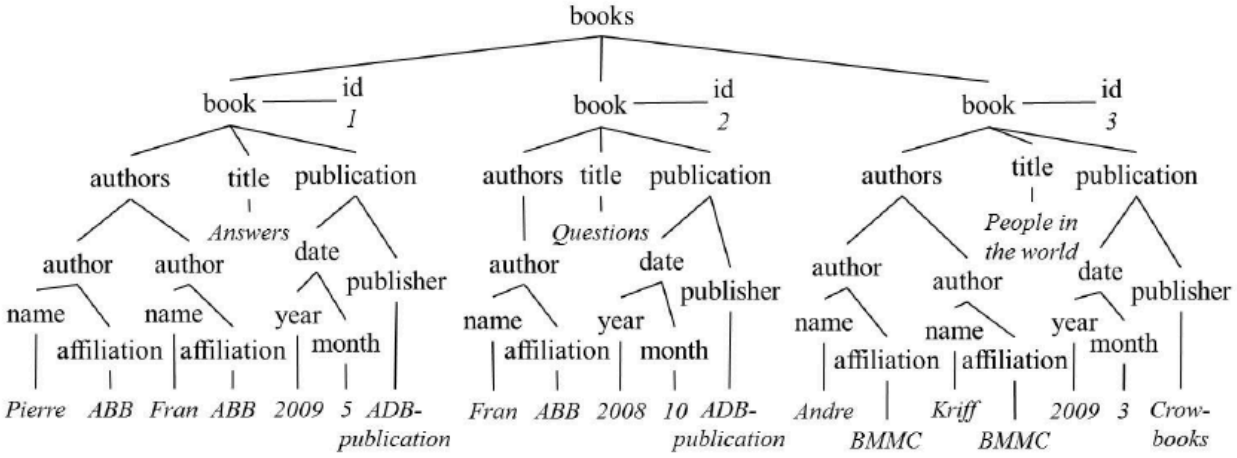
1. Background and Motivation
2. Preliminaries
3. Faceted Navigation over XML data
   - Challenges
   - Definitions of concepts
   - Operations
4. Framework
5. Experimental Results
6. Conclusion

Challenges

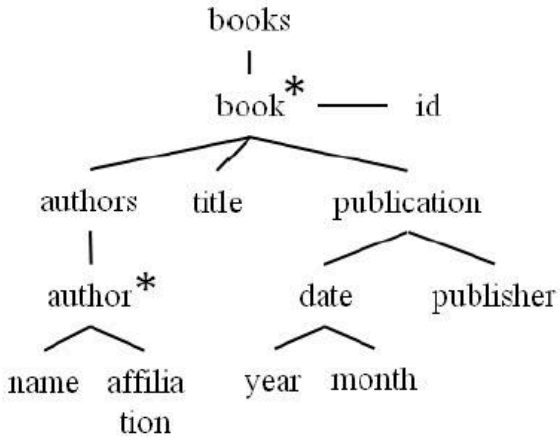
- Unlike general faceted navigation, faceted navigation over XML data has following challenges:
  - Which sub-trees or elements should be objects?
  - What are facets?
  - How to interact with XML data through faceted navigation?

Structural Information

- An overview of XML data.
- Several ways to express the structural information.
  - Schema: DTD, XML Schema, etc.
  - Index scheme: DataGuide, etc.



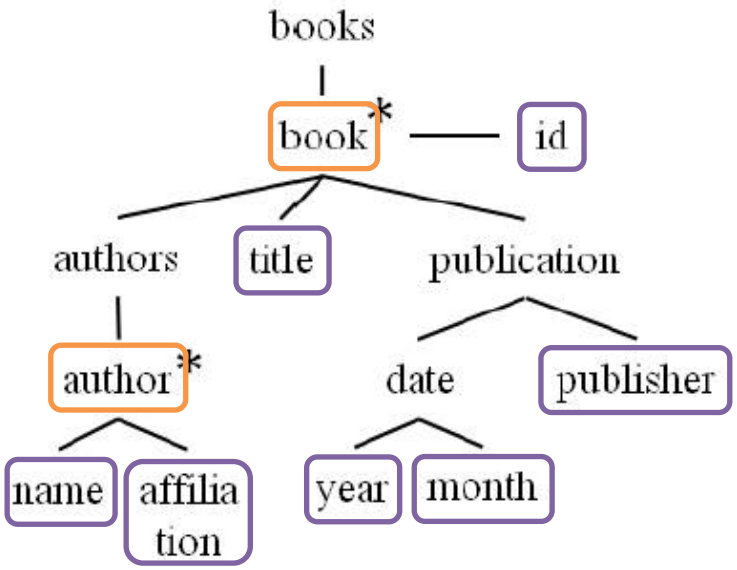
XML data



Structural information (DataGuide)

