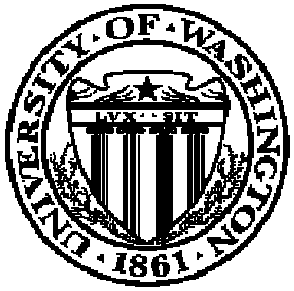


Information School

Helping Users to Use Online Resources to Resolve Information Technology Problems: An Opportunity for the Semantic Web

Melody Y. Ivory, Andrew P. Martin, Rodrick Megraw, Beverly Slabosky

Technical Report IS-TR-2005-08-02

August 26, 2005

Information School
University of Washington
Suite 370 Mary Gates Hall, Box 352840
Seattle, WA 98195-2840

Phone: (206) 543-1794
Fax: (206) 616-3152
Email: info@ischool.washington.edu
Web: www.ischool.washington.edu

Helping Users to Use Online Resources to Resolve Information Technology Problems: An Opportunity for the Semantic Web

Melody Y. Ivory, Andrew P. Martin, Rodrick Megraw, Beverly Slabosky
Information School
University of Washington
+1 206-616-6110
{ myivory, am1982, remegraw, beverlys }@u.washington.edu

ABSTRACT
Users need help systems to support their use of complex information technology (IT); however, several studies have shown help systems to be inadequate. To identify ways in which to improve help systems, we administered an online questionnaire to 107 IT users from diverse populations. The questionnaire probed users' current perceptions and use of help systems that are within software applications, web sites, and mobile devices. A major finding was that two-thirds of users reported that they use web-based content to help them to resolve IT problems; use of web-based content superseded their use of printed and electronic documentation and their communication with technical support specialists and other people. Our study also revealed accessibility issues with online questionnaire systems; we describe specific problems and how we addressed them. Based on our findings, we propose the development of a portal system to harvest help content from various sources, organize intelligently the content, and enable users to search or browse for help on specific IT problems. We consider the system to be an ideal application for the Semantic Web and advocate research and industry collaboration to develop the necessary infrastructure.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Search process.* H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Training, help, and documentation.* I.2.7 [Artificial Intelligence]: Natural Language Processing – *Text analysis.*

General Terms
Documentation, Human Factors.

Keywords
User Assistance, Online Help, Software Applications, WWW, Mobile Devices, Accessibility.

1. INTRODUCTION
The complexity of information technology (desktop applications, large-scale web sites, mobile computing devices, etc.) necessitates the development of help or user assistance systems to support users [1, 2, 3, 4, 5]. Since the early days of user interfaces, help systems have been at the forefront of the technical writing, software development, and human-computer interaction fields. Consequently, help systems have evolved from printed manuals to electronic manuals, context-sensitive help, wizards, agents, and collaborative, real-time systems [6, 7, 8, 9]. Nowadays, the Web and search engines enable information technology (IT) users to access an even broader set of user assistance resources.

Although help systems are prevalent, studies have shown that existing systems are not effective [10, 11, 12, 13]. Problematic aspects of help systems include: (1) missing or inadequate content, (2) voluminous or poorly organized content, (3) lack of conceptual information or examples, (4) use of vocabulary which users do not understand, (5) voluminous search results, and (6) unusable interfaces. Due to these issues, users are reluctant to use help systems and consequently consult them as a last resort [14, 15, 16]. Compounding the inadequacy of help systems is the fact that they are difficult to develop and there are few resources available to facilitate their efficacy [13, 17, 18].

Innovations are still needed in the area of user assistance. To identify potential innovations, we are conducting a comprehensive study to understand the current state of help from the perspectives of users, technical support specialists, and user assistance developers (i.e., professionals who create content, functionality, or interfaces for help systems). This report addresses the first part of our study—an online questionnaire that we administered to 107 representative IT users. The questionnaire probed users' current perceptions and use of help systems within three contexts:

1. Software applications that are used on personal computers (i.e., desktop, laptop, or tablet computers);
2. Web sites, applications, or search engines (the Web); and
3. Mobile computing devices or small, portable computers that allow users to store, organize, and access information that is stored on the device or online (e.g., PDA, pocket PC, iPod, handheld PC, or electronic book reader).

We begin with a discussion of related studies on help systems. We then describe our help user study and findings. A major finding was that two-thirds of users reported that they used Web-based content (frequently asked questions or FAQs, knowledge bases, etc.) to help them to resolve IT problems. Use of Web-based resources superseded their use of traditional resources like functionality that is embedded within applications or sites and printed and electronic documentation that comes with a product. Use of Web-based resources also superseded their communication with technical support specialists and other people. Our analysis revealed three types of help seekers, who are distinguished by the degree to which IT is integrated into their everyday lives; we describe the three user groups. Our study also revealed

accessibility issues with online questionnaire systems; we discuss these issues and our solutions.

Based on these findings, we propose the development of a help portal system to streamline help seeking via the Web. Consistent with the premise of the Semantic Web [19], the system would harvest help content from various sources, such as online help, web sites, and discussion groups, organize intelligently the content, and enable users to search or browse for help on specific IT problems. We describe the envisioned system and interface. We also discuss ongoing and future work with respect to understanding the current state of help and developing the portal.

Infrastructure for supporting the proposed user assistance portal is not in place. We consider this system to be an ideal application for the Semantic Web. Research and industry collaboration is needed to standardize the specification of help content and, consequently, to facilitate its aggregation and reuse. Perhaps our findings will spur work to streamline and simplify help seeking.

2. RELATED STUDIES

Users have been frustrated with help systems for a long time. From printed documentation to electronic documentation, online help, and the intelligent agent "Clippy" [20], help users continue to be annoyed and turn away from help systems that do not understand their problems. We describe related studies below.

Previous studies indicate that users of online help do not typically use the help menu, but they will respond to help information that is simultaneously context-specific, easily available, obvious to invoke, useful, and non-intrusive [21]. One study revealed that too much information is crammed into help systems [22]. Consequently, users do not know how to begin to interact with the application, misunderstand the meaning and implications of pushing certain buttons, do not know what they can do with the system, and have difficulty finding information. In another study, interviews and observations showed that most users have online help anxiety and avoid using it [23].

