

# **Theme Creation for Digital Collections**

Xia Lin, Jiexun Li, Xiaohua Zhou  
Drexel University  
Philadelphia, USA

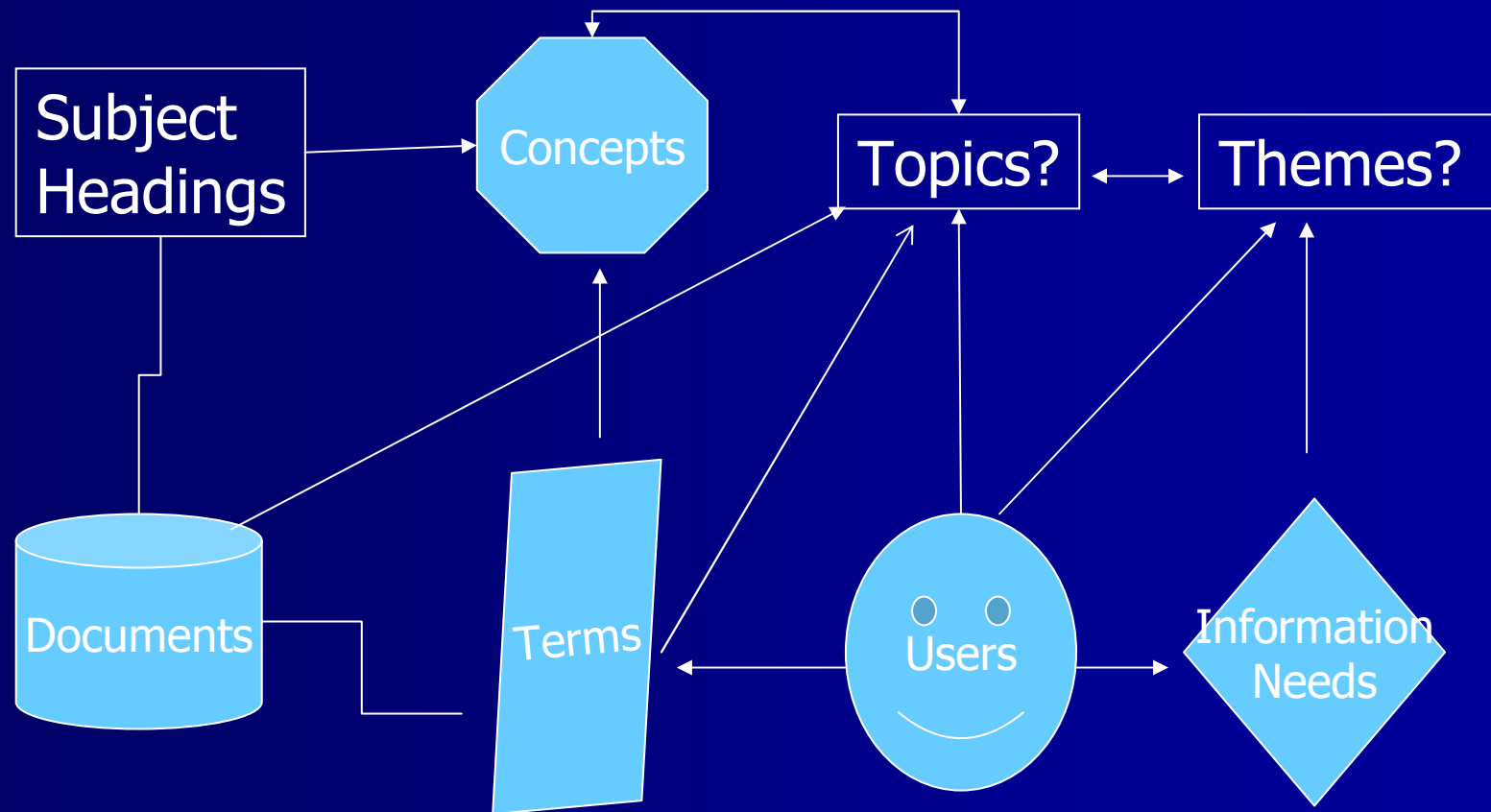
# Challenges in Concept Representation

- .....
- Representations of associative relationships
  - Concept relationships are not always in hierarchies.
  - There are many types of associative relationships
- Dynamic concept representation

# Dynamic Concept Representation

- Concepts are only meaningful when
  - specific domains or document collections are identified.
  - specific context of use is understood
  - specific groups of users are identified
- The same concept might be looked at from different perspectives when the user's information needs are different.
- How do we catch all of these in concept representation?

