

WOOD UTILIZATION RESEARCH DISSEMINATION ON THE WORLD WIDE WEB: A CASE STUDY

DANIEL L. SCHMOLDT
MATT F. WINN
PHILIP A. ARAMAN

ABSTRACT

Because many research products are informational rather than tangible, emerging information technologies, such as the multi-media format of the World Wide Web, provide an open and easily accessible mechanism for transferring research to user groups. We have found steady, increasing use of our Web site over the first 6- 1/2 months of operation; almost one-third of the Web pages accessed were publication abstract pages. By comparing subject areas of reprints requested and abstract pages viewed with the number of reprints requested, we can infer how well our research products are meeting our clientele's needs.

Government research programs are coming under increased scrutiny to justify their research direction, and to validate research project existence. One way to justify research is to monitor the needs of research clientele, and to enhance their access to and ability to adopt new technologies resulting from research activities.¹ Therefore, technology transfer efforts have become critical companions to research. These efforts can take many forms, including workshops, training seminars, newsletters, formal and specific technology transfer arrangements, professional meeting presentations and proceedings, trade journal articles, news

items, and peer-reviewed publications. Notably, the nature of research results is most often information rather than tangible products. Therefore, information technologies could provide an effective mechanism to transfer research results to potential users.

The World Wide Web provides a useful environment for communicating information. Many different types of information can be delivered: text, graphics, sound, images, movies, and most recently, interactive software applications. Basically, what the Web does is distribute information and technology that resides on a single computer to many other computers located elsewhere. This distributed information environment extends the role of the computer beyond traditional data and information processing

into communication technology, where network computers form a many-to-many communication medium.²

Our research work unit of the USDA Forest Service, Southern Research Station (Primary Hardwood Processing Products and Recycling), initiated a Web site in October 1995 to inform our clientele regarding our research objectives, accomplishments, and products.³ The primary goals of our research unit are to: 1) identify, evaluate, and develop new or improved automated primary hardwood processing technologies and hardwood products that make U.S. industry more competitive in domestic and foreign markets; and 2) develop and evaluate pallet repairs and other uses of discarded pallets to extend our timber resources. Within these overall goals, the unit has four principal areas of research: 1) develop improved and new automated primary processing technologies for hardwoods; 2) identify, evaluate, and develop new or improved products that use abundant, nonselect hardwood sawtimber; 3) evaluate the use of short hardwood lumber and green dimension made from abundant, low- and medium-quality hardwoods to supplement shortages of hardwood materials for fine hardwood uses; and 4) identify and evaluate new or improved recycling methods to repair

¹Schmoldt, D.L. 1992. Bringing technology TO the resource manager... and not the reverse. *In: ASPRS/ACSM/RT 92 Convention: Monitoring and Mapping Global Change*, vol.5. Am. Soc. for Photogrammetry and Remote Sensing, Am. Congress on Surveying and Mapping, Bethesda, Md. pp. 62-74.

²Harnad, S. 1991. Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge. *The Public-Access Computer Systems Review* 2(1):39-53.

³Our Web site can be visited at the URL, <http://www.se4702.forprod.vt.edu>.

The authors are, respectively, Research Forest Products Technologist, Forestry Technician, and Project Leader, USDA Forest Service, Brooks Forest Prod. Center, Virginia Tech, Blacksburg, VA 24061-0503. This paper was received for publication in June 1996. Reprint No. 8541.

©Forest Products Society
Forest Prod, J. 47(6):25-31.