A fairly recent study of help system features and guidelines showed that users preferred five help system features—an index, search functionality, a table of contents, balloon help, and hyperlinks within help contents—and preferred help systems that were easy to understand, contained procedural (step-by-step) content, were unobtrusive, accurate, complete and consistent, and used the user's language [9]. The study also showed that few help systems implement effectively these features or conform to these guidelines. Other recent studies document the inadequacy of help systems for assisting users with accomplishing tasks that range from word processing to HTML coding, programming, and image editing [10, 11, 24, 25].

Studies of commercial help systems have revealed similar user difficulties. In one study, a software company called a random sample of customers within a few months after they had purchased one of the company's applications [26]. They asked customers about how useful the printed help and the online help were to them. Overall, users did not find the information for which they were looking. They found the information within help screens to be incomplete, so they wanted additional detail or information. The structure of help made it difficult for them to find or locate the needed information, and the terminology of the application was different from that of the users (they did not know what the software called the problem or feature). Users had to go through numerous links to find useful information.

In the early 1990s, Apple researchers conducted a two-year study on the types of questions people ask while seeking help for their products [27]. Users found it difficult to find information or the help system did not return relevant information. Users also did not like switching contexts from work to help functions. They also found the help interface to be difficult to navigate and the help information to be poorly displayed.

In this study, we were interested in discovering whether or not these well-documented problems have been resolved. Furthermore, unlike prior studies, we wanted to assess the effectiveness of help systems that are developed for web sites and for mobile computing devices. In addition, we were interested in understanding how users from diverse populations use help.

3. HELP USER STUDY

We describe our help user study in this section. We begin with a discussion of the study methodology. Early in the study, we discovered that the commercial online questionnaire system that we were using was not accessible to users who employ screen reading software. In response, we developed an accessible questionnaire system, which we describe. We then describe study participants, present findings, and discuss study implications.

3.1 Methodology

We developed a comprehensive questionnaire based on the major themes that emerged from the prior help user studies. Common themes included: non-intuitive terminology and organization, too broad or too specific content, not knowing how to access help, navigation difficulty, and unusable interfaces. We asked users a total of 45 questions about their use of help within three domains: software applications, web sites, and mobile computing devices. We asked specific questions about how they use help, which tasks they perform when they seek help, their satisfaction with help, in which format(s) they prefer to seek help (e.g., print, online, calling someone, etc.), and their suggestions for ways in which to improve help systems. The questionnaire also asked for extensive demographic information (age, gender, income, household size, education, etc.) and background information on computer, Internet, and mobile computing device use.

We used the SurveyMonkey system [28] to implement and facilitate the online questionnaire. Initially, we created two identical SurveyMonkey questionnaires (one for the special population and one for the general user population; discussed in Section 3.3). We conducted two rounds of pilot studies, during which we observed a user as he or she completed the questionnaire and recorded completion time and issues that arose. After completing the questionnaire, the user provided feedback, which helped us to simplify the questionnaire. We reduced technical jargon, condensed questions, and clarified questions that users did not understand. After the second pilot study, we launched the questionnaire. Shortly afterwards, we discovered that the SurveyMonkey system was not accessible to users who used screen reading software, in particular JAWS for Windows [29]. Consequently, we developed a custom questionnaire system, which we describe in the next section.

We collected questionnaire responses from June to September 2004. Based on the questionnaire data, we generated profiles of users' needs and satisfaction by correlating their satisfaction levels with their demographic data (age, gender, education attainment, income, etc.). We standardized the data (i.e., converted responses to z-scores. We then ran the K-means

clustering algorithm [30] on the data to identify groups of users who had similar experiences and satisfaction with help. We present findings, based on three user groups that we identified.

3.2 The Need for an Accessible Questionnaire

Users who used the JAWS screen reading software were unable to navigate the input forms within the questionnaire. In addition, JAWS could not read the labels for answers. After a request to the SurveyMonkey team for the addition of specific accessibility features went unanswered, we investigated alternative technologies. Our initial survey of commercial and open source online questionnaire systems revealed that few of them went beyond the bare minimum of providing alternative text for images. We concluded that we needed to either develop our own solution from scratch or modify and existing open source tool.

We examined several open source projects and settled on a project called PHP Surveyor [31]. PHP Surveyor had the most complete set of questionnaire functionality of the tools that we considered, but it had only a few accessibility features and a substandard user interface. We modified a copy of PHP Surveyor to improve its user interface and to implement recommendations from accessibility guidelines published by W3C [32] and from an assistive technology expert at the university. Our changes were primarily structural (e.g., limiting the use of images, adding labels and tab indices for form objects, adding summaries for tables, and enabling keyboard shortcut keys for navigation) and aimed to improve the navigation of pages, forms, and tables. Even though we followed W3C recommendations to the best of our ability, our accessibility expert found twelve additional issues which we needed to be address to make the questionnaire compatible with JAWS. Recommendations concerned making the labels, table summaries, and button names more understandable.

We presented the modified questionnaire to a class of novice JAWS users and found that not a single user was able to complete it. While expert users of JAWS were able to navigate the questionnaire (four completed it), novice users had difficulty with forms when they used the JAWS forms mode (a specific feature of JAWS to assist users with entering data into online forms). Our observations of novice users and their feedback revealed additional changes that we could make to improve the questionnaire's intuitiveness. Users' recommendations included: presenting the questionnaire as a single page to reduce the time that they spent navigating between pages and making questions easier to navigate in JAWS forms mode.

As a specific example, users wanted keyboard shortcut keys to work differently than they are implemented currently in Internet Explorer. They wanted the shortcut key to CLICK the hyperlink with which it is associated. In Internet Explorer, the shortcut key just shifts the FOCUS to the hyperlink; the user is required to press the Enter key to click the link. We explored various techniques to make Internet Explorer work like they wanted it to work, but did not find a reliable solution. Consequently, we included detailed text within the shortcut key description: "Press Alt+s followed by Enter," rather than just "Alt+s."

Our experience demonstrates both the lack of adequate accessibility support within commercial questionnaire products and the dedication that is required to navigate accessibility guidelines and understand the use of specific assistive technology in practice. In future work, we plan to re-implement the accessible questionnaire system and distribute it freely so that people who use screen readers can participate in online studies.