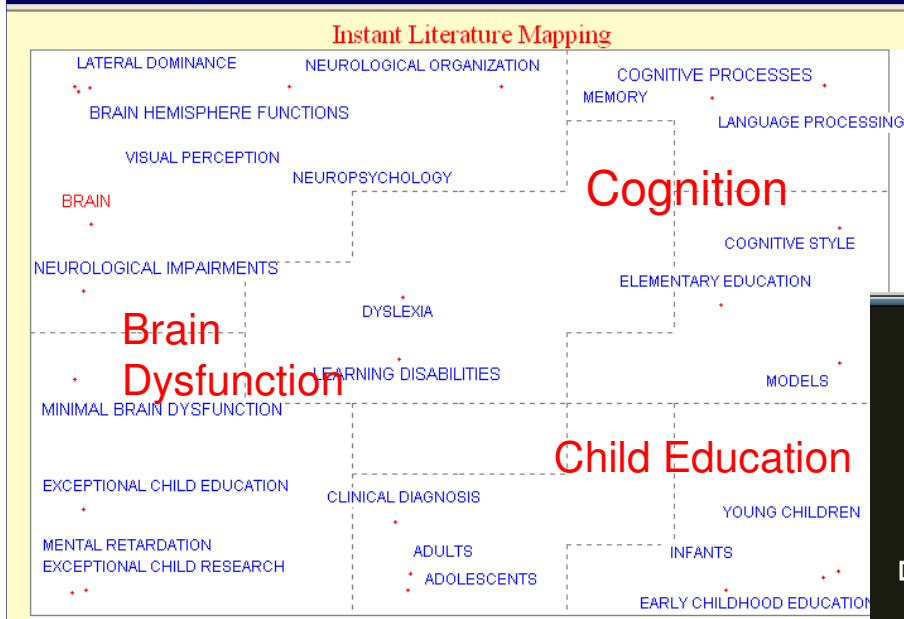
# Looking for solutions ...