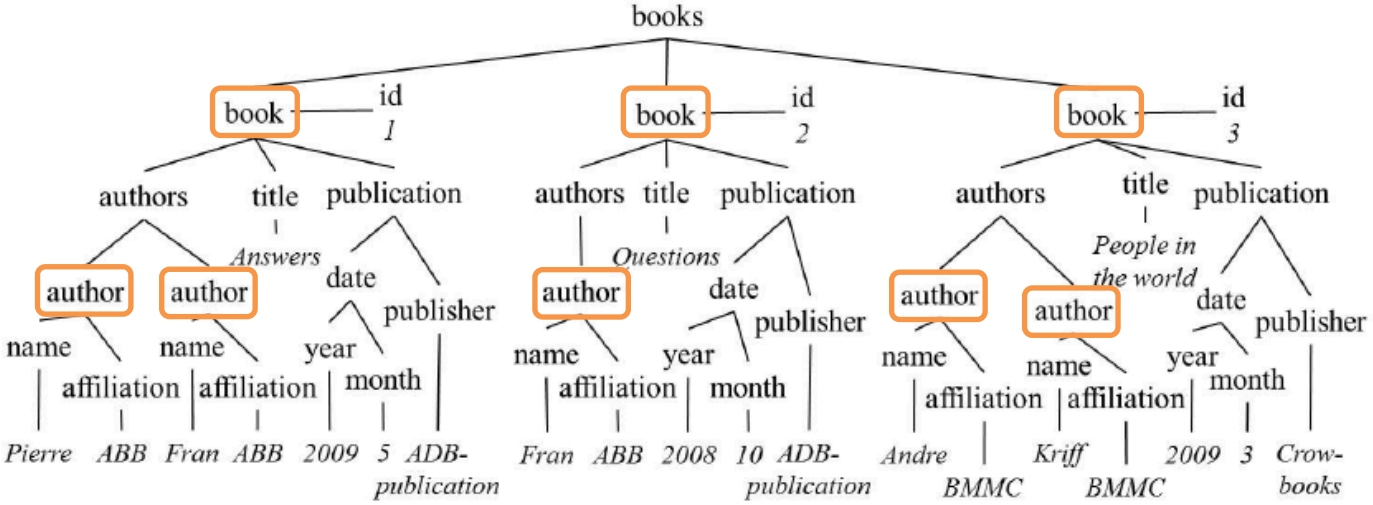
Class and Property

- Observation of Structural information
- Repeating elements seem to be objects. ➔ Class
- Text nodes show features of ancestor nodes.
- Elements which contain text node directly can be most relevant elements for it. ➔ Properties
  - each element is (in)direct descendant of the class node, and
  - there is no other class between each element and the class element.



Object and Facet

- Objects are defined as corresponding elements in XML data to class elements in structural information.
- Facets are union set of all properties exist in XML data and values of facets are existing texts in property elements.

