Unified Chemical Research Interface: A New Concept of Information Exchange via the Internet

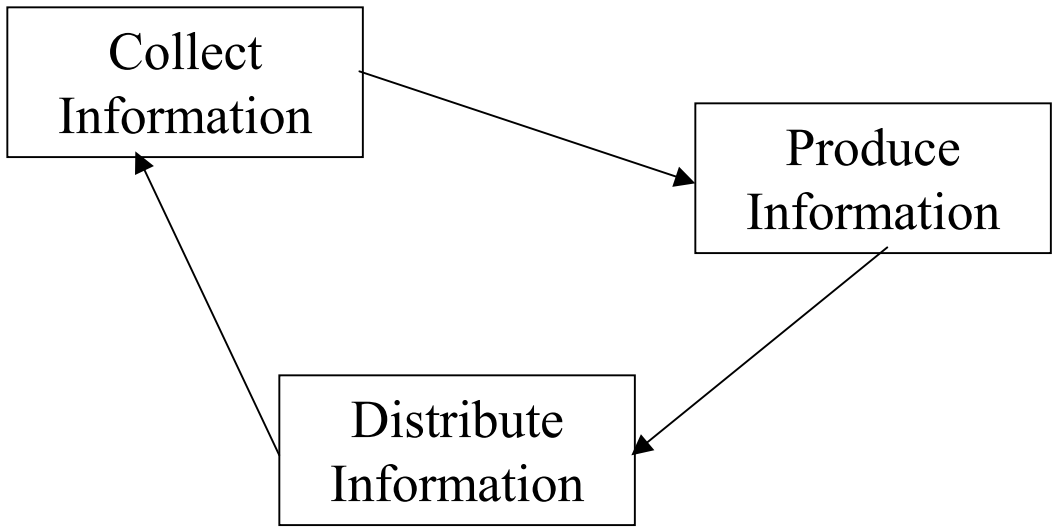
YAN HE, Department of Chemistry, University of Iowa, Iowa City, IA 52242. Email: yan-he@uiowa.edu

Abstract

Successful chemical research requires fast and efficient research information exchange, which includes information collection, production, and distribution. Current information exchange mode is publication oriented, which separates information producer and consumer from information manager and distributor, resulting in inefficient information exchange and journal crisis. To solve this problem, the concept of Unified Chemical Research Interface (UCRI) is proposed. UCRI allows researchers to build their own information management system and exchange both information and information connections through a standard but flexible mechanism via the Internet. Along with the existing information archival system, UCRI could potentially enhance both the rate and quality of research information exchange, and promote multilevel cooperation in the research community.

Information Crisis in the Information Age

- Chemical research, or any scientific research can be viewed as an everlasting cycle of information flow:



- Currently, chemical research cycles involve four major parties:
  - Research group:
  - Journal publisher & librarian
  - Database provider
  - Chemical and equipment supplier

- Journal crisis:
  - Researchers publish more articles, with more pages per article, but with less time to read each article
  - Journal publishers launch more issues and series, provide more printed and online services
  - Result: journal prices increase but the number of subscriptions decrease
  - Similar or potential information crisis:
    - Database provider: database crisis
    - Chemical and equipment supplier: catalog and product crisis
- Why?
  - Finger-pointing game
    - publishers charge for too much, libraries have too little budget, researchers publish too many articles, etc.
  - In-depth: reflection of information explosion
    - Research complication and specialization require increased information exchange, and the traditional publication centralized mode limits the rate and efficiency of information transportation

Bottleneck: Journal Articles

- Format and content of journal articles
  - Rigid format largely unchanged for over 100 years
    - Five level hierarchy
      - keywords, title, abstract, full text, reference list
    - Similar modules
      - introduction, theory, experimental procedure, results and discussion, conclusion
  - Intrinsically flexible content
    - Each article is a logic presentation of a finite set of scientific information, which could come from various areas
    - Each article is a bridge between previous knowledge and new findings, which could appear at any direction
  - The contradiction
    - Information has to be presented using certain format
    - No format can satisfy the the need of all information presentation

- Format-content separation: speed information exchange
  - Standardized format facilitates scientific report
  - Different articles can be indexed using certain criteria by librarians and publishers, sparing the researchers' effort
- Format-content separation: impede information exchange
  - Information and information relationship presented in the article do not necessary reflect all the research effort by the authors
  - Abstract and tile can not summarize all the information provided, sometimes even does not offer a clear picture of the full text
  - Keyword set and reference list are apparently structureless, though they form certain internal relationship
  - Existing centralized indexing systems archive information using a fixed tree-like structure, which can not meet the individual requirement of researchers
  - Building connections between format and content requires extended and repeated effort from researches, resulting in information loss, information bias, and inefficient information flow