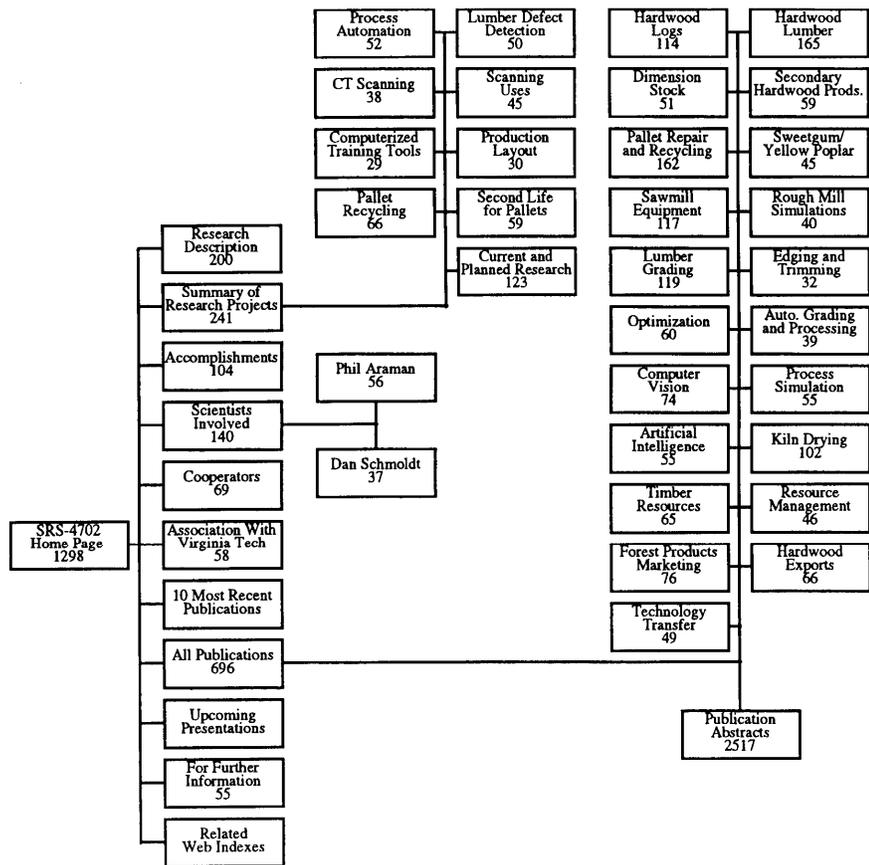


Figure 1. — The hierarchical organization of pages on the Web site begins with the top-most level at the left. Numbers of accesses for each page appear in each box.

pallets for reuse and to use discarded pallets for other products. The Web server has operated 24 hours a day since its inception (October 10, 1995), providing a large volume of information to site visitors about the research unit, including research objectives, scientific staff, current and planned research, collaborative research programs, upcoming presentations, and publication lists with abstracts.

Demographic and usage information about visitors to the Web site was logged and analyzed for the first 6-1/2 months of operation (October 1995 to April 1996). This information may be useful to future Web site developers by providing them with indications on what they can expect in terms of user activity and user interests. Usage information also provided our research work unit with valuable feedback (albeit from the Internet-literate

segment of our clientele) regarding current research direction and dissemination of research results. While not available during that initial period of operation, more recent enhancements to our site allow visitors to: 1) fill out and send us customer feedback forms; 2) download publications of reprint quality; and 3) search publication abstracts and titles for specific content.

WEB SERVER SETUP AND OPERATION

WEB DOCUMENT ORGANIZATION

From October 1995 to April 1996, the content of our Web site was organized as presented in **Figure 1**. The current organization has changed slightly since that time. Numbers appearing in this illustration indicate the number of accesses that those particular pages received during the period in which records were kept (see the next section for a description of visitor-days). At the top level resides the home page with text links to the second level of Web pages and also

graphical links to the USDA and Forest Service home pages and to home pages for Virginia Tech, the Hardwood Utilization Consortium, and ForestNet. Many of the second-level links are terminal pages. The Cooperators and Scientists Involved pages, however, link to other individuals and institutions. The Summary of Research Projects and Publications pages provide links to much of the remaining information contained at the Web site. This information includes publication lists, abstracts for all publications listed, and forms for e-mail requests of those publications not yet available on-line.