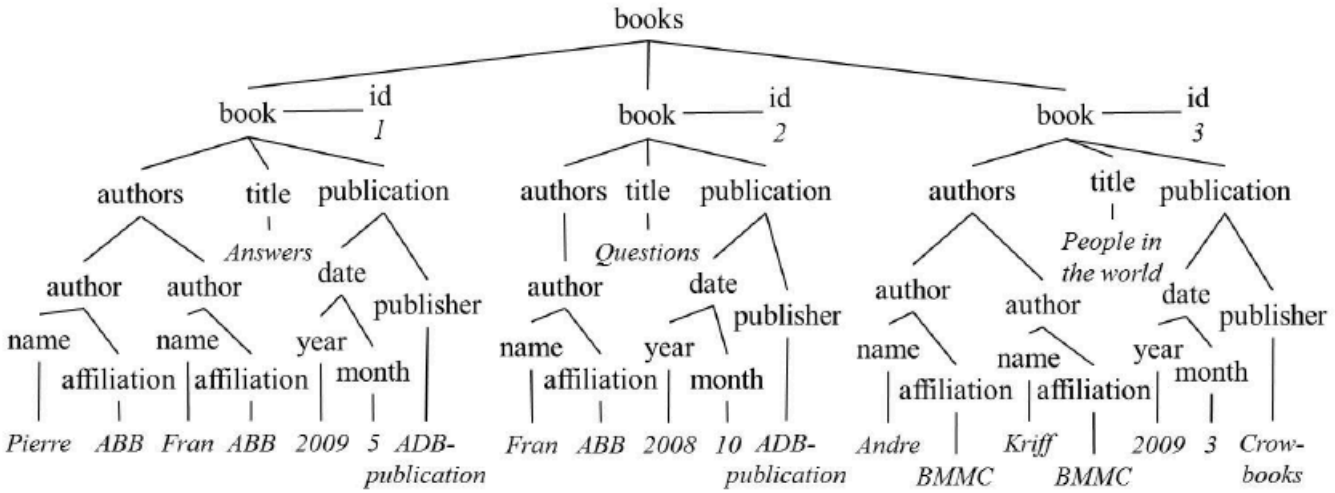


Facets
- id
- title
- year
- month
- publisher
- name
- affiliation

Operations

- We give formal definitions of interactive actions during faceted navigation.
- Operations
  - **selection operation**: A user selects a facet and its value to narrow objects down.
  - **class-based selection**: Since multiple classes possibly appears in nature, a user selects a class to narrow down.
  - **keyword-based selection**: To increase the usability of the proposal, it supports keyword search from users.
  - **path-based operations**: Nodes may have same name but different context. In the case of facets and classes, nodes of them are ambiguous. To address this problem, users can specify context of facets and classes as a path.

