

WORLD ALUMINUM ABSTRACTS COMPUTER RETRIEVAL USING QL
SYSTEMS USER COMMENTARY AND DEMONSTRATION
(RECUPERATION PAR L'ORDINATEUR DE WAA: COMMENTAIRES
ET DEMONSTRATION D'UTILISATEUR)

Audrey E. Rushbrook
Alcan Research Centre
Kingston, Ontario
K7L 4Z4

Hugh Lawford, President
QL Systems Ltd.
322 Brock Street
Kingston, Ontario
K7L 1S9

ABSTRACT

Since its inception in 1968 World Aluminum Abstracts has been put on computer-readable magnetic tape. At first only the citation and index headings were recorded, but since January 1972 the complete abstracts have been included. Using the QL System for retrieval the entire citation, abstract and index terms can be searched using free language. Miss Rushbrook comments on the usefulness of this information retrieval system from the user's point of view, and Mr. Lawford comments on QL Systems' method of retrieval and QL's various other data banks available. The system will be demonstrated on a computer terminal. (Depuis le commencement de la publication de WAA en 1968 les résumés sont été enregistrés sur bandes magnétiques accessibles par l'ordinateur. Seulement le titre et le répertoire des domaines étaient enregistrés au début, mais depuis janvier 1972 on a commencé l'enregistrement complet des résumés. Le système QL peut être utilisé pour récupérer le titre entier, le résumé et le répertoire des domaines par

WAA AND THE QL SYSTEM

voie de la banque naturelle. Mlle. Rushbrook fait une analyse de l'utilité du système QL de point de vue d'un utilisateur, pendant que M. Lawford fait le commentaire du système QL, comme une méthode de dépistage automatique de l'information. D'autre part il passe en revue d'autres banques de données QL disponibles. Une démonstration du système QL avec un terminal sera présentée.)

USER COMMENTARY AND DEMONSTRATION

The answer to the "information explosion" seems to be to find means to pinpoint information quickly and effectively from the huge mass of written words.

The ideal way to do this is to search the complete printed document. And this may come in time, as more text of journals and books is put directly onto machine-readable tape for printing purposes. But this requires a sophistication of hardware and software, and a backlog of material prepared in this way, that has not yet been achieved to make it become effective and economically viable.

Preparation of machine-readable tapes by abstracting services began in the sixties, so the earliest data banks go back to only 1966 or 1968, with a few exceptions.

World Aluminum Abstracts was developed by a subcommittee of the Aluminum Association consisting of five members from four of the large aluminum companies. This committee consisted of Dr. L. W. Eastwood of Alcoa, Chairman, and F. F. Dietsch, Reynolds Metals Co., W. B. Frank, Alcoa Technical Centre, G. Gerard of Kaiser Aluminum & Chemical Corporation, and I. H. Jenks of Aluminium Laboratories Limited (now Alcan Kingston Research Centre). The first issue was published in March 1968 and was called Aluminum Technical Information Service. From its inception, the references were put on machine-readable tape, and from the beginning of 1972 the complete abstract was also included.

WAA AND THE QL SYSTEM

The abstracts and the tapes are prepared by the American Society for Metals following guidelines laid down by the Aluminum Association subcommittee. Each of the companies represented had published in-house abstract bulletins for many years. Alcan's "Abstract Bulletin of Aluminium Laboratories Limited" was begun in 1929 and was the earliest of these. From these various experiences, the subcommittee pooled their resources and were able to say exactly what was required in the abstracts and what was required in the indexing. A thesaurus was developed which was a composite of this experience, and was designed with computer retrieval in view.

This original groundwork has made the WAA data base a very useful tool to search the literature in the aluminum and light metal field. It was designed by the users and is continuously being further developed by the users through the Aluminum Association subcommittee.

It was only fairly recently that these tapes have been used for searching. In 1972, Alcan obtained a copy of one month's tape and looked into the possibility of developing software to search it. After considerable discussions and trials, Alcan was still undecided whether to proceed with its own in-house development. At this time, the summer of 1974, QL Systems offered to put the tape up on their system. This they did, and their searching method proved good and useful, and as use increases, further improvements are being made.