Visitors can now obtain publication reprints for more than 210 publications in 21 different subject areas. During the 6-1/2 months that we analyzed visitor usage, we had approximately 180 publications; this, of course, was a continually changing value. Most publications are multiply listed within two or more areas. This makes it easier for users to find what they are looking for because, until recently, no search facility was provided by our site. There are two ways to obtain publication reprints (**Fig. 2**). First, if it is available, visitors can download a portable document format (PDF) file of the publication. Then, by using Adobe⁴ Acrobat Reader, they can read the publication on their own computer screen and print it on their printer. Printed results are higher quality than photocopying and are of equal or better quality than traditional reprints. Second, if a PDF version is not yet available for a publication, visitors can fill out a simple on-line form that is easily e-mailed to our site editor. The requested publication is located, photocopied, and postage mailed to the requester.

WEB SERVER LOG FILE

While the Web server is running, it records all server activity in a log file. Each entry in this log file records: the date and time of access, the domain name of the accessing machine, and the files downloaded. A domain name is a mnemonic (e.g., www.se4702.forprod.vt.edu) for an Internet Protocol (IP) address (e.g., 128.173.241.12). Every machine connected to the Internet has a unique IP address that identifies it, and that allows message packets to find their way from sending to receiving machine. Each Web document (page) that is accessed by a user may result in 10 to 15, or more, files being immediately transferred. Each one

⁴Adobe Systems Incorporated, Seattle, Wash.

of these file transfers is recorded as a separate entry in the log file, and constitutes what is typically referred to as a hit. All the text in a Web document is transferred as a single file, but many of the graphic objects that one sees on a page are separate graphic files (often in GIF file format). Consequently, the number of hits that a Web site receives is not a valid indicator of site activity, but rather how busy (in the sense of cluttered with detail) each page is.

For the purposes of describing and analyzing visitor usage to the Web site, we have chosen to define an alternative to hits. A single user session is what we would ideally like to record. However, because any user access may include dozens, or hundreds, of hits over an extensive period of time, there is no way for us to reliably determine if this period of activity contains multiple sessions or a single session spread out over time. Our alternative is called a visitor-day. When any user accessed any one of the web pages on a given day it was considered a visitor-day. If the same user accessed the pages more than once in a single day it was still considered a single visitor-day. Unfortunately though, because some Internet connections (in particular, those involving modem pools) generate random IP addresses, a single user accessing the site at different times during the same day might appear as two separate users, and hence would be recorded as two visitor-days. While not entirely accurate, this terminology is a more realistic measure of user activity than hit counts and it enabled us to perform an analysis of visitor usage.

ANALYSIS OF USERS AND USAGE

Demographic and usage data for users was taken from the log file generated by the Web server. These ASCII log files were parsed into a spreadsheet, where extraneous hits were removed. Extraneous hits included entries representing graphic and image files served and all entries associated with access by our unit personnel. All pages accessed by users were retained in the spreadsheet, but visitor-day access information was condensed out of the larger data set. Information on user type, Web pages accessed, and reprints requested was analyzed from October 1995 through April 1996 (approximately 6-1/2 months). Because the server was set up in October, this was considered a partial month with consis-

1. Conners, Kline, Araman and Brisbin. 1992. Reflections on the development of a machine vision technology for the forest products industry. *Proceedings, Knowledge Based Expert System for the Furniture Industry*. 6/1-6/20.
2. Araman, Schmoldt, Conners and Kline. 1995. Scanning system technology worth a look. *Wood and Wood Products*. 100(5): 138-142.
3. (130k) Occeña, Chen and Schmoldt. 1995. Procedures for geometric data reduction in solid log modeling. *MU-IE Technical Report 019507*. 5 pp.
4. Schmoldt, Li and Araman. 1995. A CT-based simulator for hardwood log veneering. *Proceedings, 2nd International Workshop/Seminar on Scanning Technology and Image Processing on Wood*. 65-75.

Check the above publications which you would like copies of, fill out name and mailing address below,

and press this button **Request Copies** to request copies of publications.

To clear the form, press this button: **Clear Form**

Figure 2. — An example of a partial publications list illustrates: 1) on-line publications (PDF files); and 2) a reprint request form.

tent hits starting around October 10th. The amount of work involved in filtering these log files and keeping track of all the resulting information manually is now prohibitive. Instead, we now use a free software program that automatically analyzes log files and categorizes similar information, except visitor statistics. We do, however, continue to track user requests for reprints manually.

