The ALCOM/NIST Heterogeneous Structures Database: Knowledge Structure for Basic and Applied Research in an Interdisciplinary Scientific Collaboration

Abstract: Scientific enterprise represents a pivotal area in knowledge organization because of its prolific use of emerging information technology and its close network of researchers in government, industry, and universities as producers and consumers of knowledge. The ALCOM/NIST Heterogeneous Structures Database (http://cpip.kent.edu/PSP) is an online information system involving university research scientists, government agencies, and industrial partners in the ALCOM/NIST Heterogeneous Structures Project. The Database is the first stage of a multiphase knowledge base of liquid crystal science, the Basic and Applied Liquid Crystal Research Database. This paper will describe the system functionality of the Heterogeneous Structures Database and the expansion of a previously existing tool, the Thesaurus for Liquid Crystal Research and Applications (http://www.personal.kent.edu/~sllslzeng/maja2.html), into the Database. The first part of the paper will discuss how the design of the Database addresses the following user characteristics from multi-disciplines involved in the project: interdisciplinarity, basic and applied research, experts and novices users, and user population groups (academia, government, and industry). The second part of the paper will report on how the authors are developing a frame of knowledge structure to meet these characteristics.

1. National Science Foundation Center of Science and Technology

The Center for Advanced Liquid Crystalline Optical Materials (ALCOM) is one of 25 NSF Science and Technology Centers and represents the largest concentration of liquid crystal research in the United States. One of five major ALCOM research projects, the Heterogeneous Structures Project involves: university physicists, chemists, and mathematicians from ALCOM (Kent State University, Case Western Reserve University, The University of Akron) and University of California at Los Angelos (UCLA); national laboratories at National Institute of Standards and Technology (NIST) and Wright Patterson Airforce Base; and industrial partners at IBM T.J. Watson Research Center. Researchers investigate the process of phase separation, which produces heterogeneous liquid crystalline materials, in order to understand the relationship between morphology and bulk physical properties. The project goal is to design and fabricate heterogeneous structures with desired optical and physical properties. The research of the five ALCOM projects covers the spectrum of leading basic and
applied research in the field of liquid crystal science.

2. Heterogeneous Structures Database

The Heterogeneous Structures Database is the first stage of a multi-phase online information system to encompass and interconnect the five major research projects in ALCOM. The long-term project, the Basic and Applied Liquid Crystal Research Database, will make available a comprehensive formal and informal knowledge base of basic and applied research in liquid crystal science. The Heterogeneous Structures Database exploits the collaborative creation and rapid transfer of knowledge between academia, government, and industry through organization and retrieval within an online environment.

The system design from which the Heterogeneous Structures Database evolves, combines the broad theoretical structure of liquid crystal research, the specific pragmatic needs of Heterogeneous Structures Project, and the Internet as our vehicle for collaboration. A team headed by Laura Bartolo established the Heterogeneous Structures Database in 1997. We recognize that researchers in collaborative projects like the Heterogeneous Structures must be able to access a variety of resources in a problem-specific manner. The information system is comprised of:

- A Web server organizing the knowledge base of the Heterogeneous Structure Project;
- An archival and retrieval gateway for polymer liquid crystal research;
- Tools to find, filter and customize information preferences;
- Mechanisms to track the generation, flow, and usage of information.

The Heterogeneous Structures Database is organized so that scientists can create, retrieve, apply, and use easily project related information. The database collects, indexes, abstracts, and updates textual and multi-media project information generated by researchers (e.g. scientific presentations, preprints, program codes, technical reports, experimental data, visualization, and modeling and simulation results). Participants contribute original research to the database through web-based submitting forms. An automated current awareness mechanism based upon personalized research profiles alerts researchers in academia, government, and industry about new research contributed to the database. The framework of the Heterogeneous Structures Database is largely grouped around three primary needs and interests of the project participants: internal materials generated within the project; participants activity; and external materials.

Internal Materials: The Recent Research and Archive components of the Database house the body of literature generated within the ALCOM/NIST Heterogeneous Structures Project. Both components arrange the documents by type of document, as well as by date of publication. Document type was determined by the range of research literature anticipated to be authored by the project researchers. At this time document type is comprised of published papers, scientific presentations, experimental data, modeling and simulation results, and visualization. In terms of publication date, the Heterogeneous Structure Project was established in 1995 and the Database archives research written during that time period. In addition, ALCOM research published from 1991 onward has been selectively included within the Database.