Now, WAA can be searched by any QL users for the cost of one dollar to sign on and a dollar per search. Recently WAA has become available on the Lockheed "Orbit" System as well, where it is searched by their software methods and follows their costing methods.

QL's system for searching is very easy to use, and has the advantage of searching not only the title and index terms, but the complete text as well. It is an on-line or "interactive" system. The user's terminal is connected by telephone to QL's computer facility in Ottawa. Each user has a unique user identification code that allows access to the system.

WAA AND THE QL SYSTEM

Every search service, including QL Systems, has its own language which consists of statements and commands which the user keys into his terminal. He formulates his request by using Boolean logic (AND, OR, AND NOT) to connect terms of interest to him. He can then look at the material thus retrieved and can have items printed, specifying the desired format.

The need for information comes in many variants. A search for articles by a certain author is very simple, and so is a search for a single subject concept. It is when several parameters of a subject that cannot be described in exact language are required that searching becomes a detective and skilled art. As the number and size of data banks increases, it becomes increasingly apparent that only the person having a substantial, even expert, knowledge of the field, a knowledge of the data banks to be searched, and the commands and intricacies of the search service used, can produce the optimum results.

The following examples illustrate different kinds of questions and information needs. The terms and strategies used for the QL System show how to locate the documents containing the information.

SEARCH: Is there anything on the fracture toughness of U.S. wrought alloy 2048?

TERMS: In selecting terms to be searched, look first for terms that are unique and necessary. In this case, the specific wrought alloy, 2048, meets these requirements. But this alloy may be expressed in many ways. It can be expressed by its composition, either by its major elemental components or by the specific amounts of each component. It can be expressed by the number 2048, or it might have the Aluminum Association designation AA in front of it. The first step then is to write down all the possibilities. Checking the "Registration Record of International Alloy Designations and Chemical Composition Limits for Wrought Aluminum Alloys" published by the Aluminum Association, we find that there is an X2048 as well as 2048, and the composition is given as 2.8-3.8% Cu, 1.2-1.8% Mg, 0.20-0.6% Mn, 0.25% Zn, 0.20% Fe, 0.15% Si and 0.10% Ti, other 0.15% max. and remainder Al.

WAA AND THE QL SYSTEM

From this composition, the additional terms Al-Cu-Mg and Al-Cu-Mg-Mn should be searched. (The X represents an experimental alloy; the X is dropped when the alloy has full registration.)

The second concept to be searched is fracture toughness. This can be done either by linking fracture and toughness with the boolean 'and' connector, &, or, since the words are usually found side by side and in the same order, they can be more precisely linked as a phrase using quotation marks, as "fracture toughness".

STRATEGY: Our terms have now been chosen, so now we choose how to link them together. Since the QL System supports searches for a phrase and one of several words, our search strategy becomes

"fracture toughness" & 2048 X2048 Al-Cu-Mg Al-Cu-Mg-Mn

This strategy results in presenting us with all articles in the data bank which contain the phrase "fracture toughness" and 2048 or X2048 or Al-Cu-Mg or Al-Cu-Mg-Mn.

To test this strategy, we can call up the term or dictionary entries by typing DICT and the term, e.g. DICT 2048. This will give the term and how many times it occurs, and in how many documents. The next entries follow immediately afterwards. The results received for the various terms were:

	OCCURS	DOCUMENTS
Dict 2048	6	3
2048-T851	2	2
20488	2	2
DICT' X2048	16	10
X2048-T851	1	1
X224.J	3	3
DICT' Al-Cu-Mg	129	105
Al-Cu-Mg-Ag	2	2
Al-Cu-Mg-Fe	1	1
Al-Cu-Mg-Mn	12	8

WAA AND THE QL SYSTEM

These results show that we should add 2048-T851 and X2048-T851 to our strategy. (The '-T851' is a temper designation for an alloy.) We could use an asterisk for a truncated ending, e.g. 2048* and X2048*, but then the term 20488 would also be picked up. However, if there is a possibility that 20488 is a typing error for 2048, then the truncated ending would be the better choice.