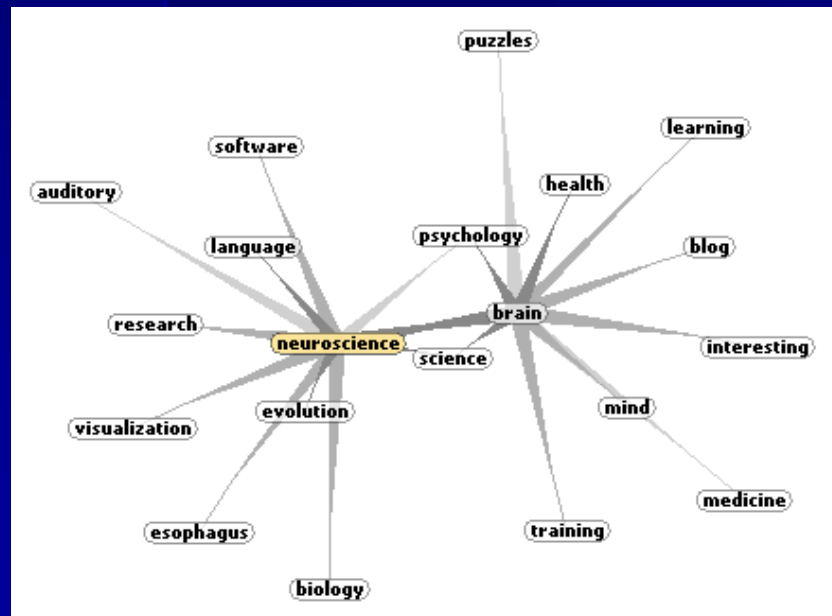
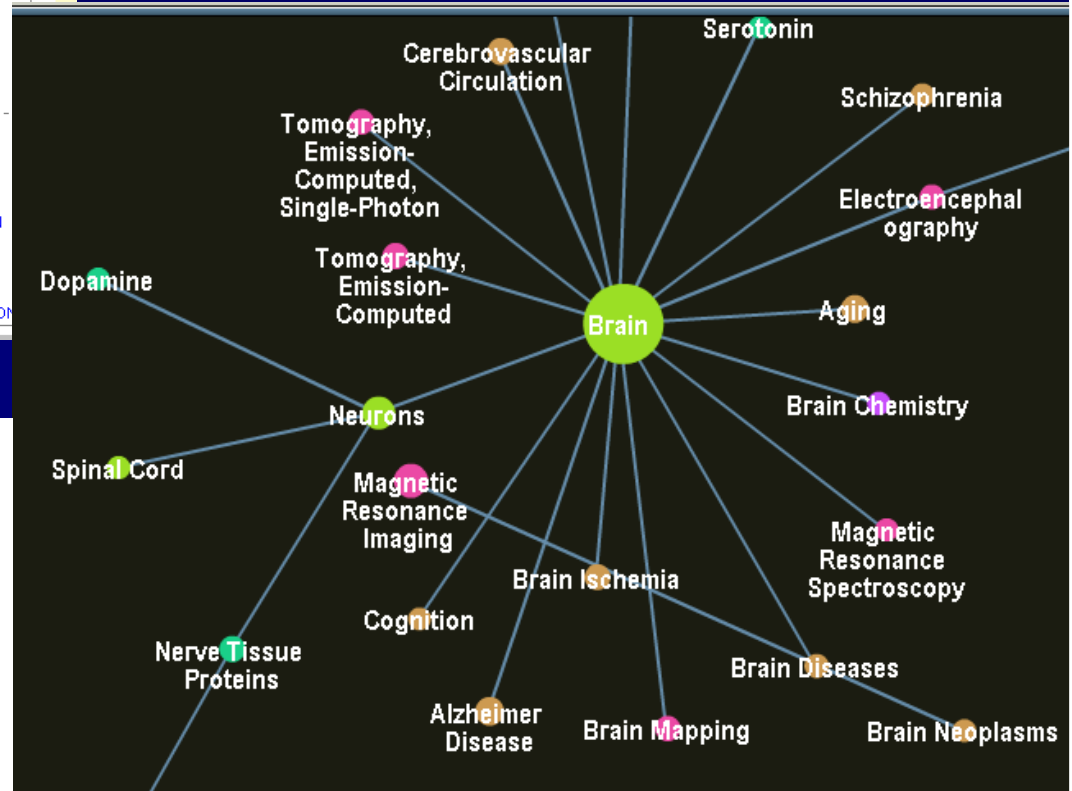
# Getting Users Involved

- Approach 1
  - Semantic Mapping + Visualization
    - Computers extract associations from documents and terms
    - Users understand them through visual displays and visual interactions
  - My ISKO2006 Paper
    - “Dynamic Concept representation through a Visual Concept Explorer”
    - Example: the “Brain” maps

