Information Processing & Management Vol. 25, No. 5, pp. 509-514, 1989
Printed in Great Britain.

0306-4573/89 $3.00 + .00

THE LIVERMORE INTELLIGENT GATEWAY: AN INTEGRATED INFORMATION PROCESSING ENVIRONMENT

HILARY D. BURTON
Technology Information Systems Program, Lawrence Livermore National Laboratory, University of California, P.O. Box 808, L-542, Livermore, CA 94550, U.S.A.

(*Received* 1 *September* 1988; *accepted in final form* 4 *January* 1989)

Abstract—The Lawrence Livermore National Laboratory developed the Intelligent Gateway software in work that began in 1975. The Intelligent Gateway creates a framework that links distributed, heterogeneous computer resources and provides a single user interface such that a "virtual information system" can be tailored to any user's idiosyncratic needs. The Gateway can provide access to external services as well as serve as resident host for installation of local databases and information processing tools. Thus, it provides an integrated information processing environment to support the complete range of scientific and technical information needs.

INTRODUCTION

The Intelligent Gateway, developed by the Lawrence Livermore National Laboratory's Technology Information Systems Program, is a machine-independent software system that provides simplified access to geographically distributed and physically unlike computer-based information resources. Through the Gateway, a knowledge worker can utilize any variety of resources without learning different access methods, logons, and communication protocols. From a single entry point, that subset of textual, numeric, graphics, or other computer-based systems that best meet his needs can be targeted through the Gateway to create a virtual information system. In addition to extensive data access capabilities, the Gateway system provides powerful analysis and processing tools to complete the creation of an integrated information environment.

Thus, the Livermore Intelligent Gateway provides an information environment featuring automated access to the diversity of scientific and technical information available in today's rapidly proliferating online world. From the literature and raw data files of basic research to the databases of the applied disciplines, the Gateway provides a single point of access to the breadth of information resources supporting science and technology and the tools to manage these data.

INFORMATION PROLIFERATION—THE PROBLEM

Current communications technology supports electronic access to computer-based secondary services such as *Biological Abstracts*, *Chemical Abstracts*, *Engineering Index*, the Commonwealth Agricultural Bureaux files, and others. But the proliferation of online services, with their idiosyncratic search systems, different operating policies, and peculiar communication protocols, can complicate the life of even an experienced information specialist, let alone the novice searcher.

Studies have shown that no one database provides adequate coverage for the majority of information needs. Thus, a combination of online databases has been shown to return considerably more relevant citations than use of a single source [1]. For example, in a study of selective dissemination of information services within the food science and technology area, less than 5% of the queries were carried out on single databases. The most popular combination involved joint searching of Food Science and Technology Abstracts (FSTA), BIOSIS, CAB, *Chemical Abstracts*, and Agricola [1].

Since each of these databases has a unique physical format and indexing approach, effective multi-file searching is complex. Furthermore, search strategies vary depending on the access approach selected. For example, searching the Chemical Abstracts database via STN, Lockheed DIALOG, and Pergamon's ORBIT requires three different sets of procedures, including different logon procedures and retrieval languages. Furthermore, each system yields citation output in a different format.

While a number of commercial products designed to ease the searcher's burden have recently reached the market, these are only partial solutions. Without exception, they facilitate access to a predetermined set of systems, which is much narrower than the full set of potentially relevant sources. Software such as Scimate, Pro-Search, and Easynet provide access to a limited set of resources easier and faster but do not provide a solution to the fundamental problems caused by the diversity of the average user's unique information needs.

Thanks to the availability of more than 3600+ online databases [2], the end user becomes inundated with stacks of citation listings and undigested streams of computer printouts. SDI services alone can provide the average user with an information input flow that is anywhere from three to ten times what he previously obtained by manual methods [3]. The tools necessary to manage this abundance of information are evolving [4], but most are not available in an integrated support package. In 1972, Bennett made the following remarks. They remain applicable today:

> There are several requirements for further development of the emerging user-interface technology. First, it is imperative to cut through the mass of inessential, application specific detail and to overcome confusion in terminology so that the basic similarity of user services required in many applications becomes clear. Second, observers of the computer scene have decried the tendency of software designers to produce each new system as if others do not exist. To be successful in projecting interactive facilities into new applications, designers must learn to build on the work of others and stop dissipation of resources. . . .[5]

THE GATEWAY SOLUTION

The goal of the Technology Information Systems Program (TISP) at the University of California's Lawrence Livermore National Laboratory (LLNL) is to provide a generalized information analysis and management capability. The Intelligent Gateway software developed by TISP provides automated access to heterogeneous, geographically distributed, computer-based resources. The Gateway is not constrained by physical distance or hardware vagaries. Operating via multiple telecommunication channels, the Gateway can be programmed to select an optimum electronic path to any resource that can be a stand-alone computer, a sophisticated system such as Lockheed's DIALOG system or Pergamon's ORBIT system, or a full-scale network such as the Defense Department's Arpanet/Milnet.