We are now ready to type S, and after the response ENTER QUERY type the terms

"fracture toughness" & 2048* X2048* Al-Cu-Mg Al-Cu-Mg-Mn

the computer responds with

YOUR SEARCH IS PROCEEDING

then

1 DOCNUMBER	2 TITLE	3 AUTHORS	4 SOURCE
5 ABSTRACT	6 KEYWORDS		

If you want all, then just press the carriage return. The first document will be presented. The carriage return must be used at the end of each document for the next one to be presented. When all the documents found by the query have been shown, e.g. 15 documents, the computer responds with

SORRY, MAXIMUM RANK IS 15;

PLEASE ENTER NEXT REQUEST

This example illustrates the use and simplicity of the QL software system.

The WAA base also contains references to patents and the term "patents" is used each time as a keyword. The following query illustrates how to locate patent references.

WAA AND THE QL SYSTEM

SEARCH: Are there any Canadian patents that deal with aluminum alloys containing magnesium, manganese, and iron?

TERMS:

Mg	magnesium
Mn	manganese
Fe	iron
Al-Fe-Mg-Mn	
patents	
Can	Canadian

STRATEGY: patents & Can Canadian & Al-Fe-Mg-Mn Mg Mn Fe
magnesium manganese iron

This results in seven documents. An examination of these reveals that all are Canadian patents, but only one refers to an aluminum alloy containing iron, magnesium and manganese. If we restate the strategy as

patents & Canadian & magnesium Mg & manganese Mn & iron Fe we find that the one precise document is presented.

Authors of papers have been asked to suggest one or two "propositions" for which their work can be considered to provide supporting evidence. While this paper represents mainly a history and description of a useful information tool rather than the results of a research project, I think it does emphasize an aspect of the new field of computer-assistance in searching for information that is often overlooked. This is, that a data base must be carefully and adequately founded; its format and delineation should be chosen with the abilities of computer-manipulation in view, and the selection of abstractors and their mandates is of paramount importance.

My proposition then is that "Standards for information input into data banks should be evolved from a scrutiny of existing abstracting services and methods".

WAA AND THE QL SYSTEM

SYSTEM OPERATOR COMMENTARY AND DEMONSTRATIONS

As the operator of a system which provides access to the World Aluminum Abstracts data base and to several other data bases, QL Systems Limited has reached the same conclusion reached by Alcan and other users of retrieval systems. The care taken in the design of a data base and in the preparation of the abstracts, keywords and bibliographic data is critical to successful use of the data base within any information system. By these standards, the World Aluminum Abstracts data base has been found to be of the highest quality. The data base has been well planned. Abstracts are prepared promptly. The bibliographic information is well organized and the abstracts are well written. The additional keywords are chosen with care. Not only is the content of the data base good, but the technical standards of its preparation are excellent. The entire data base is surprisingly free of typographic and proofreading errors. The American Society for Metals and its Director of Information, David Chafe, deserve credit for the high quality of this data base.

I would like to supplement the comments made by Miss Rushbrook as a user of the data bases operated by QL and indicate some of the considerations which led us to design the system to work in the way she has described.

QL Systems Limited originated as a research project at Queen's University, the QUIC/LAW Project, to investigate potential uses of computers by lawyers. When the project became too costly for the University to continue, the staff of the Project first offered to donate the project's programs and other work to the Canadian government and, when that offer was refused, set up a private company to carry the research into commercial operation.

The QUIC/LAW Project had placed a strong emphasis on full-text information retrieval of legal documents -- statutes, regulations and judicial decisions -- which might vary in length from half a page to more than one hundred pages. Because of the varying lengths of the documents and the lack of any significant control over the terms used, the

WAA AND THE QL SYSTEM

information system used by the Project and later by QL emphasized automatic indexing of source documents and the use of statistical weighting and ranking of retrieved documents. Many of the source documents were large and the potential sizes of legal data bases were enormous; the West Publishing Company which has adopted QL's retrieval system for a U.S. legal service had an initial test data base of over 1,000,000 documents or over 5 billion characters. So other strong emphases in system design were on reducing the indexes and word locator files required to search such large data bases and on rapid yet inexpensive search techniques. The Project's funds and QL's funds have always been limited. At the same time, a basic objective has been high volume use at the lowest possible cost. Probably because QL's system originated as a university research project, cost effectiveness has been emphasized mainly as a technique for enabling more people to gain access to information more cheaply.

