RightNow eService Center: Internet customer service using a self-learning knowledge base

Stephen D. Durbin, Doug Warner, J. Neal Richter, and Zuzana Gedeon

RightNow Technologies, Inc., P. O. Box 9300, Bozeman, MT 59718
{sdurbin, doug, nealr, zgedeon}@rightnow.com

Abstract

Delivering effective customer service via the Internet requires attention to many aspects of knowledge management if it is to be convenient and satisfying for customers, while at the same time efficient and economical for the company or other organization. In RightNow eService Center, such management is enabled by automatically gathering meta-knowledge about the Answer documents held in the core knowledge base. A variety of AI techniques are used to facilitate the construction, maintenance, and navigation of the knowledge base. These include collaborative filtering, swarm intelligence, fuzzy logic, natural language processing, text clustering, and classification rule learning. Customers using eService Center report dramatic decreases in support costs and increases in customer satisfaction due to the ease of use provided by the "self-learning" features of the knowledge base.

Introduction

Many companies small and large, as well as various types of non-corporate organizations, now find it imperative to maintain a significant presence on the World Wide Web. One of the major organizational functions that is still in the early stages of being delivered via the Internet is customer service, i.e. remedying complaints or providing answers to a particular audience. This task involves many aspects of knowledge management, at least if it is to be convenient and satisfying for customers, while at the same time efficient and inexpensive for the company or organization. On a basic level, it is essential (but not sufficient) to handle the administrative overhead of tracking incoming questions and complaints, together with outgoing responses, over different channels such as e-mail, web forms, and live chat. Beyond this, to support customer service representatives (CSRs), and to assist customers seeking help at peak load times or after hours, it is necessary to provide both a knowledge base containing needed information and a convenient, intuitive means of accessing that knowledge base. Even were it not for the expense of maintaining a large staff of CSRs available at all times, it is found that many users prefer to find answers to their questions directly, rather than take the time to compose a sufficiently detailed e-mail or wait in a telephone queue, possibly playing tag with a CSR for days before resolving their concerns. Furthermore, CSRs may experience boredom and burnout from constantly handling similar questions, and in any case are not using their skills most efficiently.

The most common response to this situation is to write and make available on a web page or pages a set of answers to frequently asked questions (FAQs). This provides a basic solution to the problems mentioned above, but, except in the simplest and most static cases, it requires continued expert maintenance to keep the FAQ list current and organized. In addition, especially if the number of FAQs is relatively large, it becomes difficult for users to navigate the FAQ pages to find the answers they seek.

Recently, a number of more conversational interfaces to knowledge bases have appeared, which may be personified as human or character "chatbots," or represented more soberly as simple input and response text fields. A user is expected to enter natural language questions and will receive replies, the quality of which depends on the level of natural language understanding the system has of both queries and items in the knowledge base. Although continuing progress is being made in the question-answering field [Voorhees & Harman 2001], the commercially available chatbots are based mainly on pattern recognition and pre-written scripts, which require a sizable knowledge engineering effort to create and maintain. We believe that some sort of meta-knowledge (as is represented by the patterns and scripts) is indeed an essential element in facilitating access to knowledge. However, it is also one of our goals to minimize the level of human effort necessary to construct and maintain the knowledge base.

Our approach centers around a dynamic database of FAQ documents, which we call Answers. Meta-knowledge relating to the usefulness of and relationships among Answers is acquired automatically as the knowledge base is used. This is used to spare the experts from most organizational upkeep, and also to make it easier for users to find Answers. By means of the architectural design, with its close coupling of end-user questions and CSR answers, the creators of the knowledge base are necessarily kept up to date on the information needs of end-users, closing a feedback loop that optimizes operation of the system.

In this paper, we describe how this approach is embodied in RightNow eService Center (eSC). After briefly introducing the overall system, we describe in greater detail those aspects of the application related to the knowledge base, as this is where most of the artificial intelligence (AI) techniques come into play. We also present the results obtained by customers using eService Center.

The RightNow eService Center Application

RightNow eService Center is an integrated application that combines e-mail management, web self-service, collaborative live chat, and knowledge management. Most customers choose to deploy it in a hosted environment, but it is also available for individual installation on multiple platforms. It consists of over 500,000 lines of code, primarily in C, but also in C++, Java, and PHP, as well as HTML. The first prototype was constructed 4 years ago; the most recent significant upgrade, to version 5.0, involved about 11 months of effort by approximately 16 full-time developers and 7 quality assurance testers.