Publication Oriented Information Exchange

- Collect information: literature search
  - Readers must fit their targets to the library and the database, e.g, form certain keywords combination and searching strategy
  - Readers form different understanding to the same article
  - Amount and efficiency of information retrieved depend on one's knowledge base, time and financial situations, and curiosity
- Produce information: research process
  - Difficult to find the specific information required, rediscovering is sometimes easier than information retrieval from existing literature
  - Difficult to justify the problem attacking strategy
  - "How do I know what I do not know?"
- Distribute information: result report
  - Organize information obtained to publishable format, resulting in some information are filtered out, and some information is split into multiple publications

Unified Chemical Research Interface (UCRI)

- Summery of chemical information crisis
  - All chemical information can be classified as data and intra- or inter- connections of data. But centralized information archival systems categorize large amount of data into limited and relatively fixed frames, while losing the dynamic connections among them. As a result, a significant portion of chemical research is devoted to rebuilding those connections repeatedly.
- The solution
  - An information transportation mechanism that allows both data and connections of data to be distributed to the community.
  - Instead of tailoring all the required or produced information to certain centralized frames, researchers will define (at least part of) their own connections of information, maintain their own database dynamically, and exchange information directly via the Internet.

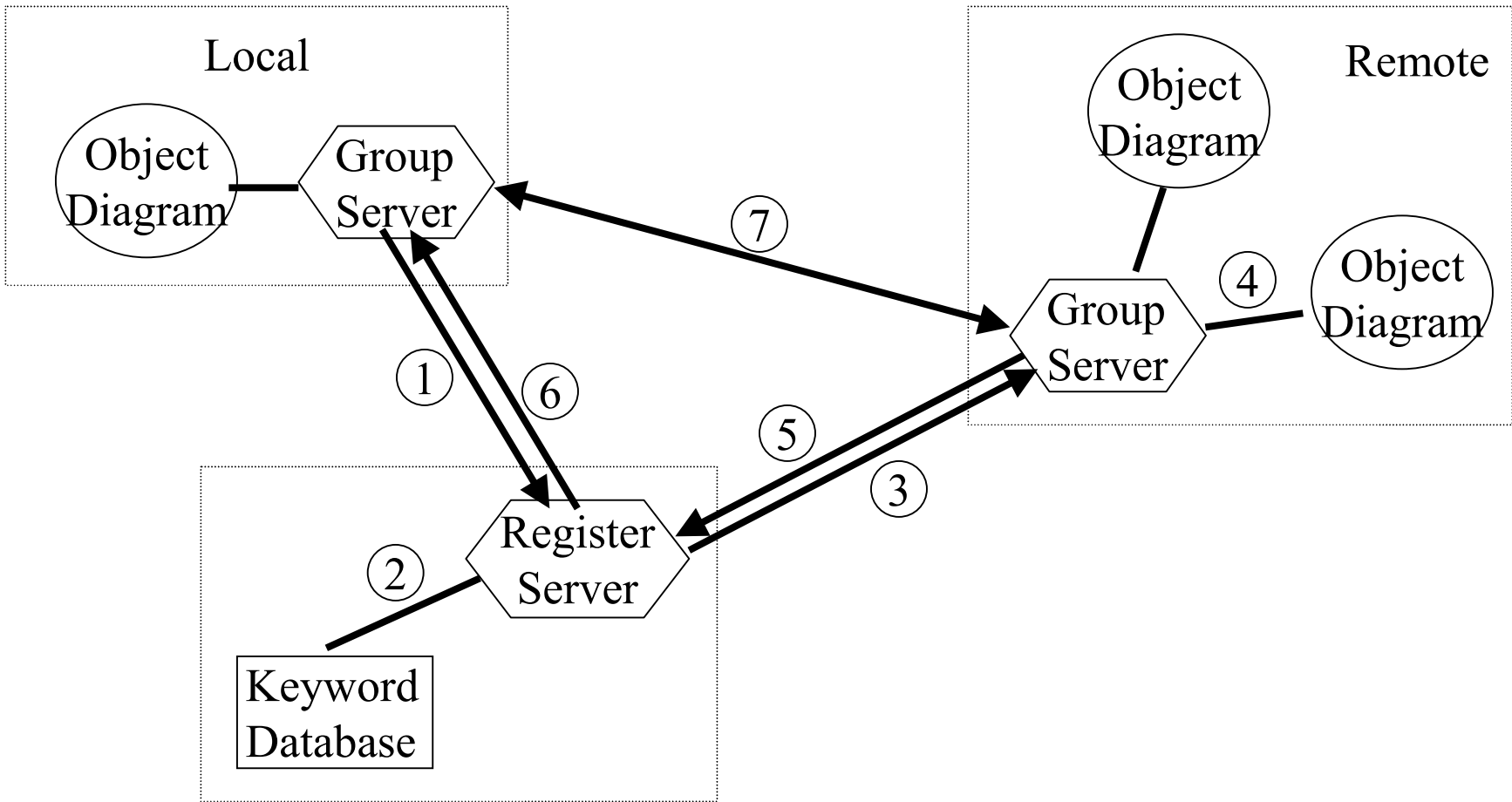
Conceptual Structure

- Data object: similar to a small article but with only one clearly defined concept and its connections with other concepts
  - Identity
    - Id, name, author & owner, date of creation and completion
  - Security: accessibility to visitors
    - Public, general researcher, cooperator, group member, etc.
  - Description:
    - Theory, substance, instrument, procedure, experimental results, etc.
  - Connections:
    - Reference &citation
    - Links to related data objects
    - Links to other research groups
    - Links to free or commercial databases
  - Management operators:
    - Change status or security, add, delete, split, and merge links, etc.

