CONF-971118--1
RECEIVED
MAR 25 1997
OSTI

An In-House Alternative to Traditional SDI Services at Argonne National Laboratory

ANL/IPD/CP--92625

Robert E. Noel
Technical Librarian

Richard R. Dominiak
Systems Analyst

Argonne National Laboratory

February 20, 1997

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under Contract No. W-31-109-ENG-38 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, non exclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

An In-House Alternative to Traditional SDI Services at Argonne National Laboratory

Abstract: Selective Dissemination of Information (SDIs) are based on automated, well-defined programs that regularly produce precise, relevant bibliographic information. Librarians have typically turned to information vendors such as Dialog or STN International to design and implement these searches for their users in business, academia, and the science community. Because Argonne National Laboratory (ANL) purchases the Institute for Scientific Information (ISI) Current Contents tapes (all subject areas excluding Humanities), ANL scientists enjoy the benefit of in-house developments with BASISplus software programming and no longer need to turn to outside companies for reliable SDI service. The database and its customized services are known as ACCESS (Argonne Current Contents Electronic Search Service). Through collaboration with librarians on Boolean logic and selection of terms, users can now design their own personal profiles to comb the new data, thereby avoiding service fees from outside providers. Based on the feedback from scientists, it seems that this new service can help transform the ANL distributed libraries into more efficient "central functioning" entities that better serve the users. One goal is to eliminate the routing of paper copies of many new journal issues to different library locations for users to browse; instead users may be expected to rely more on electronic dissemination of both table of contents and customized SDIs for new scientific and technical information.

Keywords: BASISplus, SDI, Selective Dissemination of Information, FQM applications, personal information profile, Current Contents, TECHLIBplus

Introduction

Argonne National Laboratory (ANL) is a federal research and development laboratory with more than 200 programs in basic and applied sciences. The ANL staff comprises some 1400 PhD scientists and engineers, together with an additional 3000 technicians, graduate students, visiting faculty, and support staff. Information services are integral to the research process, and access to the most recent scientific and technical literature available in the 10 subject-based distributed libraries is essential. An overall objective of ANL's Technical Information Services and libraries is to maximize the ease and effectiveness of information retrieval for all users. The implementation of ACCESS (Argonne Current Contents Electronic Search Service) and subsequent programming enhancements have kept scientists and engineers informed with specific current awareness searches at a time when information overload seems rampant.

Several factors make Current Contents an attractive current awareness package for Argonne. It covers over 6,500 titles in the sciences and social sciences, is international in scope, and it has recently included some books in addition to journal titles. In the Feb/Mar 1995 issue of DATABASE, Jaguszewski and Kempf compared the coverage and currency of four current awareness databases: CARL UnCover, ContentsFirst, Inside Information, and Current Contents on Diskette. A few of the limitations they point out with Current Contents are a result of receiving the information on separate diskettes - many of these problems have been overcome because Argonne purchases single, comprehensive tapes of data. For example, it is not necessary for the personal profiles to be run repetitively against separate sections of the database (i.e. Life Science, Engineering, Technology and Applied Sciences, etc.). It is also not the case that "journal titles are searchable only by abbreviation"; further, abstracts are searchable (in fact, Current Contents stands out as the only package that includes abstracts), and the inclusion of journal titles in the database is based on evaluation of content and format by journal experts and statistical analyses of the impact and use of material published.[1]

Electronic Table of Contents and SDI Services

Argonne began its subscription to Current Contents in 1993. Users were provided with all bibliographic data and a BASISplus system capable of searching by keywords, author, and journal title. All

[1] Jaguszewski, JM; Kempf, JL, "Four Current Awareness Databases: Coverage and Currency Compared", **DATABASE**, Feb/Mar 1995, p. 36. This method arguably has a downside; nevertheless, it is what ISI uses to decide what is worth indexing and what is not.

programming was done using FQM. The first in-house enhancement was a table of contents e-mail delivery service that allowed users to identify useful journal titles and receive the table of contents of new issues. Free delivery of journal titles has long been available via various publishers' web pages, but this route requires users to first research where important titles reside and then sign up, a much more tedious, time-consuming, and limiting process. One advantage of Argonne's system at the time was that it was a more comprehensive, one-stop service.