DEMOGRAPHIC ANALYSIS

Domain names appearing in the log file were examined to provide some indication of each user's affiliation. All users were placed into one of two categories: North American users or International users. North American users were subdivided into one of the following user types: government, non-profit organization, educational institution, Internet service provider, commercial, or unknown. Because not every machine on the Internet has a fully qualified domain name, we were unable to identify all site visitors, and therefore had to categorize some visitors demographically as unknown. International users were classified according to their country codes and then grouped into the continents: Europe, Asia, Africa, South America, and Australia / New Zealand.

The Internet has grown in a somewhat haphazard way that has partially hindered our ability to clearly establish user

demographics. While a user's country is readily established, we are less confident about the usefulness of domain names for categorizing users in North America. For example, we know of one case in particular where an educational user has accessed our site for someone with a forest products business who did not have Internet access. Even our unit's domain name contains .edu (educational), but we are actually a government site. Nevertheless, we feel that the domain-name identification of users that we were able to discern is adequate for this analysis, even if actual demographics are not entirely accurate.

USER ACTIVITY ANALYSIS

There were two areas of user activity that we examined: 1) pages accessed; and 2) publication reprints requested. We tallied the total number of accesses for each page of the Web site. These data were relatively easy to extract from the log files after extraneous hits were eliminated. This page-access statistic was expected to indicate the most popular or interesting or important pages across all user activity. In particular, we looked at which abstract pages were accessed most frequently within the various subject areas. We also recorded information about each user request for publication reprints. In addition to the requesting user's name, postal mail address, and e-mail address,

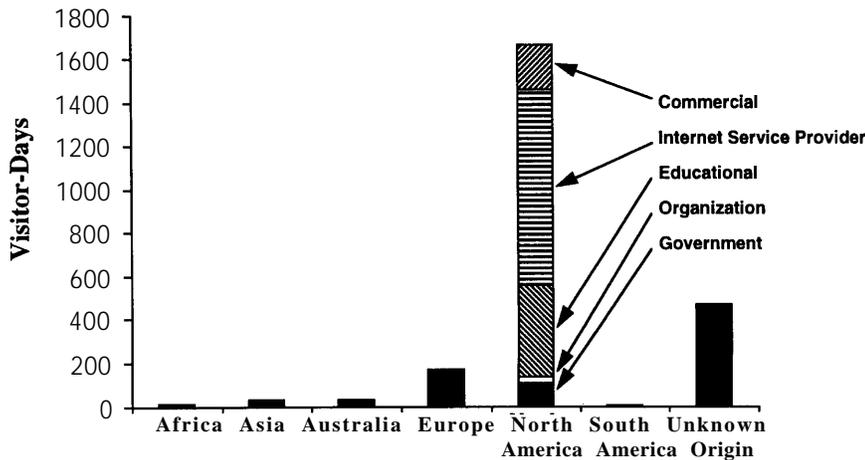


Figure 3. — Visitor-days are categorized by users' continents and by access type for North America.

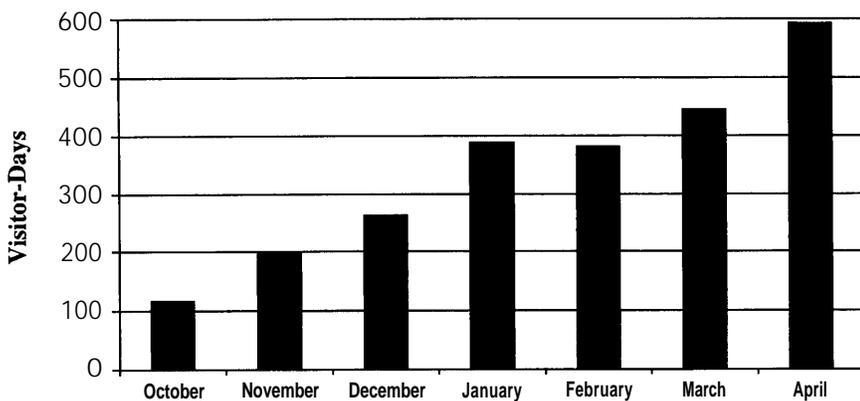


Figure 4. — The number of visitor-days is grouped by month for the 6-1/2 months of initial Web site operation.