Example of Operations



name	count
Fran	2
Pierre	1
Andre	1
Kriff	1

year	count
2009	2
2008	1

publisher	count
ADB-publication	2
Crow-books	1

…

Facets

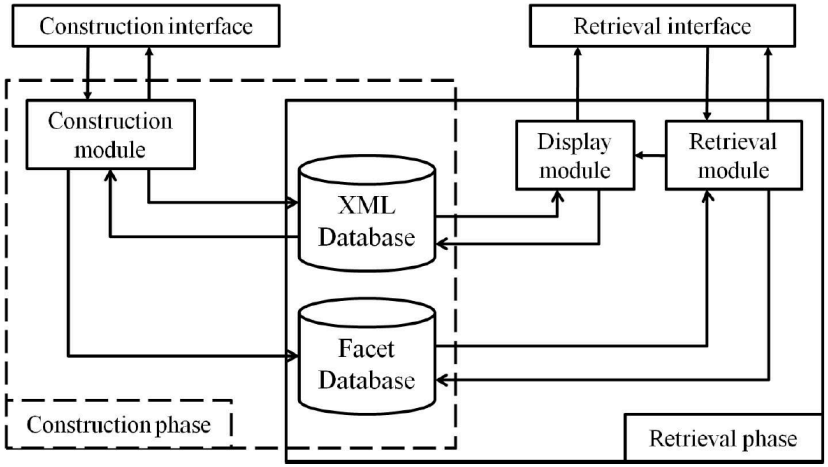
class	count
author	5
book	3

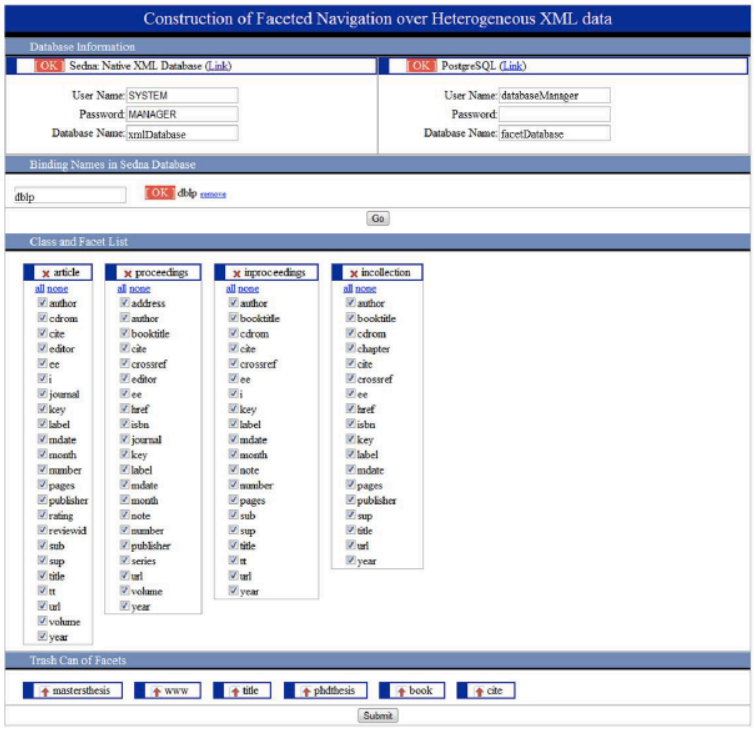
Classes

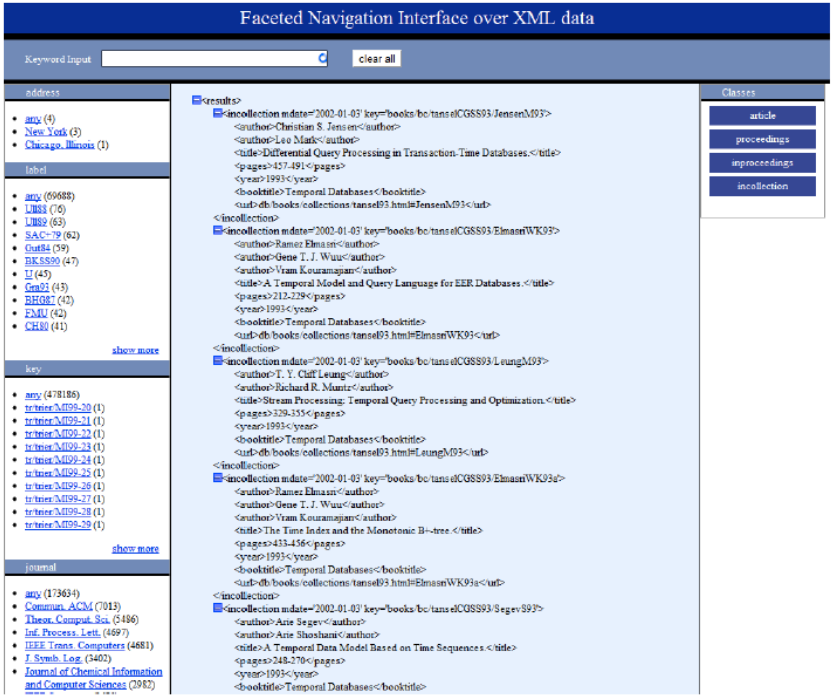
Outline

1. Background and Motivation
2. Preliminaries
3. Faceted Navigation over XML data
4. Framework
5. Experimental Results
6. Conclusion

Architecture of Framework







Outline

1. Background and Motivation
2. Preliminaries
3. Faceted Navigation over XML data
4. Framework
5. Experimental Results
   - Settings
   - Results
6. Conclusion

Experimental Settings

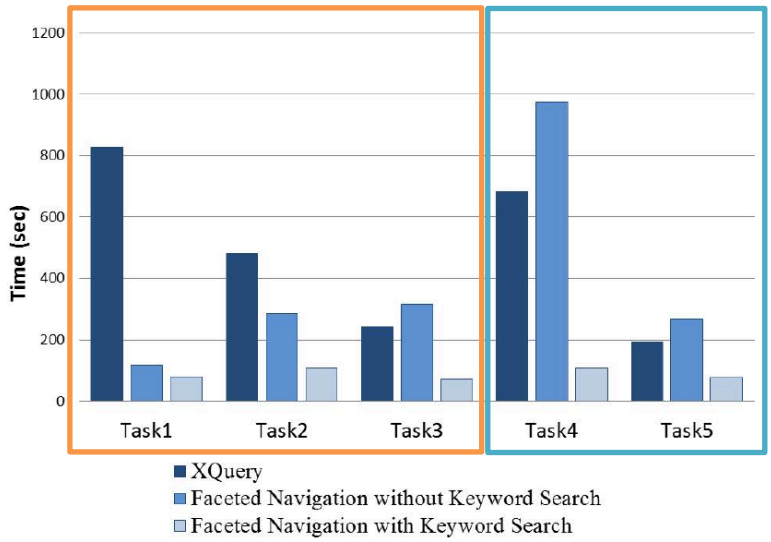
- User study with 10 examinees.
- Data: DBLP XML data
- Give 5 tasks, 3 for *exploratory tasks* and 2 for *ad hoc query tasks*, and for each task, time how long examinees take to terminate it.
- Additionally, 2 questionnaires were done for evaluation.

