M. Zeleznik

Unclassified

The Design of an Extensible Communications-Based Full Text Information Retrieval System

Prepared under Contract 82F765800,
"Development of Algorithms and Evaluation of Performance for Information Retrieval Systems"
for the Information Systems Research Division
Office of Research and Development
Central Intelligence Agency

by the

Department of Computer Science
University of Utah
3160 Merrill Engineering Building
Salt Lake City UT 84112

July 1983

Table of Contents

Introduction

This report describes the preliminary design and operation of a text information handling system capable of filling the needs of a variety of Agency users. In addition to offering conventional information retrieval services from very large databases, it supports the addition of other text processing tasks, such as word processing and electronic mail, to better support the needs of an analyst or other user. It can support a variety of indexing and searching schemes, including specialized backend systems. Other user aids, such as knowledge-based query reformulators, can also be easily added.

The user's view of the system is a series of windows on a workstation or display. Each window corresponds to a key system function, such as the entry of queries to a particular database or a view of the retrieved documents, or a word processor. The editor resident in each window allows for the searching of the information within that window and the moving of data from one window to another. Using this approach, many of the commands previously required in an information retrieval system, particularly for manipulating queries or browsing results, are merged into the existing editor commands.

The internal structure of the system is communications-based, with all processes viewed as being on a bus network. All interactions between the processes consist of passing messages on the network. The reception of a message by a process activates it, and when it has completed the desired action, it returns its results to the process designated or inferred by the original message. New processes can be easily added to the network. System performance can be monitored without affecting the system by a special process that tracks messages on the network.

While the system logical structure is one of separate machines for each processes, it can be mapped into a variety of physical implementations, including the entire system operating on a single mainframe machine. In this case, the low level data communications is performed by conventional subroutine linkages.

1.1 Overview of this Report

The remainder of this section will discuss the background of the design and its goals. It covers our view of the functions necessary for any information handling system: location of documents that match a given query, retrieval of those documents, display and manipulation of the retrieved information, and enhancements (like word processing and electronic mail) to the user environment.

II.5 References

[1] Paul J. Brusil.
Protocols for Unifying Distributed Systems in Hospitals.
In Proceedings of the Sixteenth Annual Hawaii International Conference on System Sciences, pages 14–24. University of Hawaii, 1983.

[2] Harry Katzan, Jr.
Systems Design and Documentation.
Van Nostrand Reinhold, 1976.

[3] John E. McNamara.
Technical Aspects of Data Communication: Second Edition.
Digital Equipment Corporation, 1982.

[4] Edwin E. Mier.
High-level Protocols, Standards, and the OSI Reference Model.
Data Communications :71–76,83+, July, 1982.

[5] Glenford J. Myers.
Advances in Computer Architecture.
John Wiley & Sons, New York, 1982.

[6] Nihal M. Hounou, Yechiam Yemini.
Development Tools for Communication Protocols.
1983.
Working Paper, Department of Computer Science, Columbia University.

[7] Operating Systems, Inc.
High-speed-text-search Design Contract Design Specification Document.
1977

[8] Roger S. Pressman.
Software Engineering: a Practitioner's Approach.
McGraw-Hill, 1982.

[9] J. F. Stay.
HIPO and Integrated Program Design.
IBM Systems Journal 15(2):143–154, 1976.

[10] Carl Sunshine.
Formal Techniques for Protocol Specification and Verification.
Computer :20–27, September, 1979.

[11] Andrew S. Tanenbaum.
Software Series: Computer Networks.
Prentice-Hall, Englewood Cliffs, NJ 07632, 1981.

[12] TRW.
SAFE Project System Requirements Specifications.
1982

[13] U.S. Department of Commerce.
The Patent and Trademark Office Automation Master Plan.
1982
In 3 volumes.

[14] David C. Walden, A. A. McKenzie.
The Evolution of Host-to-host Protocol Technology.
Computer 12(9):29–38, September, 1979.