## "Brain" Map in ERIC



## "Brain" Map in Medline

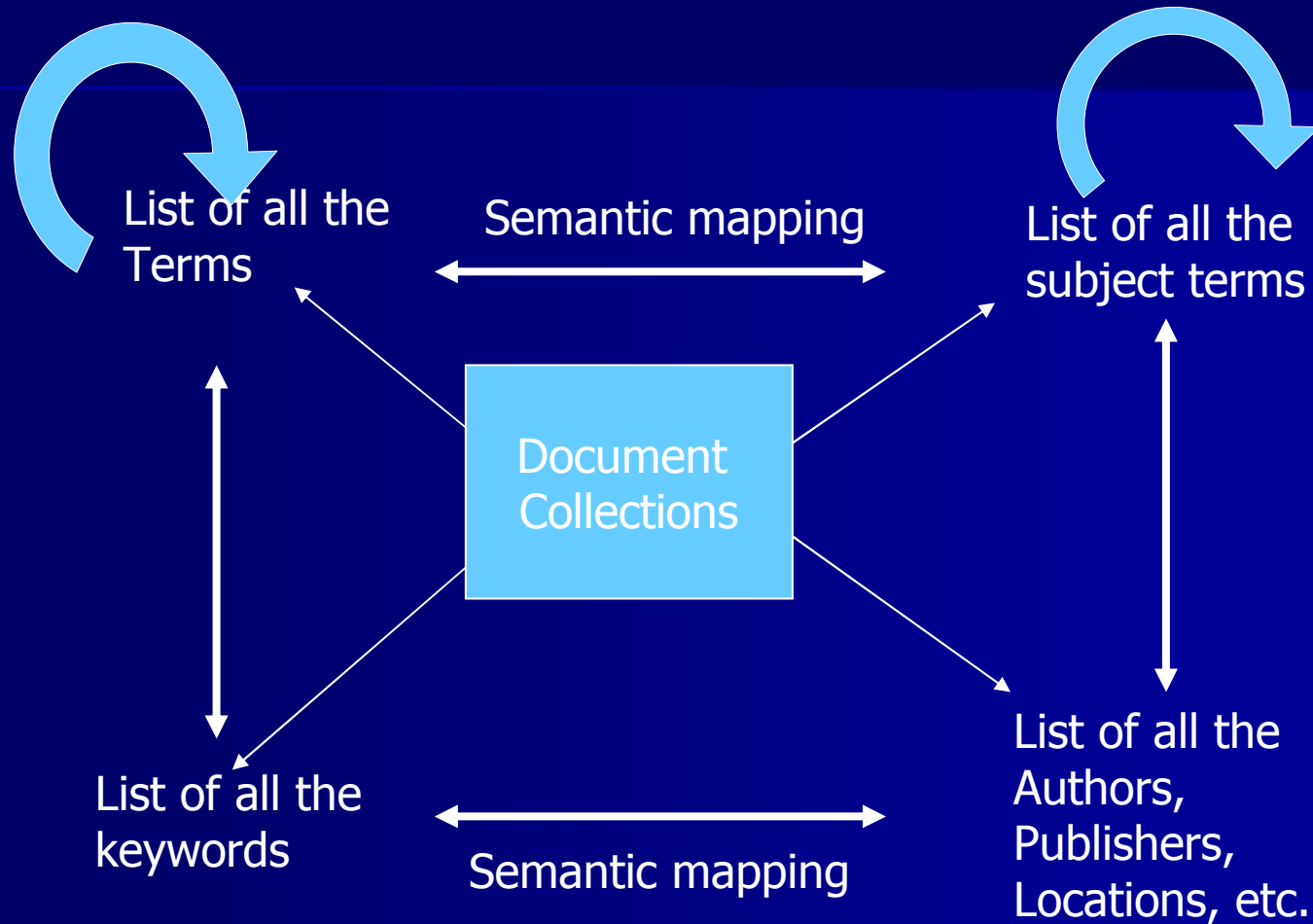


## "Brain" Map in Del.icio.us

# Getting Users Involved

- Approach 2
  - Semantic Mapping + User-initiated Theme creation
    - Collection-specific semantic mapping
    - Clustering of user search terms (from the search log)
    - User's creation of themes and topics
    - User-computer collaboration of theme editing/updating/dynamic changing

# Semantic Mapping





# Semantic Profiles

- What indexing terms are mostly likely associated with a given free-text term?
  - The mapping of indexing terms to free terms based on the statistic associations within a collection
  - The mapping reflects the nature of the collection.

Free-text Term: Health	
<i>IPL</i>	<i>IPL-Teens</i>
health	health
Mental health	fitness
health care	exercise
disease	nutrition
activism	stress
disorders	tic
public Health	panic
psychology	medicine
mental	attention
Illness	deficit
reproduction	ADHD
medicine	depression

# Topic Signatures

- What terms are mostly likely associated with a given indexing term?

Indexing Term: Health	
<i>IPL</i>	
Stress	0.0591
depress	0.0339
teen	0.0292
anxiety	0.0286
health	0.0284
disorder	0.0264
eat	0.0226
Mental	0.0221
.....	

# DEMO:

- <http://project.cis.drexel.edu/IPL/>

## Semantic Mapping of IPL Subjects

Keywords are subject terms indexers assigned to each item for subject representations. Please look up similar keywords before creating a new one.

Mapping:

```
metadata
```

**Go!**

Subject Terms	Related Subject Terms
Organizing the Web	
Standards	
Online Catalogs	
Canada	
Cataloging	
HTML	

Keywords	Related Keywords
help	
Icons	
testing	
Tests	
Cataloging	
World Wide Web	
Internet	

Click a term on the left-hand side to retrieve related keywords on the right

# Semantic Mapping of IPL Subjects

Keywords are subject terms indexers assigned to each item for subject representations.  
Please look up similar keywords before creating a new one.