Exploratory Task
In your research group, each member is asked to tell the best researcher who one thinks the best to share one's research interest among the members. In addition, you need to find the most prosperous year in terms of research achievements, as well as the year when those achievements are made.

Ad Hoc Query Task
Imagine that you are taking a course named "Systematic Languages" in that you learn several kinds of programming languages. From the next class, you will learn OCaml, and you are asked to read a paper entitled "Using, Understanding, and Unraveling the OCaml Language. From Practice to Theory and Vice Versa." Find this paper.

Results

Task1, 2, and 3 are exploratory tasks,
and Task 4 and 5 are ad hoc query tasks.

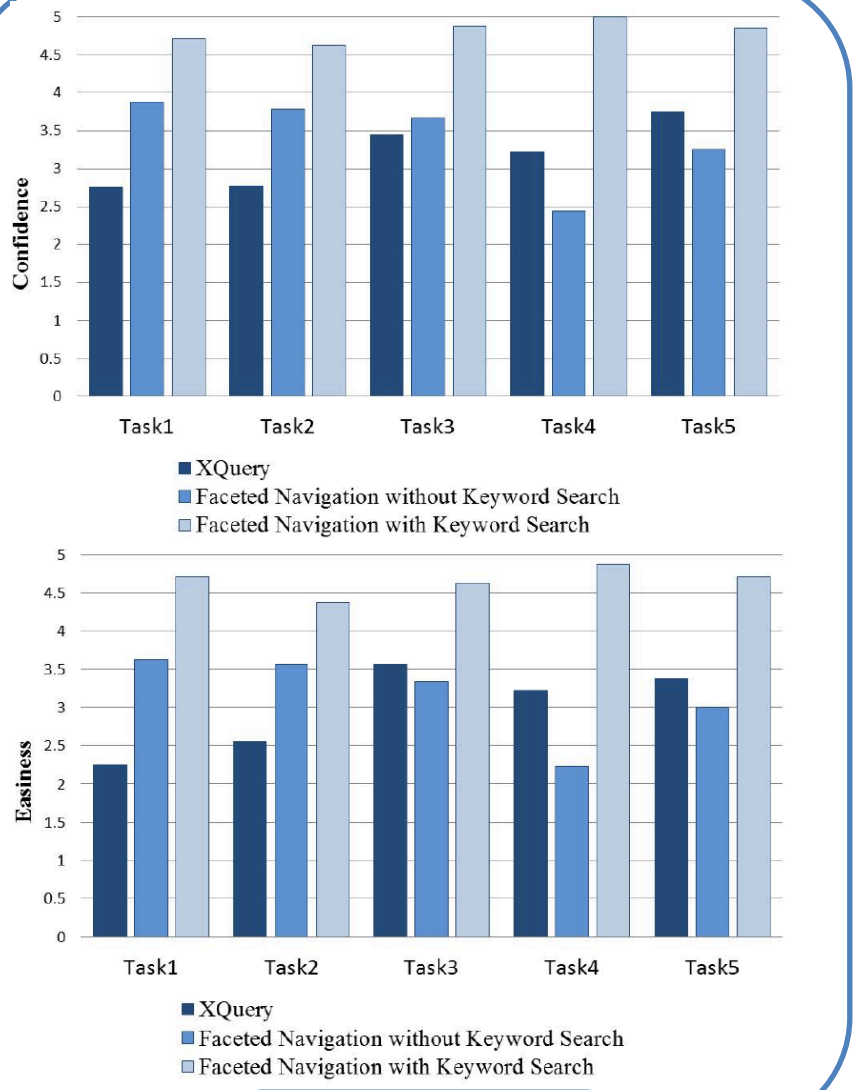


Elapsed Time

XQuery

Faceted Navigation without Keyword Search

Faceted Navigation with Keyword Search



Questionnaires

Outline

1. Background and Motivation
2. Preliminaries
3. Faceted Navigation over XML data
4. Framework
5. Experimental Results
6. Conclusion

Conclusion

- Faceted navigation for XML data.
  - Definitions of concepts:
    - Class and Property
    - Object and Facet
  - Operations
    - Selection and path-based selection
    - Class-based selection and path-based class-based selection
    - Keyword-based selection
- Experimental results show the proposed scheme better usability than XQuery and keyword search works well with faceted navigation.

Thank you for your attentions.