we noted the publications requested. From this information, we kept track of how often subject areas were requested by users, i.e., the frequency of interest in various subject areas at the site. We also kept track of how many reprints were requested from each subject area, i.e., the intensity of interest in each subject area. The combination of frequency and intensity was expected to indicate those areas of our research program that are most valuable to our clientele.

There are two difficulties associated with examining activity and interest for particular subject areas: an availability bias and a saturation effect. These are both related to unequal numbers of publications within the subject areas. Abstract page access and total numbers of reprints

requested are biased by the number of publications available in each subject area. That is, subject areas with greater numbers of publications will, naturally, tend to have more reprints requested. For example, some subject areas, such as hardwood lumber, are very broad because many publications are related to some aspect of hardwood lumber. One way to overcome this availability bias is to normalize subject area counts, for abstract accesses and reprint requests, by the number of publications in each subject. Unfortunately, this approach would ignore the fact that there is also a saturation effect. That is, someone will tire of examining abstract pages after a short time, or will only request the number of reprints that he or she can reasonably

expect to read and absorb. Users will reach a point where their attention is saturated irrespective of how important or interesting they feel the subject is.

As an example, consider subject A with 8 publications available and subject B with 25 publications available. Given that A and B are equally interesting and important to the user, requesting 4 reprints from A, i.e., one-half of the total number available, does not mean that the user will also request an equal proportion from B, i.e., 12 or 13. The latter figure would exceed the saturation limit for most users. Therefore, we need to normalize subject area counts by some value other than the total number of publications in each subject area, so that they incorporate both availability bias and the saturation effect.

We examined several mathematical relationships to adjust for both availability and saturation. Each relationship was a function of the number of publications in a subject area. These included the square root, the natural logarithm, and the base 2 logarithm. They each gave relatively similar results, so we selected the base 2 logarithm of numbers of publications in each subject area to normalize subject area counts. Our intuitive rationale for making this selection is that given indifferent interest in a subject area, i.e., neither strong disinterest nor strong interest, a user's curiosity will produce a binary sampling of publications on average. Using the example in the previous paragraph, we would normalize counts in subject A by $\log_2 8 = 3$ and in subject B by $\log_2 25 = 4.64$. Then, 4 reprints in subject A would normalize to $4/3 = 1.33$, and an equivalent number of reprints (normalized) in subject B would be, $4.64 \times 1.33 = 6.17$. This normalization approach reduces counts based on the number of publications in each subject area (to incorporate the availability bias), but by realizing that those counts are also subject to a saturation effect, normalization employing the base 2 logarithm of subject area publications does not reduce counts excessively.

RESULTS

Site visitor-day usage is presented geographically in **Figure 3**. By far, the greatest use was by North American users (70%), including Canada, Mexico, Central America, and the Caribbean. Users of unknown origin accounted for approximately 20 percent of the visitor-

days. Based on the initial unknowns that we were eventually able to identify, we suspect that most of the remaining unknowns are actually of North American origin. This pushes North American access to near 90 percent. Within North America, Internet service provider users account for more than 50 percent of total use. As previously noted, we suspect that these user types are either personal Internet accounts or small businesses, because those are the entities that contract with Internet service providers.

Over the 6-1/2 months of initial operation, visitor-days increased steadily (Fig. 4), except for a slight drop in February and surprisingly substantial jumps in usage during January and April.