Participants Activity: Project researchers contribute and collaborate in the Database through the Upload, Profiling, Discussion, and Search components of the system. Project participants upload directly their research into the Database. Textual documents are provided in as many formats (HTML, LATEX, PDF, RTF) as possible to provide users with numerous avenues for acquiring materials. Researchers select indexing terms to describe their research from the Heterogeneous Structures and Liquid Crystal Thesauri which provides a consistent
taxonomy of concepts and are described in the second part of this paper. The Liquid Crystal Thesaurus presents a broad, conceptual overview of liquid crystal research while the Heterogeneous Structures Thesaurus lays out a problem-specific arrangement of heterogeneous structure terms.

Project participants provide extended abstracts of their work to improve the selection and evaluation of retrieved documents within the Database. Research Profiles are customized for each project participant to selectively disseminate new research corresponding to user needs and interest. As researchers upload documents to the Database, other project participants are automatically and selectively alerted to the new contributions based upon indexing and abstracting profile matches. Meta-descriptors embedded within the documents map selected identifiers and vocabulary terms with researchers’ work. Users are able to query the database and retrieve relevant documents based upon single or combined search requirements.

The Heterogeneous Structures Database seeks to capitalize on the rapid change in communication technologies in the context of collaborative scientific research. Web-based chat rooms and a project discussion board enable a range of collaborative work among scientists. Researchers can work together as a group in the chat room with transcripting capabilities for documentation purposes. Project participants contact authors directly and automatically through a web board for each entry contributed to the database.

External Materials: Information, such as applications and new research, relevant to the Heterogeneous Structures Project is selectively included within the Database. Project participant can add new resources which meet the work and requirements of researchers in the project. As the resources are added to the Database the tools and information are automatically assigned to locations within the Database and researchers are notified of the new resources.

3. Knowledge Organization and Interdisciplinary Scientific Work:

The Database is designed to reflect characteristics integral to the Heterogeneous Structures Project and liquid crystal research: interdisciplinary, basic and applied research, experts and novices users, and user population groups (academia, government, and industry).

First, interdisciplinary, collaborative scientific work is conducted in the Project and archived in the Database. Liquid crystal science is inherently interdisciplinary. The information specialists and researchers collaborate to represent the structure of interdisciplinary basic and applied science as conducted in research centers like ALCOM. Physicists, chemists, and mathematicians in the project contribute experimental and theoretical research results from their respective disciplines. The Database retains the disciplinary orientation of each field and acts as a catalyst to interconnect shared concepts across disciplinary boundaries. Classification structures from physics, chemistry, and mathematics broadly organize the database contents.

Second, the goal of the Project and the Center emphasizes basic and applied research. The research is more narrowly defined by terms from the Heterogeneous Structures Thesaurus which arise from the specific questions asked within the context of the project. The thesaurus is dynamic and tries to represent the mutual interests of academe, government, and industry and the practical consequences of their research.

Third, given the interdisciplinary environment of the Project and the Center, scientists must act as experts within their disciplines as well as outside their disciplines. Researchers represent experts and novices with distinct information preferences in the interdisciplinary framework of the Project. The Database and Thesaurus are designed to supplement the scientists' familiarity of technical terminology and organizational structures so that researchers...
can function as experts in each of the disciplines represented. Tracking mechanisms explore
the problem solving skills used by experts in their disciplines and compare their ability to
transfer these skills when working as novices outside their discipline. The scientists also
represent different user groups in academia, government, and industry. A comparison of
information preferences for each of the user populations documents the relation of relevant
information to university scientists, government researchers, and industrial partners in
scientific work.

Fourth, academia, government, and industry represent three user populations with
distinct information preferences. Preliminary usage data suggest repeated patterns by the three
types of researchers. Academic researchers have demonstrated greatest interest in recent
published papers. The group also tend to hold database contributions until works are ready for
publication. Industrial partners choose experimental results and contribute to the body of
patent literature. Government participants provide modeling and simulation results which
support industrial applications.