Mapping:

**Go!**

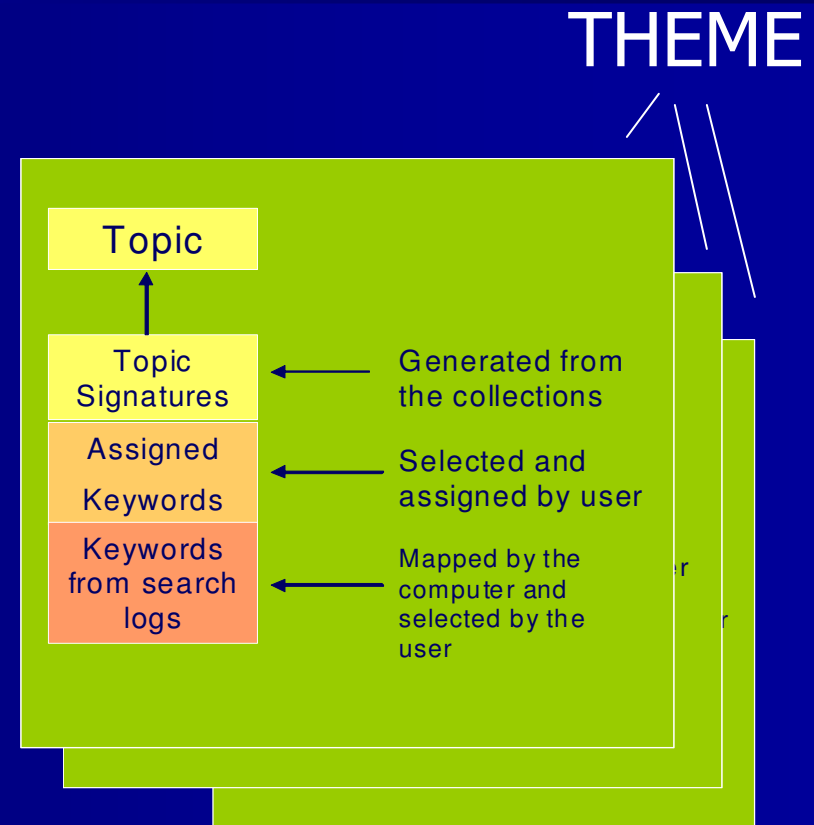
Subject Terms	Related Subject Terms
Intellectual Property	Copyright
Patents	Intellectual Property
Copyright	Patents
Non-U.S. & International	State Parties
Censorship	Law & Law Enforcement Blogs
Issues	United States History
Law & Law Enforcement Blogs	United States
Biography	Political Parties (United States)
Other Sports	State Digital Libraries in the U.S.
Non-fiction	United States Travel & Tourism

Keywords	Related Keywords
intellectual property	copyright
copyright	intellectual property
Patents	Patents
Witchcraft	United States
Poets	Turabian
Censorship	Modern Language Association
Inventors	MLA
Wicca	Chicago
Law	Supreme Court
German	Legal

Click a term on the left-hand side to retrieve related keywords on the right

# Theme creation


- An attempt to bring together three types of semantics to create dynamic representation of concepts and topics
  - semantic analysis of collections.
  - User's search patterns and search vocabulary
  - Controlled vocabularies or indexing terms



# Current Research - 1

- Develop a method of concept representation that are
  - Dynamic
    - Over collections, users, time, ec.
  - Context-sensitive
    - Same themes may be defined different for different collections and different usages.
  - Semi-automatic
    - integrates both manual and automatic processes.
  - Collaborative
    - Allows users/indexers to make changes.

# Wiki interface for theme creation



ipl LABS

Log in / Create account

[page](#) [discussion](#) [edit](#) [history](#)

## American Literature

**Themes**

The collection page for all themes of IPL.

24 articles in [English](#)

- Literature
- History
- People
- Teens
- Social Issues
- Sciences
- Technology
- Environment
- Health
- All Subjects

**Literature**

Themes under this subject.