The core of the application, from an AI perspective, is the publicly visible Answer knowledge base and the tools by which it is created, maintained, and accessed. This is discussed more fully in the following section. In addition, there is a roughly parallel set of private customer service Incidents which are fully tracked from initial creation, which can be via e-mail, web form, or live chat, through resolution and archiving. For many of our customers, the Answer knowledge base is quite dynamic and may comprise 100's or 1,000's of documents; numbers of non-public Incidents are typically much larger.

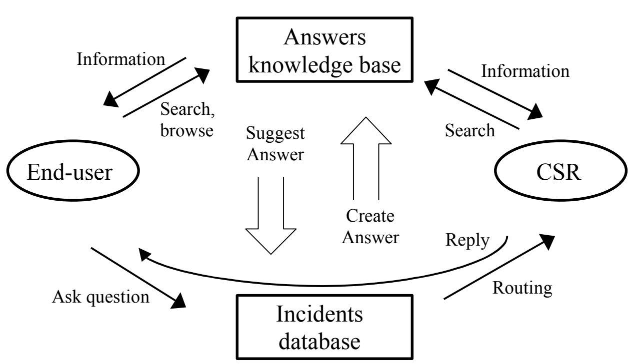

Figure 1. Principal knowledge-related transactions in RightNow Web eService Center. End-users search the Answer knowledge base for information; if they cannot find what they need, they submit a question, which is stored and tracked in an Incidents database, and replied to by a CSR. CSRs also use the knowledge base, and add to it by creating new Answers, typically suggested by frequently asked questions. Answers to questions can be suggested from the knowledge base, either to assist CSRs in forming replies or as auto-replies to end-users.

Other important features of eSC, not discussed in this paper, include extensive administrative functions, customization options, and a wide variety of reports to aid in analysis of transaction statistics, CSR performance, and website usage. One AI-related feature which we allude to briefly here is an emotional index that is determined for incoming messages, as well as in real time for agents involved in live chat sessions with end-users. The emotional index places a message on a scale from negative (upset, angry) through neutral to positive (happy), and is derived using a lexicon of rated words and a set of grammatical rules applied to the part-of-speech tagged text, produced as described in a later section. This index may be used in rules for routing incoming messages, for example, sending angry messages to a veteran CSR while perhaps providing an automated response to positive ones meeting some other criteria.

Constructing an organic knowledge base

In traditional practice, knowledge bases have been constructed by domain experts, who do their best to record, in some form of document, what they know and believe to be necessary for a given task or situation. This paradigm may work reasonably well in capturing knowledge for narrow, static subject areas, but in the case where information needs are constantly changing, the burden of frequently adding new knowledge items can become significant. Although it may be easy to predict that introduction of a new product will lead to inquiries related to that product, it is not so easy to foretell what external events, such as a new law or regulation, or new products offered by competing companies, will cause a shift of end-user information needs. In the absence of human maintenance, conventional Answer lists are brittle in the sense that they break as information becomes out-of-date or irrelevant. Our aim has been to construct a more robust framework that would use AI methods to do as much as possible, and thus require minimal human resources.

The eSC knowledge base is termed organic, because of the natural way it is seeded and evolves over time. A key element of our system is that both growth and organization are responsive to end-users' shifting demands. This is a result of the way in which eSC integrates question and answer channels, and works in the following way (refer to Figure 1). The knowledge base is first seeded with a relatively small set of Answers to the most predictable or frequently asked questions. Many end-users coming to the support website will find their answers among these, but if not, they are encouraged to submit their questions via e-mail or the web-based form provided on the support home page. As CSRs respond to these, they naturally tend to become aware of trends and commonalities among incidents. At any time, a CSR reply, or an edited and extended version of one, can be proposed as a potential Answer. Depending on organizational practices, the item could be reviewed or edited by collaborators or managers before being made a publicly available Answer. The general availability of the answer should then result in a reduction of incoming queries on that topic. Even if such queries continue, there is now an item in the knowledge base available to end-users and CSRs as a Suggested Answer. Answers are suggested by treating the end-user's message as a search query, then filtering the returned set of Answers by requiring them to be in the same cluster as the query. They can be provided automatically to end-users or CSRs.

In this "just in time" paradigm, it is the end-users and their needs which drive knowledge creation, while the CSRs' or other experts' time and effort are conserved. This means that users are more likely to be satisfied, and the CSRs will have more time to focus on the usually smaller fraction of non-repetitive questions.