The ACCESS SDI service at ANL began in early 1996. The searches themselves are designed around user-identified words and phrases (i.e. interest profiles), they automatically run on a regular basis, and are delivered by e-mail.[2] It was expected that the circulation statistics of new journals would increase after the service became available. A study comparing earlier years with 1996 circulation data revealed that the number of times a new journal was requested or checked out to a user remained relatively stable despite the existence of the new SDI service.[3] This confirmed that many staff simply use and photocopy journals in the library; while they could make an electronic request for the relevant journal to be mailed to them, it seems that most staff, once notified of a relevant article by SDI, simply walk to the library and copy it. It is more difficult to tally journal activity when information is accessed by this very straightforward, nonelectronic way. It may also be the case that the SDI service has not been fully embraced by users, and it is therefore too early to gather statistically significant data about its effects.

New journals have traditionally been routed to the distributed libraries with the assumption that users browse new issues for relevant material on an as needed basis. It is difficult to measure the payoff for scientists when journals are routed, however, some comparative observations can be made about the two methods of information distribution (e-mail SDIs vs. journal issue routing).

SDIs		**Routing Journal Issue**	
Pros	Cons	Pros	Cons
Immediate results to multiple users.	Only capable of delivering bibliographic information.	All article information available; easy scanning & evaluation of content.	Available to only one user at a time.
Reduces overall burden on library staff.	Requires librarians to learn new syntax to create SDIs; then both identify and correct errors in profile design.	Preserves status quo; users expect and rely on existing system.	Routine journal circulation requires daily attention with uncertain benefits to users.
Automated filters (SDI) exponentially increase amount of information scanned for relevancy.	Rigid keyword / keyphrase limitations; cannot interpret information.	User evaluates and interprets information directly - judgement on relevancy not constrained. (Serendipitous finds.)	Neither cost-effective nor feasible for users to scan the super-abundance of new information.
Users may append or attach results to existing private bibliographic databases.			New journals end up in the mail instead of on display for use.

How well does the ANL's ACCESS system do what it is designed to do, and is the system serving scientists and the institution better than outside commercial sources? According to Yan and Garcia-Molina, an SDI service should meet the following requirements:

[2]The role of the librarian may be shifting. Consider the following 1987 definition of an SDI where it is suggested that the librarian conducts the searches, then distributes them: "Selective dissemination of information: Regular use of current awareness searches to provide persons with the most current references to documents that match their interest profiles." Wert, Lucille M., **Encyclopedia of Physical Science and Technology**, from glossary preceding *Scientific Information Services* article, Vol. 12, p. 450.

[3] We used the first 60 days of its availability in the library as the definition of "new" journal.

1. It should allow a rich class of queries as profiles, unlike Netnews.
2. It should be able to evaluate profiles continuously and notify the user as soon as a relevant document arrives, not periodically.
3. It should scale to a very large number of profiles and a large volume of new documents.
4. It should efficiently and reliably distribute the documents to subscribers.[4]

ACCESS meets the first requirement fairly well. The index structure for Current Contents information allows for a variety of gathering techniques. Users are not limited to keyword/keyphrase Boolean searches that scan only titles and abstracts; for example, a profile may be designed to find any new articles by 10 noteworthy physicists on synchrotron radiation, or another could be designed to find any article on ceramics from Oak Ridge National Laboratory. Any combination of bibliographic fields may be included in the SDIs, and Boolean combinations of full or truncated keywords and phrases may be included. Proximity connectors and phrase searching are also available for use in SDIs (e.g., find "steam generators" within 4 words of "fracture" or "tubing"). Phrase searching is unique and a key advantage of the BASISplus software.

Profiles are run against Current Contents data only once a week; although that does not satisfy requirement 2, it is more frequent than many commercial databases available through Dialog and STN, and users are escaping some fees. For example, STN can provide a daily SDI in its CA Plus (Chem Abstracts) database, but there is a $1.30 execution fee, in addition to variable fees based on number of results and format of the deliverable document.

"It should scale to a very large number of profiles and a large volume of new documents." This is not explicitly defined, but ACCESS has been able to accommodate several hundred users over a short period of time. Perhaps more importantly, the system will shortly be available to faculty at the University of Chicago, and discussions continue as to how we may make it available to Northwestern University faculty as well. The number of new table of contents scanned each week varies; on average, 600-800 new journal issues are contained on the weekly tapes.