From Figure 1, it is apparent that publication abstract pages were accessed more than any other pages (almost 1/3 of 7,768 page accesses). This is almost twice as often as the top-level home page itself was accessed (1,298). This may

seem somewhat counter-intuitive, but by using Internet search engines, users can go directly to Web pages of information that they desire without necessarily wading down through a site's home page and other intervening pages. Figure 5a shows the distribution of abstract accesses by subject, where the values were normalized by the base 2 logarithm of publications per subject. Hardwood Lumber and Hardwood Logs are rather generic and can be ignored here. Therefore, Computer Vision, Marketing, Automated Grading/Processing, and Pallet Repair/Recycling were the most popular subjects for abstract viewing.

During the 6-1/2-month period, 628 publication documents were requested and mailed, nearly 100 per month. These requests were made by a total of 92 users, which is approximately 7 publications per user-request (we define user-request as a user who has requested something, as opposed to just a visitor). These 92

user-requests can be categorized by subject area (Fig. 5b). As noted previously, most of our 180 publications are multiply categorized, therefore, the total number of categorized user-requests exceeds the 92 actual user-requests. For example, a single user-request may ask for eight reprints: two in one subject area, three in a second area, and three in a third area. This user-request would then be tallied once in each of those three different subject areas in Figure 5b. Aside from the pseudo-generic subject areas of Hardwood Lumber and Hardwood Logs, the subjects, Pallet Repair/Recycling and Lumber Grading were requested more often than any other subject areas.

Subjects requested should not be confused with the total number of reprints requested, however. To examine total numbers of reprint requests, we can categorize the 628 reprints requested (Fig. 5c), after normalizing. More Pallet Repair/Recycling publications were re-

