

COMMUNICATIONS
FORUM

JEWEL

ELECTRONIC PUBLISHING

Harold J. Logan, Dow Jones & Co., Inc.
David Gifford, MIT
Walter Bender, MIT

April 10, 1986

Genga Arulampalam, Rapporteur

Massachusetts Institute of Technology
Communications Forum

ELECTRONIC PUBLISHING

Harold J. Logan, Dow Jones & Co., Inc.
David Gifford, MIT
Walter Bender, MIT

April 10, 1986

Genga Arulampalam, Rapporteur

ELECTRONIC PUBLISHING

David Gifford - MIT

Gifford's lecture was centered around the new system architecture he has been working on at the Laboratory for Computer Science(LCS) at MIT. Two major ideas put forward by this project are - (1) the new method of organizing databases called multi-database, (2) a new approach to communications. The goals of the project are to examine a new way of building large scale information systems and to focus on the use of low cost personal systems to provide a way of customising information services to the individual. Gifford pointed out that in the 1960s, time-sharing was invented and widely commercialized in the 1970s, while in the 1970s local area networks were invented and widely commercialized in the 1980s. He stated that they were looking for a new technology for the 1980s that could be commercialized in the 1990s and widely used as a fundamental way of distributing information. As a result, he said that they are looking beyond the current system building block assumptions of using local area networks and work-stations, to actually build and test a system that people outside of the research group could use. The two options of doing this are either to build a mock-up/prototype on something like the symbolic LISP machine or to build an IBM PC related version that could be used by a number of people. The research group he said, opted for the latter.

The system architecture differs from many others because it organizes a confederation of cooperating databases. Instead of a simple master database they use a large collection of databases. Also, they have personal databases (i.e. IBM PC), and share database services. The two systems are connected together and kept up-to-date using low-cost broadcast simplex communication. Gifford used a schematic diagram of the system(see Appendix A) to describe the path taken by the information that is transmitted. The database server contains the entire database and transmits digital information continuously over the subcarrier of a Boston FM radio station. The system provides large economies of scale because the number of users (radio receivers) does not affect the load on the server at the central site. Regarding the cost of the radio-receiver, he noted that while it cost less than \$100 at present it is likely to decline further. Because of the limit of storage space on the PC, the project provides facilities to allow each user to customize the fraction of the database they like to keep. As a result, when the user wishes to access the information that is not on the PC, the technology allows them to contact one of the servers via an autodial telephone connection.

Gifford then went on to discuss how the system organizes the databases. They use predicate databases which are accessed by independent facts that may be incomplete and the system responds with a matching set of records. Since all the information is not on one database, when a query is received the

system looks up the content label on each database and locates the necessary information. The system has a very simple theorem prover as part of the search calculus that is included in the PC software. The basic technology of putting content labels on databases is used in two different forms - (1) inside the personal database system, where a query reading module inside the IBM PC uses a set of database content descriptions to find out where to process the query, (2) inside LCS's database servers themselves. The former method is the most efficient and cheapest. The actual code is the same on the personal system as it is on the server. The content reasoning framework can also be used to attach CD-ROMs to a PC. In this situation, you put a database on a CD-ROM with a content label on it, then the PC will be able to figure out whether a query should be processed at a local database or whether it should via a telephone call access a remote server.

The transmission system utilizes error correction and encryption. One of the reasons for using encryption is in anticipation of commercialization, in which case charges could be made independently for different services, and users will have key ranges inside their PCs that enable them to select among the various types of information they are permitted to access. The encryption is done at the server side while the PC handles the decryption.

Gifford anticipates that one of the most important issues in communication will be the formalizing of interfaces to remote services. He has found that the remote procedure model is not powerful enough for their application and has accordingly extended it. He said that one of the problems with remote procedure is that this basic call return model does not handle bulk data transfer.

This project began in April 1984 with one user and was followed by 20 hand-ouilt receivers until December of last year. In December they distributed 165 systems to participants in the experiment recruited from the Boston Computer Society and other Boston area organizations. The participants fill out a monthly questionnaire. The initial sample indicates that an average person uses the system for about 40 minutes each day. The average system actually receives data from the broadcast service for about 10 hours each day. Gifford predicted that multi-tasking will soon come into PC operating systems. He said that it is now possible to run the database system in the background with an extension product like Double-DOS, though it is somewhat difficult on low power.