Not only does the Gateway make possible a selection of resources previously not available through a single access point, it also eliminates the tedious clerical efforts required to access and logon to disparate systems. Telephone numbers, baud rates, identification procedures, secret passwords, differences in terminal keyboards, and display screens are resolved by the Gateway software.

Once connected to the target system, a user may interact in the system's native mode, use a Gateway overlaid common command language, or execute a fully automated search and retrieval procedure for routine tasks. At any point, he can shift from one to the other. Additionally, at any point in the interaction, the user can open local disk files on the Gateway computer where the remainder of the session activity will be saved for subsequent review and analysis. This process can be initiated repeatedly and selectively, permitting simultaneous, multi-session downloading of desired information into user-controlled, topical files on the Gateway computer. The downloading of large files can be activated in the background mode while the user engages simultaneously in other activities on another host via the Gateway. Simultaneous connection to multiple hosts is used routinely to compare results, share files, and verify findings. For example, in a search for pesticide information, it is possible to interrupt the search to verify a registry number stored in the user's elec-

tronic mail files. Thus, a user can retrieve the stored registry number, obtain the corresponding chemical and common name from STN, and then continue his search in CIS. All of this can be accomplished in a single session with the use of commands such as "Connect STN" and "Connect CIS." Moving from one system to another or the use of various applications on the Gateway, such as electronic mail, does not require additional logon or identification procedures. The automated Gateway software manages the accesses, logoffs, and accounting information in a controlled manner with appropriate audit files.

Having simplified access to and retrieval of information, be it bibliographic, numeric, or graphic, the Gateway provides a tool box to further analyze and repackage the information. Post processing tools fall into two major categories: (1) analysis of numeric data through statistical, mathematical, and graphics software and (2) analysis and restructuring of text through translation and analysis routines [6]. The post-processing routines for bibliographic citations include reformatting to a common Gateway format. By converting files to a common form, the user can merge files from different sources. Users can add fields, delete fields, change field content, or rename fields to create files tailored to their own needs. Duplicate records can automatically be eliminated according to user-defined parameters. Analytical functions can show distributions of publications by subject, by language or country of origin, by year, or by any element contained in the downloaded record. Furthermore, the user can specify a variety of other combinations to track subject evolution, perform trend analysis, or study patterns of co-authorship, evolution of terminology, life cycles of journals, and co-occurrences of descriptors. Any discrete field can be parsed and analyzed across the total file or in conjunction with any other field. Correlation of user-specified field parameters can be obtained for up to 10 fields. More commonly used combinations include author/subject/time correlations as well as intra-field correlations such as author/author distributions.

To facilitate analysis and improve the quality of the translated file, which is the input to the analytical routines, user-controlled table-driven translators have been developed for several of the major online services. By allowing the user to modify the table that controls the translation, customized translators can be developed to meet the idiosyncratic needs of a user or the peculiarities of a given database.

Results of the analysis can be displayed in tabular, textual, or graphic form. Analytical results can also be saved as files and cumulated or incorporated in subsequent projects such as literature reviews or annotated bibliometric studies. Trend analysis, made possible by the Gateway post-processor, has been demonstrated to be extremely valuable in research planning, information program management, and research in sociology/history of science. To do such analysis manually is extremely tedious and time consuming, even on small files. The availability of computer-based information analysis tools makes such studies practical, cost-effective, and nontedious. The recent appearance in the ORBIT system of the "GET" command, which provides a frequency count of single or multifield values for a retrieval set, is a first step toward providing analytical capabilities on a commercial basis.

In addition to an extensive analytical toolkit, the Gateway also provides sophisticated electronic mail and teleconferencing capability as well as a wide variety of Unix utilities such as text editors and document preparation subsystems. The electronic mail module, which was one of the earliest of the Gateway modules developed, provides the means for automated file transfer between individuals and/or computer systems with little or no user intervention. It also includes a superset of the most popular functions found in the most common electronic mail systems. And the various Unix utilities can be used directly by users knowledgeable of the operating system, or they can be presented through menus embedded in the interface.