Regarding requirement 4, ACCESS does efficiently and reliably distribute documents to users. "A *document* consists of a collection of *words*. The set of words that can appear in documents form the vocabulary..."[5] While ACCESS has had few failures and problems in its delivery of SDIs as documents (the bibliographic citation and abstract), a more fundamental question is 'how well do ANL's libraries deliver actual documents (articles from journals) based on user response to SDIs?' In other words, unless the user can expect a relatively short turn-around time between his request for an article (solely a result of receiving an SDI) and delivery of that article, the SDIs can be more like frustrating enticements than a value-added information service.

Implementing the Service

There are currently more than 500 SDIs running at Argonne. Initially a template was created and made available to the librarians so the profiles could be accessed, edited, modified, copied, or deleted. Because the librarians were not familiar with FIND commands, they decided to use the comment field to translate the FIND command into the more familiar Dialog syntax. This turned out to be a good way to share ideas on how to set up complex Boolean searches, and it helped minimize the amount of technical information the librarian would have to have mastered were the profiles inaccessible. Below is a sample search where the user is tracking three terms.

---

[4] Yan, Tak W., Garcia-Molina, Hector, "Index Structures for Selective Dissemination of Information Under the Boolean Model", **ACM Transactions on Database Systems**, Vol. 19, 2, June 1994, pp 332-364 (333).
[5] Ibid. Yan and Garcia-Molina, p. 335.

```
+------------------------------ SDI Profile ----------------------------------+
| Profile: 7                    Title: micromechanical or micromachin*
|| Patron ID: B45042           Email Addr: sanjay_ahuja@qmgate.anl.gov
|| Library: 212                                                  Abstract? Y
||
   FIND journal,article WHERE journal.ga JOIN article.ga AND
|| article.words phrase any words
|| 'micromechanical micromachin* microfabricat*'
||
||
||
|| Comment: s micromechanical or micromachin? or microfabricat?
|| Add Date: 04/29/96 Add UID: DOMINIAK  Rev Date: 07/16/96 Rev UID: RNOEL
|+-----------------------------------------------------------------------------+
Action     > (Add,Delete,Exit,Find,Help,Match,Print,Replace,Show,>,<,?)
Member#    11          of 93 for Set 1      (For Delete, Replace and Show)
```

The "phrase any" connector enables the above search to find any article that contains any of the three words listed. Another option that would produce identical results would be to use the INCludes connector, i.e. "article.words INC 'micromechanical','micromachin'*,'microfabricat'*. The comma functions as "or"; ampersand is "and". This style of SDI is common; it contains one or a few words or phrases and keeps the user informed as new articles are published on the topic(s).

The following is more complex, and illustrates FQM's flexibility and precision.

```
+------------------------------ SDI Profile ----------------------------------+|
Profile: 194                  Title: Brillouin Scattering
|| Patron ID: B49147          Email Addr: lli@anl.gov
   Library: 223                                                  Abstract? Y
|| FIND journal,article WHERE journal.ga JOIN article.ga AND
|| (article.words PH ALL WORDS
|| 'elastic propert*','elastic pressure*','sound velocity'
|| AND article.words PH ALL WORDS 'high temperature*','high
|| pressure*','superlattice*','thin film*') OR (article.words PH ALL WORDS
|| 'elastic propert*','elastic pressure*','sound velocity' AND
|| article.words INC 'irradiation','amorphous','glass','glasses','glassy') OR
|| (article.words PH ALL WORDS 'brillouin scattering' AND NOT article.words PH
|| ALL WORDS 'amplif*','stimulat*','phase conjugat*')
||
||Comment:
||
|| Add Date: 06/19/96 Add UID: LLI       Rev Date: 07/05/96 Rev UID: LLI
|+-----------------------------------------------------------------------------+
Action     > (Add,Delete,Exit,Find,Help,Match,Print,Replace,Show,>,<,?)
Member#    28          of 29 for Set 1      (For Delete, Replace and Show)
```

Conclusion

The SDI service at ANL continues to grow, and users have consistently given positive comments about its usefulness. Although the Current Contents information cannot replace other more targeted or comprehensive files, it nevertheless has proven to be both cost-effective and beneficial to users. It is perhaps too early to detect widespread changes in user behavior and library system efficiency based on the availability of in-house SDIs. Nevertheless, these kinds of enhancements can be made to similar BASISplus systems at companies and other national laboratories, and can help organizations become more self-sufficient at information distribution.