Resource and Resource Sharing in Intelligent Information Access

Jiangping Chen
Department of Library and Information Sciences
University of North Texas
October 23, 2008
Today’s Presentation

- Intelligent Information Access (IIA) and Resources
- Purposes of Research
- Research Design
- Results & Finding
- Discussion
- Conclusion and Future Work
Intelligent Information Access

- Make use of human knowledge or human-like intelligent to meet special inquiries of information users.
  - Automatic classification and clustering
  - Question answering
  - Cross-language information retrieval
  - Machine translation
  - Information extraction
  - Text summarization
IIA is Interdisciplinary

- Computer Science
- Linguistics
- Library Science
- Information Science
IIA Research Forums

- Text REtrieval Conference (TREC)
- Cross-Language Evaluation Forum (CLEF)
- NII-NACCSIS Test Collection for IR Systems Workshop (NTCIR)
- The Association for Computational Linguistics (ACL)
- ACM Special Interest Group on Information Retrieval (SIGIR)
Motivation of Research

- IIA researchers typically make use of various knowledge resources and tools in order to design, construct, and evaluate their systems, many of these resources are developed by others.

- Insufficient understanding of the current state of resource sharing among IIA Researchers.

- Lack of an effective model of resource management to facilitate the sharing of resources.
Purposes of Research

- Identify types of resources used in IIA research and development
- Identify characteristics of these resources
- Understand how IIA researchers acquire and share resources
- Develop a framework for resource management
Exploratory Study with two stages

- A content analysis of QA and CLIR publications
  - Analyze 145 CLIR and QA research papers or reports published in 2005/2006
  - The analysis is based on a coding scheme deductively developed in this study
- An Online Survey of the IIA community
  - Online survey using SurveyMonkey
  - Call for participation sent to related mailing lists
Results (1)

- Identified 290 resources from the 145 sample papers. Of the 290 resources, 118 (40.7%) belong to knowledge sources and 171 (59.0%) are various computer systems. One resource is coded as A.3: “Other.”

- Knowledge sources include lexicons, dictionaries, thesauri, annotated or non-annotated corpora, and test collections. Figure 2 depicts the distribution of knowledge sources.

- Computer Systems include computer programs for information retrieval, natural language processing, machine translation, classification, and others.
Various types of knowledge sources used for QA and CLIR

- A.1.1 Monolingual lexicon/ontology
- A.1.2 Bilingual or multilingual lexicon/ontology
- A.1.3 Annotated corpus
- A.1.4 Un-annotated corpus
- A.1.5 Test collection
- A.1.6 Other
- A.1.7 Unclear
Various types of software systems used for QA and CLIR

A.2.1 Information retrieval system
A.2.2 Natural language processing tool
A.2.3 Machine translation tool
A.2.4 Classifier
A.2.5 Other
A.2.6 Unclear
Results (4): Characteristics of IIA Resources

- More than half of the resources (59.2%, 170 resources) can be clearly identified as freely available.
- Most researchers have used resources developed by others doing research or development. Among the 145 sample papers, only 13 (8.9%) did not use any resources developed by others.
- A resource can be used quite differently from its original intended use. IIA researchers are creative in making use of available resources.
Results (5): Resource Acquisition and Sharing

- Researchers (93.1%) employed web search engines to locate resources
  - Other channels include academic association websites, conferences, personal connections;
- Majority (77.6%) states that sometimes they can find the right resources to use, only 6 responses stating they can always find them
- Researchers face various challenges during acquiring and using of resources developed by others
- Money, reputation and ease of access rank as the top factors for choosing a resource to use
Discussion

- Issues related to the coding scheme and the content analysis method
- IIA definition and community
- Problems and solutions to resource acquisition and sharing
  - A resource repository could be the solution if it can be well maintained
  - Functions of the system should be based on the needs of the community
Future Research (1)

- Toward a resource management model applying KM process framework by Bukowitz and Williams (1999) – the tactical component
  - Get – search and access resource from the system;
  - Use – make use of resources and create new resources;
  - Learn – interact with each other to meet needs for resources;
  - Contribute – share resources by contributing them to the system;
Future Research (2)

- Toward a resource management model applying KM process framework by Bukowitz and Williams (1999) – the strategic component
  - Assess – evaluation of the resources and planning of the system;
  - Build – system design and implementation;
  - Sustain – make the system sustainable.
Future Research

- Other frameworks may also be considered
- What are the most challenging issues in the framework?
  - Sustainability
  - The use of IIA technologies to speedup the IIA resource repository development?
Any Suggestions are Welcome!

Thank You!