More importantly, the Gateway provides a framework into which any computer-based resources can be integrated. This can range from software tools such as word processors and spreadsheets to database management systems and the databases themselves. Furthermore, the resulting environment can be presented to the user in a customized manner that varies from one user to another. The menus that a given user or group of users sees on the Gateway can be tailored to the vocabulary, expertise level, and privileges of each user (group). Each user can see his own customized environment and need never be concerned

with all of the other user views that may exist elsewhere in the system. These menus can be easily modified or updated by a systems administrator.

GATEWAY IMPLEMENTATION

The TISP Gateway, which has been under development at LLNL since 1975, has been implemented on a wide variety of Unix-based computers from PC's to large super minicomputers. It supports data access via public packet switched networks (PSN) such as Tymnet and Telenet, government PSN's such as the ARPANET and dial-up over commercial and federal telephone lines. The automated communications portion of the Gateway, known as NAM [7] supports three types of external interfaces: asynchronous serial (RS-232C), modems and auto-dialers, and TCP/IP-based connections. This flexibility allows the Gateway to link incompatible systems into a network via a single access point. NAM is based on a complete interpretive scripting language that supports robust error checking, conditional branching, and multiple simultaneous connections. The combination of controlled branching and error checking allows the scripts to contain alternatives so that multiple access routes can be used to reach external hosts. Existing NAM scripts usually incorporate six or eight alternative access paths for a given resource. These capabilities enable the Gateway's NAM to far exceed available PC communications software in kinds of hardware that can be accessed, routes that can be taken, sophistication of error conditions that can be handled, and number of simultaneous error-free connections that can be maintained.

Gateway technology has provided the means to create "virtual information systems" made up of very real tools and resources presented to the user in a customized manner. Support of information access, retrieval, analysis, and presentation are critical functions of the Gateway. Work at Lawrence Livermore National Laboratory for the DoD Defense Technical Information Center has resulted in several prototype gateway systems that demonstrate the concepts discussed above [8]. Future work is directed at making these tools more generic—to create a more integrated computer-based environment in which the experienced information specialist, regardless of whether he is a reference librarian or information broker, can create evaluated, synthesized information packages tailored to his various users' needs. By modularizing the code to support generic functions and using a table-driven approach to facilitate maintenance and extension, we can provide software that is configurable to meet the widest variety of information needs. The Intelligent Gateway then provides the foundation on which we can install various tools. It also acts as a neutral navigator between the many incompatible operating systems and hardware configurations that make up the resource base we wish to access.

In 1987, the Gateway software was licensed by the University of California to Control Data Corporation through LLNL's Technology Transfer office. CDC presently markets a subset of the Intelligent Gateway software under the tradename, ASCENT.

The concept and implementation of a gateway are not entirely unique to our work. One of the closest counterparts exists in Canada in the form of iNet [9]. iNet, in fact, incorporates some of the TISP software that was made available to Bell Northern Research by TISP in the late 1970s. In the United States, there has been a recent announcement of a three-part venture to offer iNet service on a commercial basis [10].

Another operational gateway exists at the International Institute for Applied Science Analysis (IIASA) in Austria [11]. This gateway began service in July 1980 and in August 1981 installed a second node in Budapest at the Institute for Computerization and Automation of the Hungarian Academy of Sciences. This gateway facilitates communication among research collaborators and provides access from VINITI in Moscow and other organizations in the USSR via a node through the All-Union Institute for Systems Studies to most international networks and information centers and, in particular, access to AGRIS and INIS via the IAEA in Vienna.

We are faced with multiple dissimilar systems offering multiple resources. Developers of these different systems structure their retrieval and input/output processing capabilities differently. Databases change over time and may exist within the same system in differ-

ent versions. Refined systems also undergo continuous change and improvement. User needs and familiarity determine preference for one system over another. These preferences change due to introduction of new services, user education, and economic considerations. The Livermore Intelligent Gateway provides a single framework to support solutions to each of these problems. Future research will focus on developing software to manage a distributed gateway system where each node can be optimized for a particular function. Thus, hardware to support AI systems, database management, graphics, and solid modeling will be linked together in a manner that is transparent to the user [12].