Explaining the details of transmission, Gifford stated that a small antenna plugs into a receiver. This receiver delivers information to the computer at 5 kilo-bits per second. In addition to FM broadcast technology, he said that CSL had looked at TV, direct satellite, 900 MHz FSK, and cable transmission.

Walter Bender - MIT

Bender asserted that people use information systems because they have to, in order to obtain the necessary information. He pointed out that when these systems have been installed in the home they have not been very successful. People use them for awhile until the entertainment value fades. He stated that the dilemma is partly to do with the types of information and also the types of interfaces and terminals that have been provided. He added that there has to be a very tight coupling between an individual's activities (e.g. calendar, mail, schedule, etc.) and the information that has been provided, in order to make these things interesting enough, palatable and usable.

Discussing the work of the Electronic Publishing Group (EPG) at MIT's Media Laboratory, Bender said that they basically take broadcast items like newspapers, magazines, TV, billboards, radio, etc., and decouple the transmission (the broadcast) from the receiving of information. The decoupling does two things - (1) provides local storage, (2) provides local computation and local intelligence that can do filtering and also some association between various databases and items of interest. As an example he cited his computer editor which reads the electronic mail sorting out the items he would be interested in. He clearly differentiated between the distribution and storage of information, and the perusal and use of information.

The systems designed by EPG are not query-based but are designed to encourage browsing. The system gives information on what is happening and the individual browsing around gradually zeroes in on matters of interest. It is modelled to some extent on the newspaper, a useful communications vehicle that has evolved over the years. Bender used a slide (see Appendix B) to describe the facilities and functions of the system developed at the Media Lab. He compared the computer editor to the editor of a newspaper who gathers information from reporters selecting those that should be published. The computer editor gathers information that the reader would probably be interested in, and via the PC, local computing, local storage, and local database access, which act as a bandwidth expander, gives the reader something that's entertaining and informal. The system like most newspapers has a front page containing stories that the computer editor expects the reader to be interested in. The act of browsing, touching a word, or flipping through a page in a story, indicates to the computer the reader's interest. Accordingly the computer responds by highlighting other related stories. The computer observes the reader's actions and reacts trying to second guess and make suggestions. Finally, when the reader asks for a complete story it produces it with pictures to illustrate (stored in an optical disc). In addition to pictures of people, places, and things, the computer also has locally generated maps.

Some of the stories were illustrated by television broadcasts stored on tape. Bender at this point used a video to illustrate the workings of the system.

The computer editor currently facilitates the process of reading. Bender stated that computing and editing functions are now merging.

Harold J. Logan - Dow Jones and Company, Inc.

At the outset Logan asserted that the technology available today together with one's imagination can take us to places far beyond what we can at present actually apply in any sort of realistic and profitable manner. He stated that Dow Jones News/Retrieval (DJN/R) is a profitable business and one of the relatively small number of profitable businesses in the world of electronic publishing. The system he said, operates on the basis of a technology that is already 13 years old and is not very fancy compared with the new systems available today. While Dow Jones' system can't do many of the things that are available in today's technology, it still seems to succeed in addressing the needs of customers to their satisfaction.

Discussing the economics of service, Logan stated that Dow Jones charges anywhere between \$12 and \$168 an hour for News/Retrieval depending on the time of day, type of database and the speed of the modem. Of the revenue, Dow Jones has to pay for 2 or 3 kinds of basic expenses, one of which is fees or royalties to the information providers. The service consists of about 38 databases of which 12 are generated from internal news resources. Other basic expenses are telecommunication costs and the cost of people working in News/Retrieval. Of the 225 people employed by DJN/R, 70 are in the systems area, about 60 in the editorial area, 60 in marketing, and about 30 or 40 in customer services. Another contributor to overall expenses is the cost of physical infrastructure (buildings, machines, etc.,) that support news retrieval.

Logan mentioned that there were 3 categories of users: personal investors, business users, and consumers. Personal investors are people who use it because they have to, or because they can make some money using it. The service he said, was built for personal investors who now form the largest category of users. This is the kind of person who is active in the stock market making 4,5,6 or as many as 12-15 stock transactions in the year. Typically, the person has a \$150,000 portfolio or more. The second largest customer is the business user. He said that if DJN/R is to become the size and power of business they want to be, they have to "figure-out" who the consumer category of customer is, what they want delivered through this kind of system, and at what price.