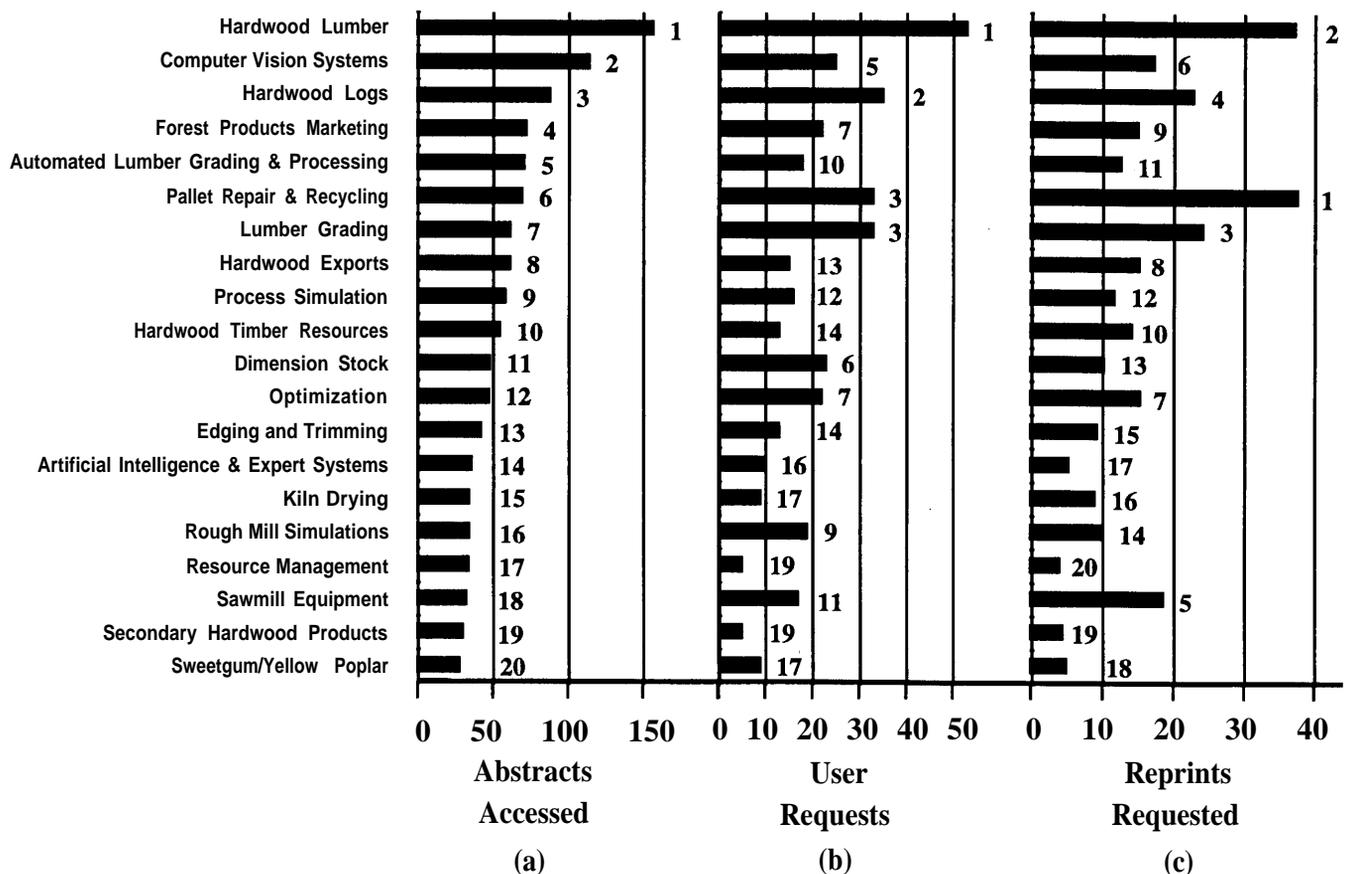


Figure 5. — Abstract pages accessed (a), user requests for reprints (b), and reprints requested (c), are categorized by subject area. Numbers adjacent to the bars are ranks. The values in (a) and (c) are arrived at by dividing actual values by the base 2 logarithm of publications in each subject.

quested than either Hardwood Logs or Hardwood Lumber even though the latter two are relatively generic subjects. Lumber Grading, Computer Vision Systems, and Sawmill Equipment also received many reprint requests. While Dimension Stock and Rough-Mill Simulations were relatively common (upper 50th percentile) *subjects* for user-requests (**Fig. 5b**), relatively few *publications* were actually requested (**Fig. 5c**) for these subjects, which means that many users requested only one publication for these subjects, whereas they requested multiple publications in other areas. Therefore, while those two subjects were requested with a relatively high *frequency*, the actual numbers of reprints requested indicated that the *intensity* of interest for those two subjects dropped relative to other subjects. We can also compare **Figures 5a** and **5c** and note that the subjects Computer Vision, Marketing, and Automated Grading/Processing rank high for abstracts viewed, but drop somewhat in reprints requested. Pallet Repair and Recycling and Lumber Grading shift in the opposite manner.

Requests for reprints contained more demographic information about users than was available from domain-name analysis alone. From **Figure 6** we can see that most user-requests were denoted as either business related or personal. Although we asked that reprint requests include business information, we expect that many of the personal requests were actually unidentified business requests. Relatively few requests came from educational institutions, but a substantial number came from outside North America.

CONCLUSIONS

Visitor-day usage increased over the short time during which we examined the Web site's activity. As more forest products professionals gain access to the Internet, we expect usage to continue rising, at least into the near future. We expect that site activity will level off eventually at some plateau that is commensurate with both Internet access by forest products professionals and with general Internet activity.

We do not have detailed and accurate demographics on this Internet segment of our research clientele or how they might relate to our larger clientele population. This demographic gap is partly due to the current technology and partly due to pri-