4. Conceptual Representation of Liquid Crystal Research

The authors have transferred these characteristics into the expansion of a previously
existing liquid crystal thesaurus for the Heterogeneous Structures Database. The Thesaurus for
Liquid Crystal Research and Applications was developed in 1993 by Maja Zumer in
collaboration with physicist Slobodan Zumer, a noted liquid crystal. The Thesaurus
(http://www.personal.kent.edu/~slisl/zeng/maja2.html) provides an alphabetic and hierarchical
display of liquid crystal terms based upon a broad conceptual representation of research in the
field. The newly developed Heterogeneous Structures Thesaurus in the Database bridges the
broad representation of the Liquid Crystal Thesaurus and the problem-centered focus of the
Heterogeneous Structures Project. Broad concepts from the Liquid Crystal Thesaurus, such as
«orientational ordering» gain greater relevance to project researchers when combined with
natural language terms drawn from project research, such as «polymerization induced phase
separation (PIPS).»

The Thesaurus for Liquid Crystal Research and Applications was developed in 1993 to
provide systematic overview of current terminology. It was based upon a broad conceptual
representation of research in the field with no particular application in mind.

It was designed as a faceted thesaurus to cover several aspects of liquid crystal research
(both basic and applied): phases of liquid crystals, theoretical models, ordering, experimental
methods, textures, dynamics, phase transitions, surface phenomena, and applications. The
choice of facets and the depth of coverage within the facets is of course to a certain extend
subjective, showing an individual scientific view. On the other hand there is a generally
accepted terminology of the field, the one that is given in the important textbooks. That
terminology was followed throughout the Thesaurus.

The terminology used was based on expert advice and textbooks on liquid crystals.

The hierarchical display is the one usually consulted because it is compact, easy to
understand intuitively and gives a good overview. It is only the alphabetical display though
which shows the whole functionality of this Thesaurus as a tool. It is much richer, because it
includes both the equivalence (synonyms) and associative relationships. With the use of
related terms important connections within and among facets are made. For instance specific
phases of liquid crystals are connected with appropriate theories that describe them. It is this
kind of connections that guide the user even further into the understanding of the field beyond
a simple choice of appropriate terms.

A thesaurus in general is a very powerful general tool: it can be used on its own as a

Advances in Knowledge Organization, Vol.6(1998)
means of standardising the terminology and at the same time an important tool for information retrieval. If the search engine supports both browsing and display of the terms in the thesaurus and extends the searching to include both hierarchical and associative relationships (on demand) into the query, the thesaurus can be used to its full potential.

The Thesaurus for Liquid Crystal Research and Applications was meant to be a starting point for more specific, in-depth vocabulary of narrower, more specialized fields of research. Two approaches are possible: to extend the existing thesaurus by adding new facets and/or add additional terms, or to build a separate thesaurus or dictionary only loosely based on the existing one. The latter was chosen for the HSP Database.

The HSP is a hierarchical dictionary covering a specific area of liquid crystal research. Some facets are given in more detail, the others are omitted. The terminology chosen reflects the usage in the published research within the project.

5. General Framework for Thesaurus Building

Our present work involves integrating the Thesauri for Liquid Crystal Research and Applications and Heterogeneous Structures Project Database. The assumptions include:

- Design a thesaurus structure based on what is relevant for users and generators of information within a specific subject domain.
- Organize representation of knowledge or classification as closely as possible to the authors’ and users’ terminology of subject domain.
- Therefore, thesaurus is meeting point of generators’ and users’ models of subject domain and information needs by representing user needs and information content.
- Such a design facilitates communication and information transfer between generators and users.
- Builds upon a close relationship among work, research text, researchers, and systems.
- Structure is conceptual organization of particular subject domain based on generators’ texts and users information needs.
- Subject domain includes compendium of knowledge and tasks, jobs, as well as functions.
- Thesaurus as intermediary structure between generators and users of information within IR system.
- Knowledge, user intentionality, own purpose are critical elements in developing effective thesaurus.

6. Methodology for Thesaurus Building

- Develop single concept terms as the unit of analysis from handbooks, encyclopedias, on liquid crystals.
- Collect and compare identified terms from generators’ writings.
- Determine number of occurrence of terms in text to chart relevance.
- Identify terms from users’ research profiles with query statements.
- Validity of suggested terms and categories come from outside texts recognized by users and generators.
- Therefore, criteria for organizing relevant to users, that is close to their model of the subject field and their work.
References: