

Economics and usage of a corporate digital library

David Alsmeyer

BT Advanced Communications Research

Abstract

This paper analyses the usage made of the journals and books offered by the BT Digital Library, looking in detail at the usage of the 3,500 people at BT's development site at Adastral Park, near Ipswich, and the impact of this usage on purchase decisions in the library.

The Digital Library offers over 800 electronic journals to its user community. Journals are held on the Library server, where possible, with the Library now offering the IEEE's IEL, Bell & Howell Information & Learning ProQuest collections and selected Elsevier journals, supplemented by links to publishers' web sites for other titles. The Digital Library has recently begun to introduce electronic books, with the offering of twenty O'Reilly books.

During 1999, 1,091 users from this site read journal articles on the Digital Library. A total 9,108 papers were read 12,919 times. This is a somewhat lower level of repeat usage of articles than that reported by Mackie-Mason, et al. In their discussion of the usage patterns in the PEAK experiment's generalised subscription model, a usage level of 2.6 accesses per article is reported compared with 1.4 accesses in BT. The difference may reflect different reading patterns of corporate library users or of software engineers to more general academic communities.

The concentration of usage by journal within the IEEE's IEL collection and the selection of Elsevier journals available in the Digital Library has been analysed. IEL showed a high concentration with 21% of titles delivering 80% of the articles read. The Library's Elsevier collection showed a much lower concentration, with half (ten) of the titles needed to deliver 80% of the articles read. The ProQuest collection has not been available long enough for a similar analysis to be carried out.

Usage throughout 1999 was relatively static with peaks in March and October. Accesses per potential user were between 0.25 and 0.35 a month, higher than those reported in the PEAK experiment.

Electronic access to journals is supplemented by a document delivery system so that users can request documents not available online. The system detects when requests are made for documents that are not yet available online, but will become available in the future. Users get the choice of being added to the waiting list to be alerted when the article becomes available online rather than go ahead with the request for a photocopy. A significant number of users choose to wait for the online version, even though they could have a copy of the desired article almost immediately and without extra monetary cost to themselves.

Introduction

BT is one of the world's leading providers of telecommunications services. Its main products and services are providing local, long distance, and international calls (with direct connections to more than 230 countries); providing telephone lines, equipment, and private circuits for homes and businesses, providing and managing private networks, and supplying mobile communications services. Like other companies in high-tech sectors, BT is experiencing major pressures on costs and products, requiring a challenging combination of cost-cutting and innovation to maintain competitiveness.

Over the last six years the BT Library, in response to these pressure, has transformed itself from a large, conventional corporate library with 750 journal titles into a Digital Library with a small physical presence. The Digital Library now offers more than 1,100 journal titles to its users, with more than 800 available online. In comparison, the Library offers 250 journals in paper format.

The Inspec and ABI/Inform databases act as gateways to these journals, using software developed by BT's knowledge management research team for searching, current awareness and collaboration features. These online journals are supplemented by an end-user driven, disintermediated document delivery system. Users, upon finding an abstract of a paper they wish to read, can request a photocopy from BT's document supply provider, which will be delivered to their desktop.

In the physical library, the librarian could readily see users when they were floundering in their search for information and discretely offer assistance. In the Digital Library, users are relatively invisible to the librarian. The Digital Library is developing methods for understanding its users behaviour more effectively.

Studies of user behaviour are intended to highlight problem areas on the server, which can be redesigned to make them easier to use, and to develop ways of automatically profiling users' interests and work areas. The unspoken purpose of this analysis is also to develop a compelling case showing how effectively the Library supports BT's business processes.

Methodology

Although no detailed study of the usage of the Digital Library has been undertaken, it is important to understand how users access the collection, what problems they find there, and what their usage patterns are. Data about where they are in the organisation are important because of the need to allocate costs and charges.

The Digital Library studies the log files of the library server and uses a number of channels to encourage user feedback.

Log file analysis

The difficulties of log file analysis are well documented. [Wright, 1999] The log files are huge and processing them can be time-consuming. Unless user authentication is required, logs record only machine addresses and not personal identifiers. Each server transaction is logged, so a user retrieving a page with five graphics is recorded in six lines in the log file. Users share machines at cybercafes or they operate behind proxies or they use shared IP addresses, so that the IP address cannot readily be tied to a single user.

A number of packages are available to assist in log file analysis (Busch, 1997). These packages report statistics such as the total hits on the server, the number of Not Modifieds (304's), Redirects (302's), Not Founds (404's), Server Errors (500's), the number of unique URL's served, the number of unique client hosts accessing the server, the total kilobytes transferred, the top one second, one minute, and one hour periods, the most commonly accessed URL's, and the top 5 client hosts accessing server. These deliver a higher level of management information than the librarian needs and are not used in the BT Library

Wright describes techniques for grouping unidentified readers into “constituencies”, based on their usage patterns [Wright, 1999]. These constituencies, such as robot checkers, users checking the What's New page, new users, or demonstrators, are identified by analysing the server's log files, can then be used to observe navigation of the site and spot usability problems.

The BT Digital Library studies its log files to determine the extent of the usage of the resources, to ensure that its resources are used effectively. For example, the comparative cost of providing a journal online on the server and the alternative costs of providing articles via standard document delivery can be evaluated. Perl scripts are used to extract meaningful usage data, concentrating on the html pages and pdf files read and the server's cgi-scripts run, and ignoring less meaningful traces of usage.

The BT Digital Library requires users to log in to access certain information sources, to make it easier for it to ensure that it meets publishers' access restrictions. Some publishers, for example, require the Digital Library to serve its publications to UK-based users. Because users are required to log in, log files can be analysed to identify personal usage patterns.

These data sets are being used to develop collaborative filtering and user profiling services.

Qualitative feedback

The Digital Library strenuously encourages user feedback, although it has not yet carried out formal user surveys. Less formal methods of obtaining feedback are used, such as user meetings and publicity events, e-mailing users when they have had their password reset, and user feedback links on the server.

Usage of the Digital Library

Journals

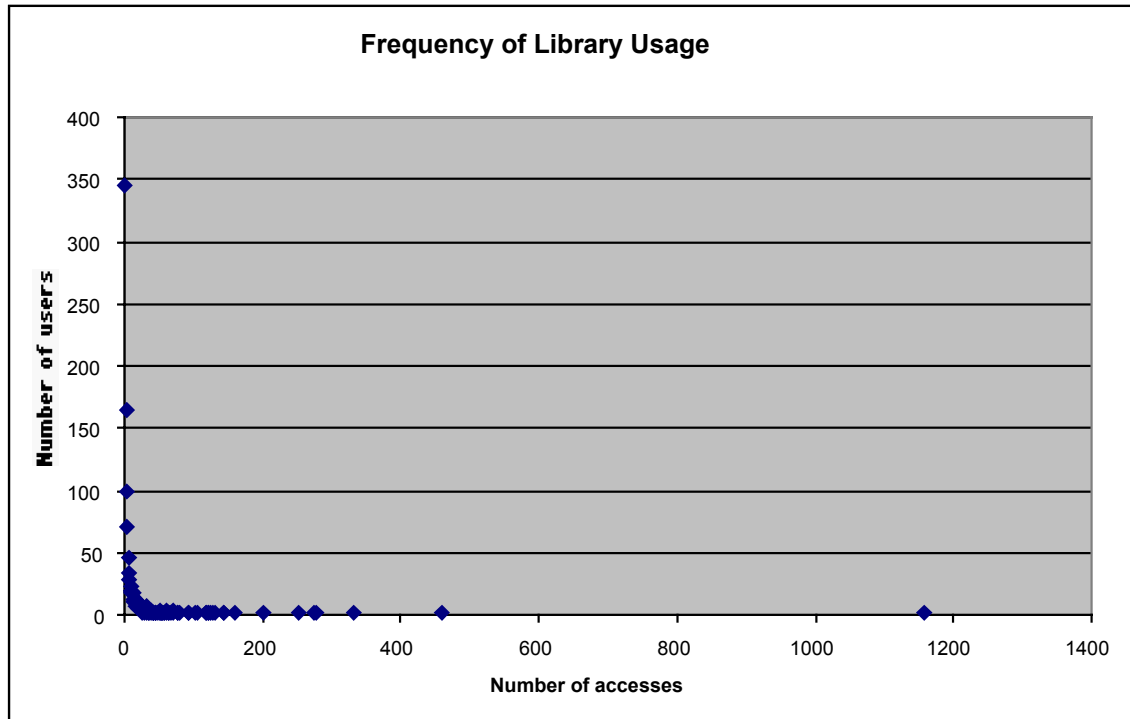
The BT Library offers 800 online journals to its user community, within the limits of the publishers' agreements. The collection of online journals is supplemented by a disintermediated document delivery service providing what might be called near-online journals. Articles not available online on the Library's server can be requested in a straightforward way and are usually delivered in two-three days.