- American Authors
- American Literature**
- Banned Books

[All Subjects](#) > [Literature](#) >> **American Literature**

**American Literature**

**Description:**

The study of American literature and American authors.

[Show more themes](#)

**Topic 1: Author Names** [\[edit\]](#)

Topic 1: Author Names	Related Resource
<p><b>Keywords:</b> <a href="#">American Poets</a>; <a href="#">Minority American Authors</a>; <a href="#">American Authors</a>; <a href="#">American Playwrights</a>; <a href="#">Female American Authors</a></p> <p><b>Other Keywords:</b> <a href="#">American Poets</a>; <a href="#">Minority American Authors</a>; <a href="#">American Authors</a>; <a href="#">American Playwrights</a>; <a href="#">Female American Authors</a></p>	<p><b>IPL URL:</b> <a href="#">Voices from the Gap</a>; <a href="#">Today in Literature</a>; <a href="#">The Whitman Archives</a>; <a href="#">Literary Kicks</a>; <a href="#">Mark Twain</a></p> <p><b>Other URL:</b> <a href="#">American Literature Anthology</a>; <a href="#">American Authors</a>; <a href="#">PAL: Perspectives in American Literature</a>; <a href="#">American Authors on the Web</a></p>

ipl labs

- Main Page
- mission
- member institutions
- people
- Help

projects

- IPL architecture redesign
- New Search Interfaces
- Metadata authoring tools
- IPL Themes
- Semantic Mapping of IPL collections
- Digital References
- Map-based public library Locator
- Suggest a Project

ipl links

- IPL Site
- Hypatia
- IPL Training

search

[Go](#) [Search](#)

toolbox



# Current research - 2

- Test the theme creation process through IPL
  - A group of students work together to create themes for IPL
    - Analyze three months of search logs together
    - Each describes a particular “theme” he or she “reads” from the search log.
    - User the semantic mapping tool to create and enhance the representation.

# Browsable Theme Interface

The Internet Public Library

Search the IPL

  
[Advanced Search & Help](#)

**About the IPL**  
**Subject Collections**  
Arts & Humanities  
Business  
Computers  
Education  
Leisure  
Health  
Law & Government  
Regional  
Science  
Social Science  
**Ready Reference**  
Almanacs  
Dictionaries  
Encyclopedias  
More Ready Reference...  
**Reading Room**  
Books  
Magazines  
Newspapers  
**KidSpace**  
**TeenSpace**  
**Special Collections**  
Election  
Literary Criticism  
POTUS (U.S. Presidents)  
Science Fair

**Themes**

**Literature**

- [American Authors](#)
- [American Literature](#)
- [Banned books](#)

**History**

- [History of United States Military Conflicts](#)
- [Presidents of the United States of America](#)

**People**

- [Influential Americans](#)
- [American Leaders](#)
- [U.S. Presidents in Context](#)
- [Biographies](#)

**Teens**

- [Raising and Nurturing a Teenager](#)
- [Teen Entertainment](#)

**Social Issues**

- [Race & Ethnicity](#)
- [Public Policy Issues](#)

**Sciences**

- [Science Fair Projects](#)
- [Topics in Science](#)

**Technology**

- [The Intersection of Computers and Libraries](#)
- [Personal Internet Entertainment](#)
- [Web Service Hubs](#)
- [Emerging Technologies](#)
- [Entertainment/ Social Networking](#)

**Environment**

- [The World We Live In](#)
- [Environmentalism](#)
- [Geography](#)
- [World Locations](#)

**Health**

- [Diabetes](#)
- [Health Disorders and Prevention](#)

# Current research - 3

- Develop an IPL metadata creation tool with semantic supports.
  - Utilize the semantic mapping and theme creation process
    - Suggest associative lists for keywords and subject terms
  - Integrate multiple subject-representation objects
  - Support collaboration

## Semantic Mapping

Page 1 of 1

Go!

## 4446

[illegible]

## The IPL General Collection

Edit	Save	Delete	Marked
------	------	--------	--------

Freetext to Keyword ▼

Go

[illegible]

Go

Search for terms used in previous user queries

# Future Research

- Dynamic concept representation
  - It needs more research!
- Use of Coding schemes
  - SKOS, Topic Maps, and others
- Evaluation
  - “Aboutness” is very difficult to evaluate.
  - “Usefulness” might be the criteria!