HEDL-SA-2416FP

DALIS: A COMPUTER-ASSISTED DOCUMENT-RETRIEVAL SYSTEM FOR THE FFTF

Weeonna G. Harves

May 12, 1981

Association pour l'avancement
des sciences et des techniques
de la documentation
(ASTED)
September 1981
200 Promenade du Portage, Hull (Quebec), Canada

HANFORD ENGINEERING DEVELOPMENT LABORATORY
Operated by Westinghouse Hanford Company, a subsidiary of Westinghouse Electric Corporation, under the Department of Energy Contract No. DE-AC14-76FF02170
P.O. Box 1970, Richland, Washington 99352

MSW

DISCLAIMER

DALIS: A COMPUTER-ASSISTED DOCUMENT RETRIEVAL SYSTEM FOR THE FFTF

by

Weeonna G. Harves

Westinghouse Hanford Company

Richland, Washington

The FFTF (Fast Flux Test Facility) is a liquid sodium cooled, fast flux reactor designed specifically for irradiation testing of fuels and components for liquid metal fast breeder reactors. The Department of Energy and the Nuclear Regulatory Commission require that all pertinent documentation for maintenance, operation, and safety of the FFTF be readily accessible and retrievable, both during initial startup and for the lifetime of the plant. That amounts to a lot of information which has to be retrievable. An early study estimated that it would take approximately 12 million pages of text and 750,000 engineering drawings to put the FFTF into operation, and more documentation is being generated daily. At the time this estimate was made, there was no centralized retrieval system for engineering design, research, constructional, or operational paperwork for the FFTF project.

With this in mind, the Hanford Engineering Development Laboratory (operated by Westinghouse Hanford for the Department of Energy) began an in-depth study on document storage and retrieval. This study was headed

by engineer Michael Theo of Westinghouse Hanford Quality Control, and was carried out with the help of the Stanford Research Institute and with that of Dr. Virginia Sternberg, a Westinghouse Librarian with 22 years of experience working with nuclear power documentation.

The resulting recommendation called for a documentation and retrieval system using computers, microfilm equipment, and a method of indexing for rapid retrieval. The software package preferred was a System 2000 with a computer capable of supporting it. However, for reasons of economy, the company turned to a minicomputer with an expandable data processing system that could also provide fast access to a large, on-line data base, and that could drive a microfilm reader/printer. The minicomputer has three disc drives, each capable of storing 200 million characters, or the equivalent of 300,000 records (documents). The minicomputer can also service 32 CRT's working at the same time.

The indexing system finally developed is called the DALIS system, short for Document and Location Indexing System. This system was designed by an engineer (Michael Theo) for use by engineers. DALIS uses descriptors and keywords to identify each document in the system. The descriptors give such information as document number, date of issuance of the document, the title, the originating organization, and the microfilm or hardcopy location of the document. The keywords are words or phrases that describe

the content of the document and permit retrieval by means of a computer search for documents with the stated keywords. This system has several advantages:

1) It provides a centralized indexing system for the whole plant.

2) Keywords can be used in combination with descriptors to narrow down a search. For example, after searching for all documents dealing with "seismic analysis", an engineer could narrow the search down by asking for all documents dealing with seismic analysis and issued after June 1978. The engineer or other user could narrow the search down still further with the use of other descriptors.

3) The system uses terms with which the engineers are familiar, which means that the system can and will be used by the people it is intended to serve.

4) The system allows users to search for specific subjects as well as general ones, e.g., "cold leg check valves" can be searched for as well as "valves." Of course, this advantage can apply in reverse: it can be advantageous sometimes to be able to search for the general as well as for the particular.

5) Keywords can be technical for the expert and semi-technical for the non-scientist. For example, where an engineer might search for a "mechanistic analysis," a clerk might search for a simple "report."

6) With this system, work could begin on indexing documents before the computer equipment had arrived.