A noticeable impact of the move from a physical to the Digital Library is in the distribution of the user base. In 1994 the Library served the research community almost solely. In spite of current awareness bulletins, which were distributed throughout the company, and a document delivery service to supplement this, approximately 90% of the Library's usage was from the Adastral Park site. By the end of 1998, when ACM, IEE, and IEEE journals became available on the server in addition to the in-house journals and a selection of titles from Elsevier, this figure had gone down to 61%. In 1999 the Library's collection was enhanced with the addition of material from ABI/Inform and the balance has shifted so that only 40% of users come from the BT's Adastral Park site.

As part of this study, the usage of the 3,500 potential users at Adastral Park was examined. These users are readily traceable, because they use relatively static IP addressing, allowing a more detailed study of individual usage.

In 1999, 1,091 users from BT ACE read 9,108 journal articles from the Digital Library 12,919 times. (These figures exclude journals from the ACM Digital Library and from the selection of other journals available only on the publishers' sites, where usage data is not available.) In comparison, the Library had 1,500 users registered for access to the physical Library and lent fewer than 8,000 documents in the same period.

Usage patterns in this data set reveal the expected kind of distribution, with most users using the Library a few times.



IEL

The IEEE's IEL product offers IEE and IEEE journals and conference proceedings in Adobe Acrobat format. These are received monthly and loaded onto the Library server, so that data on usage are available for study. In the BT implementation, user registration is not necessary to access these publications, so user analysis is limited to studying access by IP address. Data on usage are available from November 1998 which is too short a period to do more than speculate on seasonal variations beyond noticing the obvious drop in readership at the Christmas/New Year period.

A mean of 39 users read IEL papers each week, with a minimum of 7 and a maximum of 56. These users read 11 papers each, with a minimum of 4 and a maximum of 38.

The IEL collection offers a typical usage pattern, with 80%% of journal usage concentrated on 21% of the titles in the package.

ABI/Inform

ABI/Inform offers a wide range of management and trade journals. Some of these, such as *Harvard Business Review* or trade journals in the telecoms area of key interest to the Library's user base. ABI/Inform has been fully available on the Digital Library for two months which restricts the possibility for usage analysis.

During this limited period, a mean of 97 users a week have read ABI/Inform papers, with a minimum of 25 and a maximum of 214. These people have read 3 papers each, with a

range from 2 to 6 papers read per user during this limited period.

Elsevier

The Digital Library holds a collection of more than 20 Elsevier titles online. Unlike ABI, IEL, and ABI/Inform which offer a package of publications, our Elsevier collection is based on the set of paper journals the Library took in paper form. These were selected as being core journals, based on Library usage and on the Library's understanding of BT's research interests. Because access to these journals can be more easily recorded than to those in paper format, usage patterns can be tracked more easily, advising the Librarian which titles are no longer appropriate to hold in the Library's collection. In addition, the Library's document delivery system records the publishers of documents requested, allowing the Librarian to monitor new titles for inclusion in the Library. Three titles are heavily used, but even these are not accessed at all in 30% of weeks. BT's recent decision to stop research in the speech processing area is reflected in the drop in accesses to journals in these areas. In between these two extremes are the majority of journals that are used sporadically. In spite of the additional data on how frequently these journals are used, collection management is still difficult because much of the usage data is sparse.

Books

The Library has made a limited start to providing online books to its userbase. 24 computing books from O'Reilly on Perl, Java, Unix, and networking were made available in 1999. 87 users a week access one of the O'Reilly books online, with the number of users ranging from 8 to 179 in any one week. These books, serving as reference books for problem-solving, as well as textbooks are ideal for online publishing. They have certainly produced the most excited and positive unsolicited feedback.

Conclusion

The corporate librarian is pressurised to demonstrate their value to their parent organisation, leading to efforts to reduce costs and improve library usage. In BT's case, these pressures resulted in outsourcing labour intensive activities such as document delivery and replacing paper-based publications with online versions. Susan Rosenblatt is reported as commenting that "available information drives patterns of usage" [Odyzko, 1997]. BT's experience bears this out. Making more information available on the intranet increases library usage both by local users choosing online access in preference to using the Library in person and by remote users who previously had no practical means of access.

References

Busch, DD, Count your blessings, Internet World, June 1997

Mackie-Mason, JK, et. al, A Report on the PEAK Experiment, *D-Lib Magazine*, July/August 1999

Odlyzko, AM, Silicon dreams and silicon bricks: the continuing evolution of libraries, *Library Trends* 46 (no. 1) (Summer1997), pp. 152-167

Wright, MJ, Constituencies for Users: how to develop them by interpreting logs of web site access, *AAAI Spring Symposium on Intelligent Agents in Cyberspace*, March 22-24 1999, Stanford, CA, USA.

Zhang, Z, Evaluating electronic journals services and monitoring their usage by means of WWW server log file analysis, *VINE*, no.111, 1999 pp 37-42