3.3 User Populations and Recruitment

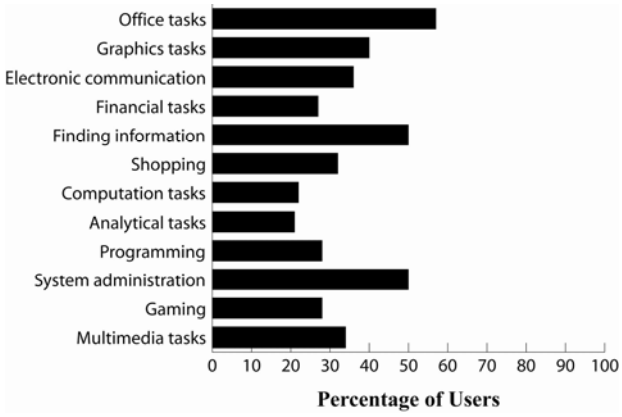
Our primary goal was to gather input from a diverse population of IT users, including people who are considered "traditionally underrepresented." Thus, we specifically targeted racial/ethnic minority groups (i.e., African Americans, Native American Indians, Latinos, etc.) and people who speak English as a second language, have low incomes, and have visual, cognitive, or other impairments. By ensuring that under-studied groups are included in our study, we can assess how well help systems serve a broader group of everyday Internet and IT users. Also, by including users from these under-studied groups, we have the opportunity to find out whether user assistance systems are accessible to users who have non-traditional needs.

We recruited IT users from the general population of Seattle, WA to participate in the study. We had three recruitment strands: (1) recruitment targeting the general population, (2) recruitment targeting diverse users, and (3) recruitment targeting blind users who use screen readers to navigate web pages. We sought to recruit fifty general population users (respondents to the first questionnaire) and fifty special population users (respondents to the second and third questionnaires).

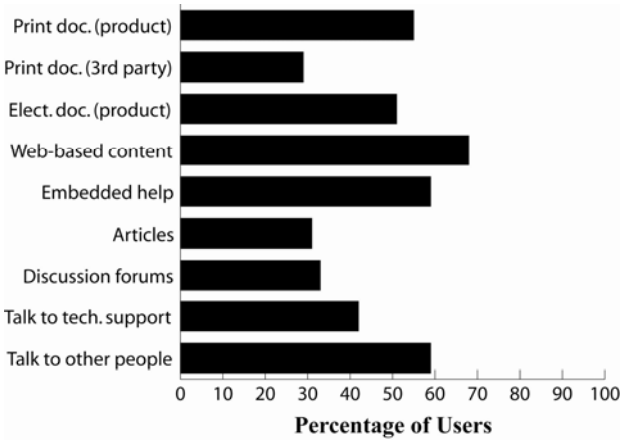
To accomplish our objective of recruiting from diverse user populations, we targeted organizations that served the needs of each group. For the general population, we posted flyers at libraries and community centers in the Seattle/King County area. We also posted information within Craig's List for the Seattle area; Craig's List is an online community for posting classified advertisements [33]. For the special population, we contacted computer labs, organizations, and libraries which served underrepresented populations. We compensated study participants with a $10 Amazon.com gift certificate.

3.4 Participants

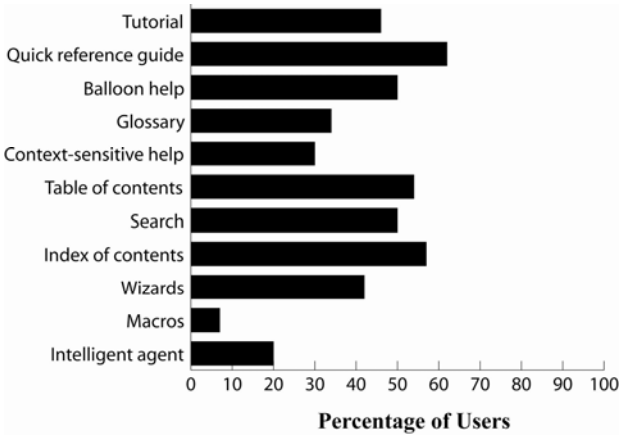
Our recruitment efforts resulted in 107 questionnaire responses. Of the 107 responses, 49 participants (46 percent) were from our special populations: (1) users who have some physical or cognitive impairment (24%), (2) users who are non-native English speakers (23 percent), (3) users who have low incomes based on U.S. government guidelines (14%) [34], and users who are from racial/ethnic minority groups (12%). Our user population was 60 percent female and had a median age of 35; 64 percent of users had a Bachelors degree or higher. Of particular note, 12 percent of the users had a visual impairment and another 5 percent had a learning impairment.



(a)



(b)



(c)

Figure 1. Tasks for which users seek help (a), ways in which they seek help (b), and use of embedded help (c).

Users were experienced with computers and the Internet: 98 percent used a computer and accessed the Internet at least once a day, half had been using computers for 11 years or more, and half had been using the Internet for 8 years or more. They used desktop computers (91%) with Windows operating systems (95%); some users used laptops (45%) and Macintosh operating systems (20%). They used computers for work or school (91%) and leisure (87%). They used the Internet for work or school (84%) and leisure (90%); 62 percent used a telephone modem to connect to the Internet. Mobile computing device use was not popular among our respondents: 62 percent had never used a mobile device, but 19 percent used one multiple times a day. The mobile device of choice was largely a PDA (83%).

3.5 Results

We report trends across users with respect to the use of and satisfaction with help. We then describe the three groups of help seekers that we identified and their help-seeking behavior and satisfaction for PC applications, web sites, and mobile devices. We conclude this section with a discussion of study implications.

3.5.1 Use of and Satisfaction with Help

The majority of users used help systems. Twenty percent sought help for PC applications at least once a day. Another 39 percent sought help for PC applications at least once a month. Users also sought help when using the Web: 21 percent sought help at least once a day, while an additional 27 percent sought help at least once a month. Overall, users sought help for some tasks at least once a month; Figure 1a shows that users sought help primarily for office (i.e., document and spreadsheet creation), information-seeking (i.e., Internet searches), and system administration tasks (i.e., installing hardware or software).