Logan then used a video tape to describe the type of service that is currently offered by DJN/R - ranging from stock prices to weather and sports information. The personal investors' interest is primarily in stock prices, transactions, etc, while the business user is interested in stock information as well as company financial and historical information, and such details as airline schedules, prices etc,. For the consumer the system provides information regarding sports, news, weather, education, movies, etc,.

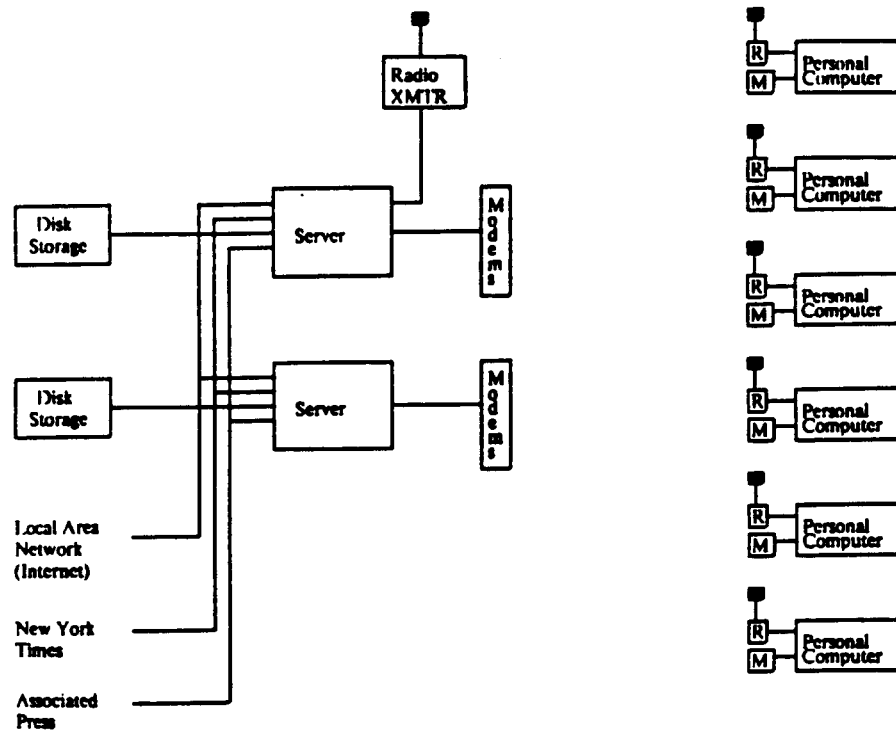
SPEAKERS' COMMENTS AND RESPONSES TO QUESTIONS

A member of the audience commented that the DJN/R service appeared attractive though it was expensive. He inquired as to whether in the long-term services would be available nearer the cost of newspapers and on a fixed price basis. Logan in response said that the 25 cents paid for the newspaper is subsidized by the advertiser. He said that the advertising represents 75 cents in every dollar earned by the newspaper. The DJN/R system has no advertising and unless there is an alternate revenue stream, the producers of this and other similar services will not be able to bring the prices down. With reference to advertising, he said they were involved right now with a couple of companies investigating the possibilities. However, one thing they are determined not to do, is to use 'intrusive advertising'. i.e. they do not want to force the user to look at advertisements. An alternative that is being considered is to list advertising related to particular services in a separate database to cater to a user's particular interest/need.

Responding to a question about the heuristics used by the system, Bender said that it was a very complicated problem and he was not doing any 'article understanding' in his system, but he mentioned that there were some people who did a good job at building templates for specific areas in terms of understanding articles etc. Bender said that his system worked primarily on a keyword basis and had a lot of heuristics.

Russ Neuman commented that in the wake of the withdrawal from business of the two prominent videotex market experiments just last month (Viewtron in Florida and Gateway in California) many are claiming that there is really no market for this type of business. A member of the audience suggested that the two companies have gone out of business because once the initial novelty fades out people are not willing to pay monthly charges for a system which does not have a good user-interface. The reason suggested for the DJN/R service being popular in spite of its archaic technology, is that it serves a special niche market of investors. People perceive that they can make money using the DJN/R service and are willing to pay for it. After all,

it is tax-deductable.



System Block Diagram