THE FUTURE

The availability and number of intelligent gateways will undoubtedly increase in the future. The need for a gateway or intelligent "front-end" to link heterogeneous systems at geographically distributed locations is a logical extension of current technology.

In particular, the new crop of powerful super-micro and super-mini computers and broad-band, inexpensive satellite communications and relatively inexpensive disk storage now make it possible to establish regional information centers at nominal cost even in areas previously unreachable. The gap between the information "haves" and "have nots" need not continue to widen. The international network of agricultural research centers has taken a first step in this direction [13]. These centers, which are located in many developing countries, are now using electronic mail to communicate among themselves and transfer information obtained from the commercial online retrieval systems. While telecommunications costs may still be high and quality may be variable, it is an important first step. Utilization of Gateway software would allow the information disadvantaged to augment their current efforts with broader access and more comprehensive processing followup. Other groups in the less developed countries could similarly benefit from this technology.

The complexity of the world of available computer-based resources is clearly increasing at an intolerable rate for the average user. This situation poses a dilemma. There is a growing need for research and information and the tools to manage it. Utilization of an intelligent gateway to optimize access to and analysis of the information resources capable of supporting such research—bibliographic, numeric, and predictive—promises dramatic improvements in this critical area.

Unlike the distinction Landau makes between front-end buffer gateways (Easynet) and host-to-host link gateways (MDC-DIALOG-OCLC [14], we see the Livermore Intelligent Gateway as more than just a buffer or a link between computers. The ability to support a multiplicity of uniquely specialized information processing environments linking heterogeneous, distributed resources comprises a powerful yet flexible toolbox. And work currently underway will certainly move the current software in the direction of a "knowledge gateway" as defined by Hawkins *et al.* [15]. The Livermore Intelligent Gateway creates and supports the necessary environment where the full spectrum of information access and processing tools can be provided.

Today's information specialist has a wealth of potential resources to draw from. The computer has made these accessible from even the most remote location. The advent of intelligent gateway technology now provides both the means to bring together relevant subsets of these disparate resources, as well as the means to fully utilize them via analytic and post-processing software. Such "virtual information systems" will enable the information specialist to increase his productivity, extend his technical capability, and better utilize and manage his information resource environment.

Acknowledgement—The work described in this paper was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

REFERENCES

1. Burton, H.D. Utilization of the FSTA database in an agricultural environment. Proceedings of the Symposium on Food Science and Technology Abstracts, Conference Series No. 2, International Food Information Service; 1980 October 21-23; Berlin.

2. Cuadra Associates, Inc. Directory of Online Databases, Vol. 9, No. 1. New York: Cuadra/Elsevier; January, 1988.
3. Burton, H.D. Multi database searching in agriculture: A cooperative, computerized service. Special Libraries, 69(7): 244-249; 1978.
4. Burton, H.D. The changing environment of personal information systems. Journal American Society for Information Science, 36(1): 48-52; 1985.
5. Bennett, J.L. The user interface in interactive systems. Annual Review of Information Science and Technology, 7: 189; 1972.
6. Burton, H.D. Use of a virtual information system for bibliometric analysis. Information Processing and Management, 24(1): 39-44; 1988.
7. Rosenthal, R.; Lucas, B.D. The design and implementation of the National Bureau of Standards Network Access Machine (NAM). NBS Special Publication 500-35, June, 1978.
8. Cotter, G.A. The DOD Gateway Information System: Prototype Experience. Alexandria, VA: Defense Technical Information Center, Cameron Station; 1986.
9. Cunningham, I.; Raiswell, J. Gateway to online information services. Telesis, 1: 2-7; 1983.
10. Betts, M. Trio targets info services. Computerworld, 21(35): 14(1); 1987.
11. Labadi, A.; Sebestyen, I. The IIASA TPS/70—X.25 Gateway—Network promotes international flow of scientific information. Computer Networks, 7: 113-121; 1983.
12. Balson, D.A. CGNET: A data transfer network for the CGIAR. Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists, 32(1): 39-45; 1987.
13. LANCE Planning Document, Technology Information Systems Program. Lawrence Livermore National Laboratory, 1988.
14. Landau, H.B. A database producer's view of online vendor gateways: A threat or an opportunity. Information Services and Use, 7(4-5): 163-165; 1987.
15. Hawkins, D.T.; Levy, L.R.; Montgomery, K.L. Knowledge gateways: The building blocks. Information Processing & Management, 24(4): 459-468; 1988.