Even while the system still was operated as a research project at Queen's, its use had been extended from law data bases to scientific and technical data bases. The retrieval system was used for searching the PIP or Pollution Information Project files sponsored by the National Research Council and the WATDOC files sponsored by Environment Canada. Later, the PIP files were replaced by Pollution Abstracts, again sponsored by Environment Canada, and Oceanic Abstracts, sponsored by QL itself.

Still later, Environment Canada added the D-REF or Data Reference files (which describe holdings and conditions of access to data collections) and the Selected Water Resources Abstracts data base. Most recently in the environmental area, WATDOC and the Quebec Government have sponsored a data base in the French language references to the St. Lawrence River. The WATDOC system of data bases is based on a remarkably effective network of cooperating and contributing abstractors in federal and provincial agencies, in university and other research centres, and in private industry.

Two data bases concerned with metals technology are being operated by QL -- the World Aluminum

WAA AND THE QL SYSTEM

Abstracts data base described by Miss Rushbrook -- and Metals Abstracts or Metadex.

QL also is sponsoring the operation of the Inform data base, or collection of abstracts from 250 business and management periodicals.

In parallel with the operations of these scientific, technical and business data bases, QL has continued to work with federal and provincial justice departments and with the Canadian Law Information Council to improve the use of the retrieval system with data bases of statutes and judicial decisions. Although the law data bases currently offered are limited to the statutes of Canada, of British Columbia and of New Brunswick and to relatively small collections of decisions of the Supreme Court and Federal Court of Canada, we anticipate a major expansion of the law data bases during the next year. Perhaps most important, we expect later this year to be operating a data base of all federal regulations currently in force.

Almost since the outset of the QUIC/LAW Project, we have worked with the Canadian House of Commons to develop computerized processes for the editing and printing of Parliamentary publications, including Hansard. Development of this computerized process has included plans to save the machine-readable text used in the editing and printing process and to use it to create data bases. On an experimental basis, we now are operating for the House of Commons data bases of written questions and responses and of oral questions and responses during the daily Question Period. In both cases, these data bases extend back to the beginning of 1973, when the first machine-readable text of Hansard was generated.

In another recent development of interest to many Canadian users, we have agreed to the operation by the Globe and Mail, using QL's retrieval system, of a data base which will consist of the full news and other editorial text of that newspaper, starting with January 1, 1976.

As this brief outline has indicated, QL places a strong emphasis on the development and operation of

WAA AND THE QL SYSTEM

Canadian data bases. Of the 21 data bases we currently are operating, 16 are Canadian. We welcome suggestions concerning other data bases of interest to our Canadian network of users -- currently more than 120 terminals from one coast to another. Free communications circuits currently are provided to Vancouver, Calgary, Edmonton, Toronto, Ottawa, Fredericton and Halifax, and other cities are to be added.

As a Canadian company, we are trying to make our system bilingual. Obviously we will continue to have data bases in which all the documents are only in the English language -- particularly data bases from the United States of interest to Canada. But we are starting to build data bases in which all the documents are only in the French language. At present, five of our data bases are French-language data bases. The systems messages to the user are available to him and he can make system responses in the language of his choice.

Finally, I should mention that our network is known as the Shared Information Service -- and I should stress the importance we attach to the sharing of information and to the sharing of services and of the costs of providing information. We have developed contractual arrangements with government agencies and other data base owners to enable them to share their data bases with others -- some of the others including users who normally use information from other fields of knowledge. We share the costs of operating the data bases with the data base owners and we share the revenues produced by use of the data bases.

Only through an emphasis on cooperation and sharing can a full range of information services be made available to all Canadians at a reasonable cost.