When users do encounter a problem, 68 percent turned to the Internet to find a solution (Figure 1b). Users used Web-based content more so than using product manuals (printed or electronic) or the help system that is embedded within an application, site, or device. We refer to the help system that is accessible within an application, web site, or mobile device as its embedded help. Despite not favoring the use of embedded help, 57 percent of PC users frequently or always knew how to access embedded help. On the contrary, less than half of users frequently or always knew how to access embedded help within web sites. Only 36 percent of mobile computing device users knew how to access embedded help. What forms of embedded help do users use most frequently? Quick reference guides and the index of help contents were the most popular (Figure 1c).

We asked users a series of questions to help to identify areas of potential problems (Table 1). Of particular note, 47 percent of users found that it was sometimes difficult to quickly and easily find the information that they were looking for within a help system. Once they found that information, another 47 percent were only able to understand this information sometimes. As much as two-thirds of users had experienced some trouble when using help systems. Table 1 suggests why users turned to the Web for help, they frequently found information quickly and easily and always found information to be understandable.

Satisfaction with help was mixed. Only 9 percent of users were upset with the fonts and graphics that are used within help systems for PC applications; the majority was satisfied with the appearance of help. On the other end, comprehensiveness of information, amount of information displayed, and intuitiveness of the terminology all had users who were dissatisfied or very dissatisfied (29%).

Table 1. Users' experiences with help systems within the three domains (median frequencies). Only users who used mobile computing devices are included in the mobile device column.

Aspect	PC Applications	Web Sites	Mobile Devices
Find sought info	Frequently	Sometimes	Sometimes
Find info quickly and easily	Sometimes	Frequently	Sometimes
Find info to be accurate	Frequently	Sometimes	Sometimes
Find info to be understandable	Sometimes	Always	Sometimes

We did not find major differences in help satisfaction or use based on our special populations; however, clustering analysis did reveal three groups of users who had similar characteristics.

3.5.2 Three Types of Help Seekers

Our K-means clustering analysis of questionnaire responses revealed three distinct groups for 106 of the 107 users. One user was a super user (i.e., used all types of computers and applications) and was satisfied with all aspects of help; the user did not fit within any of the groups, so we did not include him in our analysis. We characterized the three groups of users based on the degree to which IT is integrated into their everyday lives: low, moderate, and high. Users from our special populations were spread among the three groups.

- The *low-integration group* consisted of 32 users who represent a mixture of newer Internet and computer users—usually students—and people who have jobs or education levels which do not require them to be familiar with many aspects of using computers and the Internet.
- The *moderate-integration group* consists of 40 users who are mostly business people and others who have some college education and use computers primarily for work.
- The *high-integration group* consisted of 34 users who have jobs or life circumstances which put them in constant contact with computers (e.g., computer programmers, digital media editors, and people who are self employed).

Table 2 shows some of the distinguishing characteristics of users within the three groups; all differences were significant according to a one-way analysis of variance. We also found a marked difference in their satisfaction with current help systems. We describe help seeking for users within these groups below.

3.5.3 Help Seeking by Integration Levels

Half of the users in the low-integration group sought help less than once a month or not at all. An additional 25 percent sought help at least once a month to once a week. Users were likely to seek help for the following five tasks: office (50%), system administration (44%), finding information (41%), graphics (35%), and gaming (32%). The four preferred approaches for getting help were: using Web-based help (69%), using printed documentation that is packaged with the product (63%), having conversations with people who are not technical support specialists (63%), and using embedded help (56%).

Table 2. User group characteristics (median values).

Aspect	Low-integration	Moderate-integration	High-integration
Age	25	29	43
Degree level	Bachelors	Associates	Bachelors
PC use	8-10 years	8-10 years	11 years or more
Internet use	5-7 years	5-7 years	8-10 years
Mobile device use	once a week	once a week	multiple times a day
Laptop PC use	no	no	yes

Similar to low-integration users, users in the moderate-integration group were infrequent help users: 50 percent sought help less than once per month and 25 percent sought help either once per month or once per week. When they sought help, they were likely to seek help for seven tasks: office (46%), finding information (37%), system administration (34%), multimedia (29%), financial (25%), graphics (25%), and electronic communication (25%). Of the nine possible strategies for getting help, they preferred four approaches: using Web-based help (54%), using printed documentation that is packaged with the product (41%), having conversations with people who are not technical support specialists (41%), and using embedded help (41%).

In contrast to the other two groups, over two thirds of the users in the high-integration group sought help at least once per month; the remaining third sought help at least once per day. They sought help for the following five tasks: finding information (76%), office (74%), system administration (71%), graphics (62%), and online shopping (62%). Of the nine possible strategies for getting help, they preferred four approaches: using electronic documentation that is provided with a product (88%), using Web-based help (85%), using embedded help (79%), and having conversations with people who are not technical support specialists (76%).

3.5.4 Help Seeking for PC Applications

Users in the high-integration group knew how to access embedded help; only 3 percent rarely or never knew how. In the low- and moderate-integration groups, 32 and 21 percent of users rarely or never knew how to access embedded help. This finding is expected, given that the two groups were infrequent help users and did not use embedded help as their primary resource. It is not clear whether they are infrequent help users, in particular for embedded help, because they cannot find the help functionality. We plan to examine this issue during observation sessions.

We found that low-integration users rated embedded help systems the lowest (i.e., sometimes) with respect to finding information sought, finding information quickly and easily, and finding information to be accurate or understandable (Table 1). Moderate-integration users rated help systems the highest (i.e., frequently) on all four dimensions. We show in subsequent sections that their ratings were consistent across computing domains; this trend is surprising given that they were infrequent help users. We will examine this issue during our planned observation sessions. High-

integration users concurred with moderate-integration users on finding sought information and finding information to be accurate; they concurred with low-integration users on finding information quickly and easily and finding information to be understandable.

The three groups had distinct levels of satisfaction with help systems within PC applications. Users used a 5-point Likert scale, where one is very dissatisfied and five is very satisfied, to evaluate seven aspects of help systems: information comprehensiveness, amount of information, arrangement of information, appearance of information, ease of navigation, usability/accessibility, and intuitiveness of terminology. The moderate-integration group was the most satisfied, being satisfied on all dimensions (median of 4.0). Here again, their satisfaction was consistent across computing domains. The low-integration group was the least satisfied (median of 3.0); they were dissatisfied (median of 2.0) with information comprehensiveness, ease of navigation, and intuitiveness of terminology. The high-integration group was fairly neutral (median of 3.0), possibly due to the broad range of tasks they carry out, their use of laptops and mobile devices, or the frequency with which they use help. They indicated satisfaction (median of 4.0) with the appearance of information.