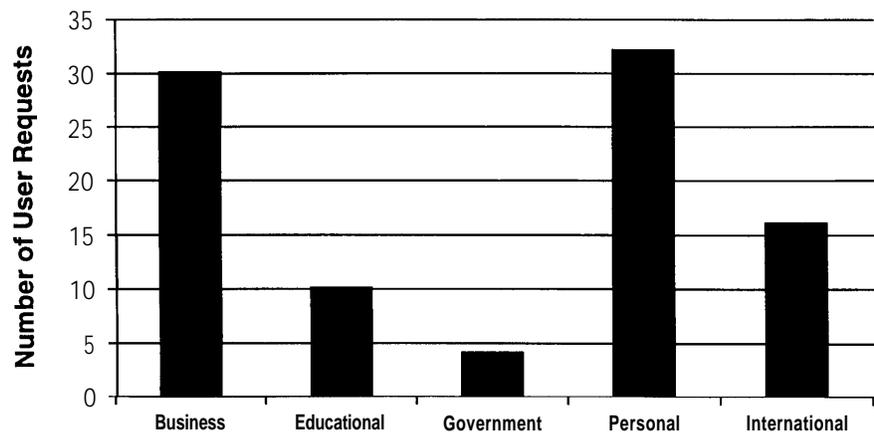


Figure 6. — Users requesting reprints are categorized according to domain name demographics within North America and internationally.

vacy issues. Nevertheless, the mix of government, educational, personal, business, and international connections that we received is not dramatically out of line with what we might expect for our overall clientele group. Therefore, we feel that it is possible to draw some reasonably valid conclusions from this cyber-survey.

On the publication order pages, we advise users to request no more than 20 publications at once. While user requests have adhered to that advice, there is a tendency for users to request many (7 pubs per request, on average) because we have made ordering almost effortless. Once a publication order has been received by us, the appropriate publications have to be located and copies have to be generated and mailed. This process is time consuming and expensive. Consequently, we have recently begun to make publications available electronically, as PDF files, which is a more cost-effective method. The time required to scan and process each publication into a PDF file varies greatly; however, results are equal to or better in quality compared to professional reprints.

We have continued to keep track of reprint requests even after April 1996. In the 14- 1/2 months from October 1995 through December 1996, we have sent out over 2,900 publications to more than 375 users, as a result of Web site visitation. This amounts to approximately 50 reprints per week. In March 1997, however, we mailed out over 530 reprints, with an additional 270 downloaded as

PDF files. The total time required to locate, copy, and mail all these documents underscores our interest in making publications available on-line.

By closely examining **Figure 5**, we can gain some insight into how well our research is being received by this Internet segment of our user community. The subject of Pallet Repair/Recycling is of interest to many of the users requesting reprints. In addition, many reprints were actually requested for publications in this area. This association seems to indicate that the topic has frequent interest (a popular subject, **Fig. 5b**) and the research being conducted is noteworthy and valuable (high reprint popularity, **Fig. 5c**). However, other subjects such as Rough Mill Simulations and Dimension Stock appear to have a negative association, i.e., high subject interest but relatively few research reports requested. This might indicate that the listed publications do not meet clientele needs, and that we might want to revise our research direction or, alternatively, include Web page links to other Internet sources of this information. In this particular circumstance, for example, the Robert C. Byrd Hardwood Technology Center has continued and extended rough mill simulation work begun initially by our work unit, and therefore we include a link to their site for more information on this subject.

In contrast, Hardwood Exports has an opposite association; although it is not extremely popular as a subject (13th), research reports in that subject were re-

requested more frequently (8th) than expected given its popularity. This mix of relatively high intensity of interest and relatively low frequency of interest for Hardwood Exports might indicate that those publications have done a good job of informing the users with an interest in that area. Abstracts accessed (**Fig. 5a**), may give some indication of users' desire to be informed about certain subjects. But, while this desire is high for Computer Vision, Marketing, and Automated Grading/Processing (**Fig. 5a**), the techni-

cal details presented in the actual publications may not adequately satisfy that need for our user audience. Hence, the number of reprints actually requested (**Fig. 5c**) for those subjects is lower than might be expected from abstract viewing statistics. In light of these results, we might want to re-examine the types of studies being conducted in some of these research areas, as well as how research results are reported, e.g., the number of technical vs. popular articles that are published.

The relatively large number of reprint requests from business users (**Fig. 6**) is both surprising and encouraging. In the forest products community, this group constitutes our largest segment of users. Their interest in our work unit's publications indicates that we are conducting research that is important to them. Also, the relatively good agreement between frequency and intensity of interest in most research subjects suggests that our studies have effectively addressed technology transfer to our user groups.