Development of this keyword indexing system began in mid-1975. A survey was performed to determine which documents were to be included in the system and which were to be eliminated. The survey also ascertained what kind of documents would be included in the system, and what items of information would be important to have for each document type.

From this survey, a matrix of 24 data fields was developed. Fifteen of these fields are searchable (i.e., one is able to find documents by searching one or more of these fifteen fields.)

Each field can have an unlimited number of entries, provided that no word or phrase contains more than 45 characters. Thus, a single large document can, in theory, have hundreds of keywords assigned to it. This ability to enter as much information about a document as is needed gives DALIS increased flexibility and usefulness.

A data input sheet was designed. A display with a similar format appears on the CRT screen for easy data entry and data retrieval.

A company name list for originators and receivers of documents was developed, and a list of document types was begun. A keyword thesaurus was started that rapidly grew from a few dozen keywords to its present size of 19,000 keywords. The document type list, the company name list, and the keyword thesaurus are ongoing, and have been established as "authority files" on the data base. The authority files permit greater control of the data base by preventing misspellings and improper items.

An acronym list was also compiled and is constantly being updated, both for use of DALIS and for that of the laboratory as a whole.

One of the areas which required a great deal of experimenting was the work-flow pattern for data entry into the system. One of the more difficult areas was determining how the information would be input to the computer. Several options seemed feasible:

1) Keypunch the data and enter it into the system via tape.

2) Input directly to the system from the documents, using a CRT.

3) Use non-technical clerks to add descriptors to a data input sheet (title, document number, author, etc.), with analysts adding the keywords, and a third group inputting via a CRT.

4) Use analysts to do all the preparation (writing descriptors and keywords down on DALIS input sheets), and forward the input sheets to another group for input via CRT.

At the beginning of the project, a group of clerks wrote descriptors onto DALIS input sheets; analysts verified their work and added keywords. Another group called Data Entry then entered the data to the system via a CRT. It soon became apparent that the analysts were spending more time correcting the work of the clerks than it would take them to do it themselves, and so this approach was dropped.

The analysts then did all the data-sheet preparation and forwarded the data sheets to Data Entry. This method has some success, but management of the input sheets became a problem. While this method is still used on special projects, for routine data, data input clerks enter descriptors into the computer directly from the documents at a CRT, with analysts verifying their work and adding the keywords via a CRT. This direct-entry, direct-revision system reduces the time required to enter a document into the system and to verify its correctness.

A basic resource library was created at the beginning of the project, and it has been steadily expanded. Various equipment lists, chemistry and physic handbooks, scientific dictionaries, the final safety analysis report for the FFTF, and many other reference works are on hand to help the analysts learn more about the FFTF and its documentation.

Perhaps the hardest job throughout the project has been the selection of the staff to do the keywording. It was a new type of job; experienced keyworders were not available. Was it necessary to utilize keyworders who had an education in the subject field (like nuclear engineers) or maybe someone who had experience working in it? It quickly became clear that this would be impossible.

A job description and salary scale was developed with the help of Personnel Relations, and the position was made an exempt position with a college degree required. The position was called Documentation Index Analyst. Persons hired in the past have been engineers, librarians with a Master's degree, English majors with or without a computer science minor, chemistry majors, biology majors, psychology majors, and education majors with a science background. Most analysts didn't know what FFTF stood for when they started out, let alone the difference between an ECN, an SNR, a stress analysis, a thermal analysis, or a hot leg and a cold leg (in relation to a reactor). And only a few knew anything about computers.

Librarians are generally the best keyworders, as they are trained to catalog and they enjoy reading--a keyworder's job involves a lot of reading. In the DALIS training program, every analyst spends the first two weeks on the job going through a specific set of instructions before actually doing any keywording or data entry. For the next three months new analysts work very closely with experienced analysts, in a sort of buddy system. During their first six months on the job, they are on probation; very few have failed to qualify at the end of that time. The analysts, through their reading and analyzing of documents, are able to acquire a comprehensive knowledge of the laboratory and of the FFTF project.