3.5.5 Help Seeking for Web Sites

Irrespective of the user groups, users were more likely to not know how to access the help system within web sites as compared to PC applications: low-integration (50%), moderate-integration (28%), and high-integration (18%). Low-integration users rated help systems the lowest (i.e., never) with respect to finding information sought and finding information to be understandable (Table 1); they rarely found information quickly and easily or found information to be accurate. Moderate-integration users rated help systems the highest (i.e., frequently) on all four dimensions. High-integration users rated help systems in the middle (i.e., sometimes) on all dimensions, except for finding information to be understandable (median of 4.0).

Users' satisfaction on the seven dimensions was somewhat similar to their satisfaction with PC help systems. Low-integration users were neutral on all dimensions (median of 3.0); they were more satisfied with information comprehensiveness, ease of navigation, and intuitiveness of terminology for web help systems than for PC help systems. The moderate-integration group was the most satisfied, being satisfied on all dimensions (median of 4.0). The high-integration group was fairly neutral (median of 3.0), but they indicated satisfaction (median of 4.0) with the information comprehensiveness, amount of information, and arrangement of information.

3.5.6 Help Seeking for Mobile Computing Devices

Mobile device users were represented at different levels within the three groups: low-integration (25% of 32 users), moderate-integration (25% of 40 users), and high-integration (62% of 34 users). Similar to the Web domain, mobile device users were likely to not know how to access the help system: low-integration (50%), moderate-integration (30%), and high-integration (43%). It is important to note that the high-integration group represents the largest number of mobile device users. Low-integration users rated help systems the lowest (i.e., sometimes) with respect to finding information sought and finding information to be understandable (Table 1); they rarely found information quickly and easily or found information to be accurate. Moderate-integration users rated help systems the highest (i.e., frequently) on all four dimensions. High-integration users rated help systems in the middle (i.e., sometimes) on all dimensions.

Users' satisfaction on the seven dimensions was somewhat similar to their satisfaction with PC help systems. Despite the lower ratings on finding sought information and finding information to be understandable, low-integration users were neutral on all dimensions (median of 3.0), except for two—appearance of information and ease of navigation (median of 2.0). The moderate-integration group was the most satisfied, being satisfied on all dimensions (median of 4.0). The high-integration group was neutral (median of 3.0), but they were more satisfied (median of 3.5) with appearance of information and ease of navigation. The latter findings are surprising given the small screen sizes of mobile devices. Perhaps their adoption and frequent use of mobile devices (Table 2) plays a role; we will explore this issue in our observation sessions.

3.6 Discussion

There are two important and somewhat surprising trends within our study data; both trends are relevant to the World Wide Web. Using the Web to find help for problems encountered when using PC applications, web sites, or mobile devices is the most popular strategy for problem resolution overall. In addition, it is the most popular method used to solve problems among less integrated users (low- and moderate-integration groups). At the same time, web sites often have help systems themselves, but the systems are considered helpful only sometimes.

We think that users turned to the Web most often because it could potentially handle their natural language queries. Term ambiguity was consistently the lowest rated aspect of help systems; it seems plausible that searching with search engines like Google and browsing FAQs that are written by people who do not use jargon are reasons why users favor the Web over embedded help. In fact, users were more likely to seek help by talking to other people than by talking to technical support people. These findings motivated our idea of a Web-based user assistance portal, which we describe in the next section.

At the same time, many users—especially users who are less dependent on computers for their income or degree (low-integration group)—found that the Web is a difficult place in which to find help. Users in all three groups reported that they need help when they attempt to find information online. It bears restating that all the users who completed our questionnaire did so online and are daily Internet and computer users. Casual users are likely to be even more frustrated with the help that is available. We think that developers need to pay more attention to the design of web sites, as well as Web-based help systems in general, so that help features are prominent and effective.

4. TOWARD A HELP PORTAL

Our study showed that users are increasingly turning to the Web, rather than embedded or structured help, when they need assistance with IT problems. Simply put, the help content that users find most helpful does not always exist in a structured help system. The user assistance portal project is an attempt to use machine learning technology to merge this useful content into the structured help system and to make it more accessible to users.

4.1.1 The Case for a User Assistance Portal

Responses to the help use questionnaire indicate that when users seek help for IT problems, they often search the Web for solutions. Empirical evidence suggests that the traditional structured help systems (e.g., Windows Help or electronic manuals) suffer from common problems: Help systems often use jargon that hinders the information retrieval process for non-expert users, content in traditional help systems evolves slowly (perhaps only updated with new versions of software), and structured help content is homogeneous in style and construction (perhaps not supporting different learning styles and abilities). By contrast, online sources are potentially updated continually, are highly heterogeneous (possibly supporting a greater variety of learning styles and abilities), and may be written in language that is more familiar to non-expert users (as many authors of online help content may be non-expert users themselves). These reasons (and possibly others) explain why users reported using Web-based content instead of traditional help systems [35].

The downside of unstructured help sources is that their inherent heterogeneity can make the information retrieval process more complex and time consuming. A user is on her own to identify sources that are potentially relevant to her need and then to query these sources and find a solution from among the results. It is a process that may be fraught with frustration, if the user does not query the proper sources or if the user struggles with information overload [36, 37]. In our study, users in all three groups reported that they sought help while trying to find information online.

Many commercial software applications now have replaced embedded help systems entirely with online content (e.g., Adobe [38]). Others allow users to post comments and updates to online help content (e.g., MySQL [39]). More recently, applications are including links to discussion forums (e.g., Microsoft Office [40]). In all of these cases, users miss out on a large body of help content that exists outside of the vendor's domain.

We think that there is an opportunity to benefit all users by making it easier and more systematic to access relevant unstructured help content, even outside of vendors' domains. The goal of the user assistance portal effort is to unify help content from traditional structured sources (i.e., embedded help) with content from unstructured sources in a manner that enriches the original body of help content and benefits all users.

4.2 Proposed System

We borrow the term "annotation" to indicate a link from a structured help source to an unstructured help source which contains useful and relevant content [41]. We wish to take the content from structured help systems and annotate it with links to useful, unstructured help content. By doing so, users' information retrieval processes can be streamlined and simplified. As annotations become part of the structured help system itself, a user can query and locate this information just as they would locate structured help content. Annotations may also implicitly bridge the jargon barrier by automatically associating structured, jargon-heavy help content with unstructured help content that is expressed with language that is familiar to users.