- Object diagram: 2D- or 3D- graphic representation of intra- and inter-connections of data objects
  - Different projects and the expansion of each project in each research group generally follow some scientific logic, which can be illustrated by a finite set of data objects and their connections
  - Project diagram:
    - Functional relationship of data objects, indicating what projects are developed in the research group
    - Sub-diagram: derived from certain fields of data objects
      - keyword diagram, chemical diagram, theory diagram, instrumentation diagram, reaction diagram, procedure diagram, reference diagram, publication diagram, etc.
  - History diagram
    - Temporal relationship of data diagrams, indicating how and why certain projects are developed in the research group
  - Both project and history diagrams vary with research progress.

- Group server: administrator of information within the group and gateway of information exchange with the community
  - Possess a unique group identifier
  - Maintain information
    - data objects, object diagrams
    - local databases of keywords, chemicals, references, etc.
  - Release information
    - Submit releasable information, i.e., all or part of the keywords, to the register server, and modify or withdraw registrations.
  - Accept and process information request
    - Perform structural search of object diagrams
    - Provide BBS or inter-group email service
    - Generate different views of object diagrams
    - Maintain visiting records
  - Request Information

- Register server: transfer station of information flow
  - Provide registration service of keywords, chemicals, references, etc.
  - Accept, process, redirect initial information requests



The Mechanism

- Step1: The local group server accepts a request form a client , i.e., a certain combination of keywords, and submits it to the register server.
- Step2: The register server searches its keyword database and find the remote group servers that have registered these keywords.
- Step3: The request and the signature of the local group server sending the request are directed to the remote group servers.
- Step4: Based on the signature of the request, the remote server assigns a security level to it, and performs a structural search of the project diagram.
- Step5: If a match of both the security and keyword combination is found, a command for the request is issued and sent to the register server.
- Step6: The addresses of the remote group servers that grant the request and the resulting commands are returned to the local server.
- Step7: The local group server sends the special command and subsequent requests of the clients directly to the remote group server, and the remote server generates and returns corresponding views of the object diagrams.

- Technical feasibility
  - Concept similar to Unified Modeling Language (UML)
  - Scale down of B2B technology and chemical information service
    - Indexing, internal database management, structural search, group-group communication, customized view, etc.
- Standardization & flexibility
  - Unified mechanism of information exchange and management, saving individual's effort in designing and advertising web sites
  - Flexible information content, structure, relationship, accessibility, and visibility
- Security & publicity
  - Individual research group has the full control of what, how and how much information will be released to the community
  - Anonymous requests are denied to prevent irresponsible acquisition and usage of information
  - The "black box" in research processes can be gradually eliminated

Researcher-oriented information exchange

- Collect Information
  - Keyword search not only returns matching hits, but also points out the research group and processes that produces these results, the consequence is more information in smaller number of hits.
  - Solving dilemma: without understanding an author's work, one does not know whether to read the author's articles; but without reading his(her) articles, one can not understand his(her) work.
- Produce Information
  - Specific information required can be easily found, and initial hypothesis can be easily verified, avoiding rediscovery
  - More innovations and new ideas could come up by studying the research process and information management of others
- Report Information
  - Results can be presented and distributed in a logical manner consistent with one's research process

Win-win situation

- Save resource & promote cooperation
  - Expansion of information sources and speeding up of information flow will save both time and money in the research process
  - Increase information exchange will increase mutual understanding and promote multilevel cooperation in the research community
- Solve journal crisis and other information crisis
  - UCRI will not replace existing journals and archiving system, because it can only serve as a guidance of individual's research, detailed reports still need to be presented as a publication.
  - Researchers will read more articles , because they will need less time and efforts to find and understand the information they need
  - Less number of articles need to be published, because researchers have more channels to present their achievements
  - The result: journal prices decrease and # of subscriptions increase
  - Similarly, database crisis and product crisis will be solved.

- Huge market
  - Developer can gain large profits by managing register servers, providing training and other services, publishing patterns, etc.
  - UCRI can further become United Scientific Research Interface (USRI), which will have a huge user base if being adapted by the scientific research community.
- Economic and social catalyst
  - Large amount of information transportation will increase the demand for IT products, i.e., hardware, software, services, which could potentially revive the currently weak IT industry.
  - Scientific research process and result become more understandable to the general public, which will increase their support to science.
  - The general concept, i.e., building personalized but consistent functional and historical information relationship around individuals and making them available through a unified mechanism via the Internet, can be expanded to other economic and social areas.