What is the result of this work? One big project goal was to index and keyword 250,000 documents and have them retrievable by CRT before July 4, 1978. That goal was achieved, and our data base has been expanding ever since. Nearly 550,000 document locations are now retrievable via a CRT at four satellite stations located throughout the plant for the convenience of the engineers. These are staffed by trained operators, with the most difficult searches referred back to the Data Base Administrator.

DALIS can find documents that answer such questions as these:

1) "I need the analysis that was done on the containment building, along about 1970 or 1971. Do you have it?"

2) "Give me a listing of all the drawings that show the locations of all the ladders in all of the buildings, except Bldg. 405."

3) "My boss needs a copy of the report he wrote on velocity profiles, because he lost his. Do you have it?"

4) "Who made the thermocouple stamped with the serial number 4TEXYZ?"

5) "We need the calculation sheets that pertain to the 'C' section of the operating deck. We're desperate because if you don't have them, we're going to have to re-do all those calculations, and that's about a 3-month job for 2 engineers."

6) "How many SNR's are there against Purchase Order 12345? I need copies."

7) "I need all the documents on heat controls for argon cover gas in loop 1 in the primary system."

8) "How many letters did we receive from DOE in the first half of 1978 on the subject of valve safety?"

The crucial factor in the success of a document retrieval system, whether documents are stored in microfilm or in hardcopy, is the ease of

document retrievability through the indexing system. Any storage system is worthless unless documents can also be retrieved easily. DALIS has proved to be a workable, reliable, and viable retrieval system. The satellite stations now receive between 3000 and 4000 requests per month. The DALIS data base has been expanded to include not only the FFTF documentation but all of the documentation at Westinghouse Hanford. In addition, the system is also flexible enough to be used for other data bases. At present, 45 other DALIS-type data bases (much smaller, of course) are operating, serving such functions as accounting for FFTF spare parts, tracking drawings and tracking moveable property. While we have had our share of problems, we are constantly striving to upgrade and improve our system.

-EXAMPLE OF A SEARCH: FIND A STRESS ANALYSIS ON PIPING DONE BY BECHTEL AFTER 197
5 AND REFERENCED IN THE FINAL SAFETY ANALYSIS REPORT._

```
-SEARCH
S:
S: FSAR
  2831 HITS.
S: AND STRESS ANALYSES
  126 HITS.
S: AND PIPING
  18 HITS.
S: AND DATE>750000
  4 HITS.
S: AND ORIG=BPC
  1 HIT.
S: D_
```

-OR THE SEARCH STATEMENT MAY BE ENTERED THIS WAY:_

```
S: FSAR AND STRESS ANALYSES AND PIPING AND DATE>750000 AND ORIG=BPC
  1 HIT.
S: D_
```

```
                          RECORD DISPLAY PROGRAM                PAGE 1
                        HIT #1 - RECORD 051490765
 1)DOCNO    BR58535039
 3)REV      0
 4)DATE     770930
 5)TITLE    VERIFICATION FOR PIPING FLEXIBILITY COMPUTER PROGRAMS USED IN FFTF
            STRESS ANALYSIS
 6)ORIG     BPC
 9)XREF     FSARQ112.30.4R001,TI75001-19,EROB10268,DDTB13344
11)DOCTYPE  REPORT
12)LOC      051490765
24)KWDS     FFTF,FSAR,APPLIED TECHNOLOGY,COMPUTER CODES,STRESS ANALYSES,PIPING,
            PIPING FLEXIBILITY ANALYSES,PISOL 1A COMPUTER CODE,PISOL 3A
            COMPUTER CODE,PIPESD COMPUTER CODE,ME 632 COMPUTER CODE,ME 101
            COMPUTER CODE
```

-ONLY FIELDS WITH INFORMATION IN THEM ARE DISPLAYED._