We envision a system where structured help content coexists with relevant and useful unstructured help content in the form of annotations (Figure 2). Annotations may also indicate what type of source it comes from (knowledge base, newsgroup, etc.), when it was last updated, how many users found it useful, and so on; such information may help a user to discern which annotations will be most useful. As we continually and automatically annotate structured help content, help can become a living, dynamic system.

4.3 Proposed Use of Machine Learning

As depicted in Figure 2, our approach to building this system is to apply machine learning to classify unstructured help content into the hierarchy defined by a structured help system. Machine learning techniques have proven effective for text classification in many domains [42, 43]. Such techniques most often use supervised learning, whereby a classifier is trained on pre-classified documents to accurately classify documents yet unseen. These techniques lend themselves particularly well to our problem. Structured help systems provide a pre-existing pre-labeled body of content that we can use to train a classifier.

Recent research on text classification techniques has shown that it is possible to hierarchically classify documents with greater accuracy than non-hierarchical methods [44, 45]. Other research has found successful methods for classification with limited training data [46]. But little research has been done at the intersection of these techniques [47, 48]. Our research will examine developing successful text classification techniques for when the amount of labeled training and test data is limited, the hierarchy is complex, and classes closely resemble one another (e.g., in an email application, the help content for sending a message and receiving a message is very similar). We are also dealing with a corpus of text that is far more heterogeneous than the popularly used Reuters news collection [49].

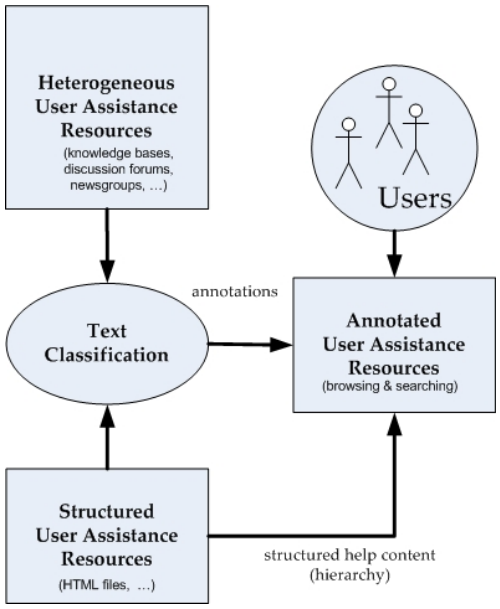


Figure 2. Architecture of the proposed help portal system.

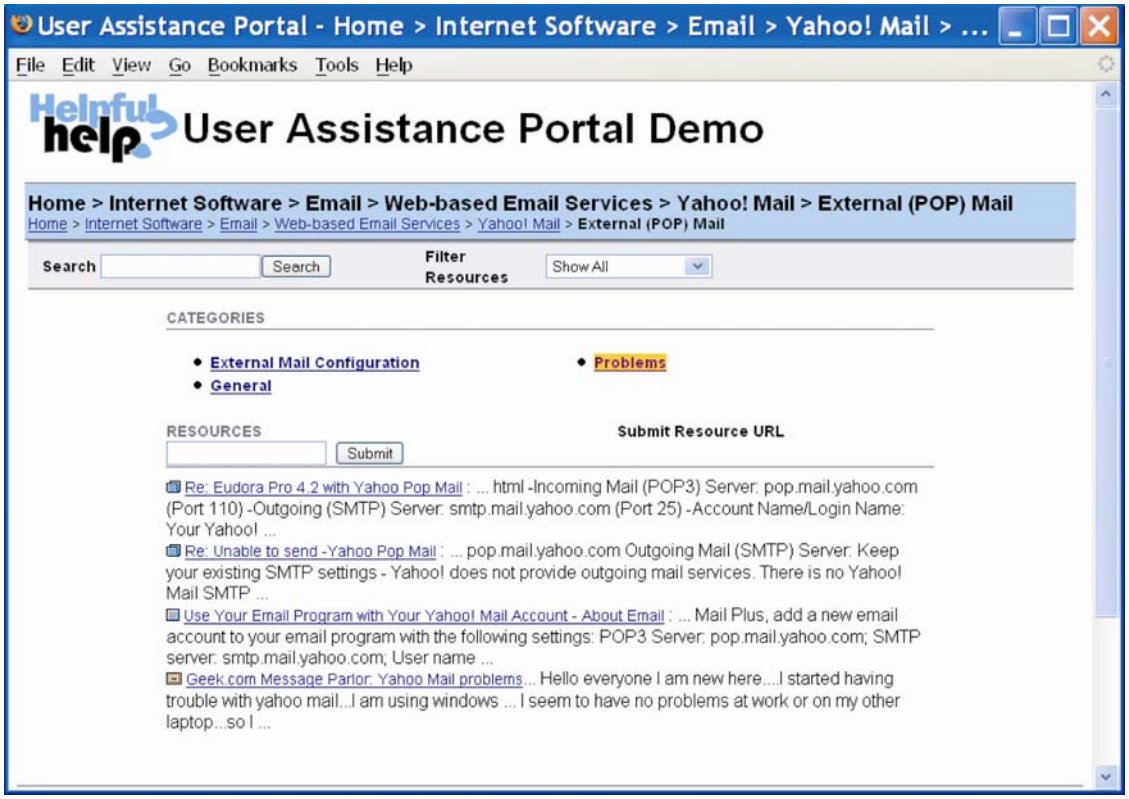


Figure 3. Screen snapshot from our user assistance portal prototype.

We are developing a text corpus to examine the use of existing or the development of novel text classification algorithms; our text corpora should be a useful dataset for other researchers. Our corpora includes structured and unstructured help content for three email applications: Microsoft Outlook [50], Yahoo! Mail [51], and Eudora Internet Suite for Palm OS [52]. The corpus consists of help content taken from Web-based and Windows help sources. Based on our survey of the most commonly used and useful unstructured help sources, we are currently gathering and cleaning content from newsgroups, vendors' knowledge bases, and discussion forums. We developed a rule-based tool to extract the main body of text from these sources by eliminating markup tags and irrelevant text. We hand-labeled categories for testing.

Once we complete the corpus, we will conduct text classification experiments. By comparing results across structured help systems and computing domains, we intend to determine how the volume and hierarchical nature of structured help content affects the success of our text classification techniques. Another aim is to use the experiments to possibly provide guidance to help content authors and architects who may wish to support automated retrieval of their unstructured help content.

4.4 Proposed Interface

We view the annotation system as part of a Web-based user assistance portal. To explore our portal concept, we constructed a Web-based prototype to demonstrate its interface and to describe its envisioned functioning. We solicited feedback from help users during follow-up focus group sessions. Preliminary feedback confirms that users prefer a help system which integrates unstructured sources to one which does not. We will continue to develop the portal prototype and evaluate it during future studies with users, developers, and technical support specialists.

Figure 3 depicts a screen from our proposed user assistance portal. Key features of the portal's interface include the ability to browse a structured help system in a hierarchical manner, the ability to search both the structured help and annotations, and the ability to filter the annotations displayed by type (newsgroup, discussion forum, or knowledge base). From any point within the structured help hierarchy or search results, users can display relevant annotations within the area that is just below the structured content which they augment.

4.5 Implications for the Semantic Web

The primary difficulty in developing the user assistance portal system is finding a classification technique that can accurately classify unstructured documents from heterogeneous Web sources into a hierarchy defined by a structured help system. For most applications, including the email applications within our study, the leaf nodes in a hierarchical help system are not enough to build an accurate classifier. For this reason, we will only be able to provide the most accurate annotations at higher levels of the hierarchy. The vast body of help content that is lying outside of structured sources overwhelms the user and necessitates the use of automated technology to locate solutions to problems. But what if the unstructured content from knowledge bases, newsgroups, and discussion forums carried semantic information? Specific semantic information which indicates the type of source, problem area, and solutions would greatly improve the ability of automated systems to find and accurately classify help content. We advocate the adoption of a Semantic Web standard for help

content so that it will be easier for automated systems to find and aggregate it.

5. CURRENT AND FUTURE RESEARCH

We administered the first round of online questionnaires to user assistance developers (i.e., any professional who plays a role in the development of the content, interface, or features that relate to a help system) and technical support specialists. The two questionnaires examine how technical support specialists provide user assistance and how user assistance developers create help systems. We will analyze this data and contrast it to the user data that we discussed in this paper.

We conducted focus group sessions with help users to explore help use patterns in more detail. Our preliminary analysis of users' discussions corroborates questionnaire findings, in particular our proposed explanations for why users turn to the Web for information to resolve IT problems. Users also evaluated the help portal concept and demo during focus group sessions. Overall, users thought that the concept would help to streamline the help-seeking process. They also provided various suggestions for improvement, such as making search more prominent in the interface and enabling users to rate and comment on help sources.

Our future research on the user assistance portal includes identifying the optimal text classification techniques for classifying unstructured help content into the hierarchy defined by a structured help system. Specifically, we plan to compare the accuracy of multiple text classification algorithms (including Bayesian and SVM methods [30]) and to develop strategies for improving classification accuracy under conditions of limited training data. We also plan to investigate unsupervised learning techniques for text classification.

6. CONCLUSION

We developed a questionnaire on the use of and satisfaction with help systems within three computing domains: PC applications, web sites, and mobile computing devices. We administered the questionnaire to 107 IT users who represent diverse populations. During the questionnaire administration, we uncovered accessibility issues that pervade online questionnaire systems (commercial and open source); we discussed specific issues and the custom solution that we developed to mitigate these issues.

We found that two-thirds of users used the Web to resolve IT problems, more so than they used embedded help. We identified three groups of users, who were distinguished by the degree to which technology is integrated into their everyday lives. They also differed with respect to their use and satisfaction with help. Overall, users were most satisfied with help systems within PC applications and least satisfied with help systems within mobile devices. Given users' preference for locating help content on the Web, we proposed a user assistance portal to support their needs. The portal is an ideal application of the Semantic Web, but its successful development requires research and industry collaboration to standardize the specification of help content.

7. ACKNOWLEDGMENTS

This research was funded by a grant from the National Science Foundation (IIS-0414385) and supplementary funds from the University of Washington's Information School. We thank the participants who completed our online questionnaire. We thank Dan Comden and Randy Hayhurst and students at the Special Technology Access Resource Center for their feedback on our accessible questionnaire system. We thank Jonathan Rochkind for his assistance with a preliminary analysis of study data and Phil Edwards for early contributions to the questionnaire design and recruitment. We also thank the anonymous reviewers.

8. REFERENCES

[1] Randall, N. and I. Pedersen. Who exactly is trying to help us? The ethos of help systems in popular computer applications. In Proceedings of ACM SIGDOC Conference. 1998: ACM Press.

[2] Foster, G. Online help systems: Learning while working. In International Conference on Computers in Education. 2002: IEEE Comput. Soc.

[3] Farkas, D.K., Our little help machines and their invisibilities. ACM Journal of Computer Documentation, 2001. 25(4): p. 141-144.

[4] Johnson-Eilola, J., Little machines: Understanding users understanding interfaces. ACM Journal of Computer Documentation, 2001. 25(4): p. 119-127.

[5] Covi, L.M. and M.S. Ackerman. Such easy-to-use systems!: How organizations shape the design and use of online help systems. In Proceedings of Conference on Organizational Computing Systems. 1995: ACM Press.

[6] Dominick, J., et al., Portal help: Helping people help themselves through animated demos. 2003, University of North Carolina at Chapel Hill.

[7] Delisle, S. and B. Moulin, User interfaces and help systems: From helplessness to intelligent assistance. Artificial Intelligence Review, 2002. 18(2): p. 117-157.

[8] Houghton, R.C., Jr., Online help systems: A conspectus. Communications of the ACM, 1984. 27(2): p. 126-133.

[9] Purchase, H.C. and J. Worrill, An empirical study of on-line help design: Features and principles. International Journal of Human-Computer Studies, 2002. 56(5): p. 539-567.

[10] Krull, R., et al. Usability trends in an online help system: User testing on three releases of help for a visual programming language. In Professional Communication Conference. 2001: IEEE International.

[11] Schweibenz, W., Does help really help? [terminology-related problems with online help systems]. NFD Information - Wissenschaft und Praxis, 2000. 51(4): p. 199-207.

[12] Haramundanis, K., Commentary on "little machines: Understanding users understanding interfaces". ACM Journal of Computer Documentation, 2001. 25(4): p. 128-131.

[13] Kearsley, G., et al. Online help systems: Design and implementation issues (panel). In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1988: ACM Press.

[14] Krull, R. Why wise users may not read computer documentation. In Proceedings of the Society for Technical Communication. 1993.

[15] JoAnn T. Hackos. Let's stop writing documentation and start working for the users. In Proceedings of the Society for Technical Communication. 1999.

[16] Ramey, J., How people use computer documentation: Implications for book design, In Effective documentation: What we have learned from research, S. Doheny-Farina, Editor. 1988, The MIT Press: Cambridge, MA. p. 143-158.

[17] Silveira, M., M.C. Dias, and V. Quental. Helping the composition of help contents. In Advances in Natural Language Processing, Third International Conference. 2002: Springer-Verlag.
[18] Patrick, A. and A. McGurgan. One proven methodology for designing robust online help systems. In Proceedings of the 11th Annual International Conference on Systems Documentation. 1993: ACM Press.
[19] Berners-Lee, T., J. Hendler, and O. Lassila, The semantic web - computers navigating tomorrow's web will understand more of what's going on--making it more likely that you'll get what you really want. Scientific American, 2001. 284(5): p. 10.
[20] Eric Horvitz, et al. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In Fourteenth Conference on Uncertainty in Artificial Intelligence. 1998. Madison, WI: Morgan Kaufmann.
[21] Grayling, T., Fear and loathing of the help menu: A usability test of online help. Technical Communication, 1998. 45(2): p. 168-179.
[22] Parush, A.P.a.D.K., Online help: Too much of a good thing? STC Usability SIG Newsletter, 2001. 7(3).
[23] Christina M. Downs and A.F. Jackson. Interactive help: Adapting content for multiple users. In STC Annual Conference. 2001.
[24] Dutke, S. and T. Reimer, Evaluation of two types of online help for application software. J Comp Assist Learn, 2000. 16(4): p. 307-315.
[25] Hsu, J. Help systems and markup coding: An investigation of the impact of help systems on user effectiveness and productivity using a markup language. In 33rd Annual Meeting of the Decision Sciences Institute. 2002: Decision Sci. Inst.
[26] Smart, K.I., M.E. Whiting, and K. Bell DeTienne, Assessing the need for printed and online documentation: A study of customer preference and use. The Journal of business communication, 2001. 38(3): p. 30.
[27] Sellen, A. and A. Nicol, Building user-centered on-line help, In The art of human-computer interface design, B. Laurel and S.J. Mountford, Editors. 1990, Addison-Wesley Pub. Co.: Reading, Mass. p. 143-153.
[28] SurveyMonkey.com, Surveymonkey.Com - powerful tool for creating web surveys. 2004. http://www.surveymonkey.com/
[29] Freedom Scientific, Jaws for windows overview. 2003. http://www.freedomscientific.com/fs_products/software_jaws.asp
[30] Witten, I.H. and E. Frank, Data mining: Practical machine learning tools and techniques with Java implementations. 1999: Morgan Kaufmann.
[31] PHPSurveyor, Phpsurveyor. 2004. http://phpsurveyor.sourceforge.net/
[32] World Wide Web Consortium, Web content accessibility guidelines 1.0. 1999. http://www.w3.org/TR/WAI-WEBCONTENT/
[33] craigslist, Seattle/tacoma online community. 2004. http://seattle.craigslist.org/
[34] Department of Health and Human Services, Annual update of the HHS poverty guidelines, in Federal Register. 2004. p. 7335-7338.
[35] Enbysk, M., Don't make the call: Go online for tech support. 2003, Microsoft: Redmond, WA. http://www.microsoft.com/smallbusiness/issues/technology/networking/dont_make_the_call_go_online_for_tech_support.mspx
[36] Gates, L., Getting control of data. Application Development Trends, 2002. 9(7): p. p 47-49.
[37] Montebello, M. Information overload-an IR problem? In Proceedings. String Processing and Information Retrieval: A South American Symposium. 1998 : IEEE Comput. Soc.
[38] Adobe Systems Inc., Adobe systems incorporated. 2004. http://www.adobe.com/
[39] AB, M., MySQL: The world's most popular open source database. 2004. http://www.mysql.com/
[40] Microsoft Corporation, Microsoft office online home page. 2004. http://office.microsoft.com/en-us/default.aspx
[41] Marshall, C.C., Toward an ecology of hypertext annotation. 1998, Xerox Palo Alto Research Center: Palo Alto, CA.
[42] Joachims, T. A statistical learning model of text classification for support vector machines. In SIGIR. 2001: ACM.
[43] Sebastiani, F., Machine learning in automated text categorization. ACM Computing Surveys, 2002. 34(1): p. 1–47.
[44] Koller, D. and M. Sahami, Hierarchically classifying documents using very few words. 1997, Computer Science Department, Stanford University: Stanford, CA.
[45] Dumais, S. and H. Chen, Hierarchical classification of web content. 2000, Microsoft Research: Redmond, WA.
[46] Nigam, K., A. McCallum, and S. Thrun, Using EM to classify text from labeled and unlabeled documents. 1999, School of Computer Science, Carnegie Mellon University.
[47] Loukachevitch, N.V., B.V. Dobrov, and M.S. Ageev, Text categorization tasks for large hierarchical systems of categories. 2002, Research Computing Center of Moscow State University: Moscow, Russia.
[48] Boyapati, V. Improving hierarchical text classification using unlabeled data. In SIGIR. 2002.
[49] Lewis, D.D., Reuters-21578 text categorization test collection. 1997. http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html
[50] Microsoft Corporation, Microsoft office online: Outlook 2003 home page. 2004. http://office.microsoft.com/en-us/FX010857931033.aspx
[51] Yahoo! Inc., Yahoo! Mail - the best web-based email! 2004. http://mail.yahoo.com/
[52] QUALCOMM Inc., Eudora Internet suite 2.1 for the palm computing platform. 2004. http://www.eudora.com/products/unsupported/internetsuite/index.html