Navigating a self-learning knowledge base

It is widely understood that knowledge comprises not only facts or data, but also relationships among them, as

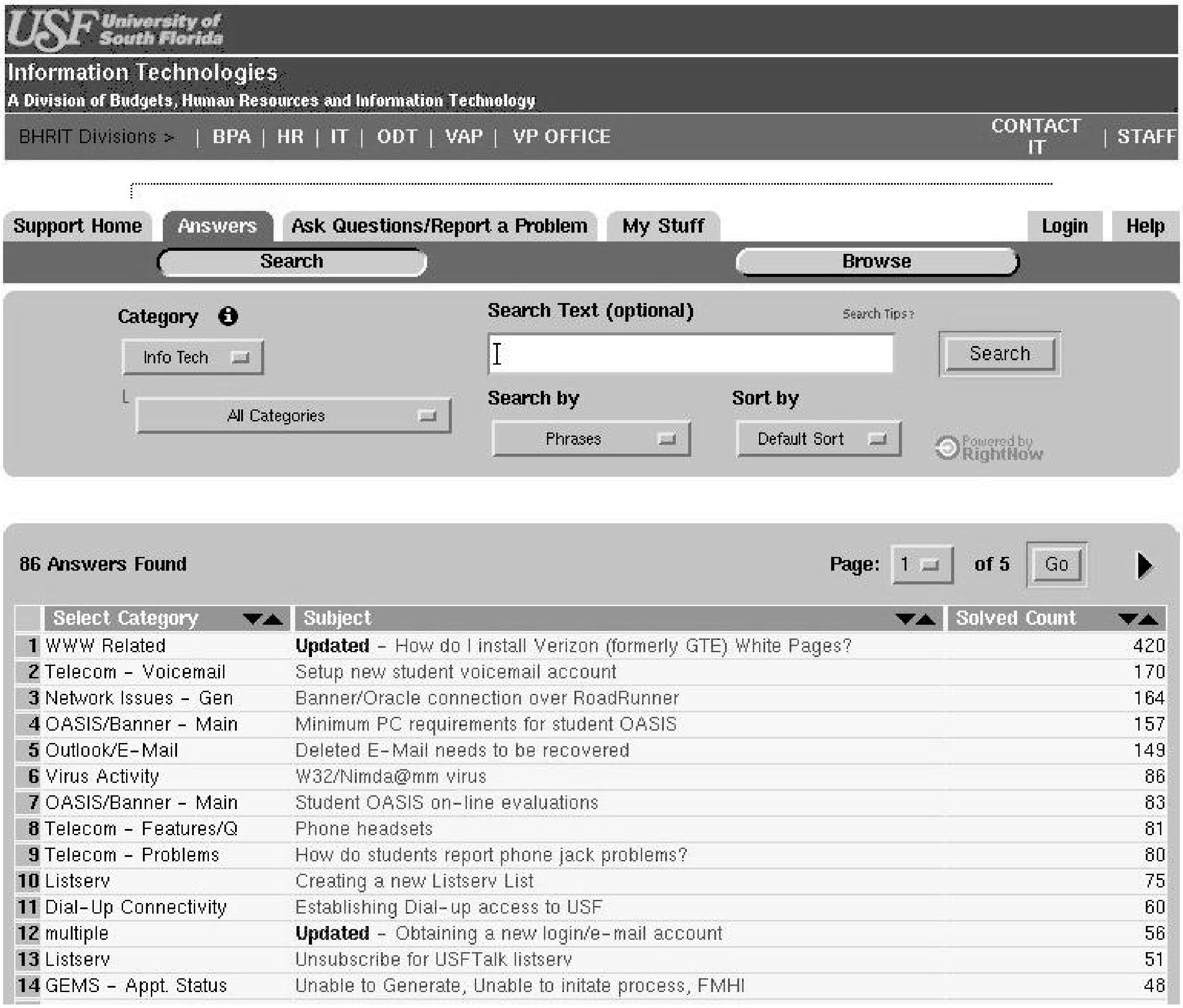

Figure 2. Portion of the web browser display from the eSC support page of the University of South Florida Information Technology division. The page is configured to list by default the historically most useful Answers (highest solved count). As a result, there is a high probability that a relevant Answer can be viewed with a single click. Users may search in various modes by entering search text, or they may view a browsable interface to the knowledge base by clicking the Browse button (see Figure 3). The Ask a Question tab provides a form via which questions can be submitted to support personnel (e.g. CSRs), and the My Stuff tab leads to a personal page with information on the status of any questions submitted. Some end-users may have access to more privileged information via the Login tab.

well as perspective on their importance, relevance, etc. A knowledge base organized to reflect such meta-knowledge forms a much better match to human user habits and expectations, and is consequently easier to use. We call the eSC knowledge base "self-learning" because it acquires this meta-knowledge through a number of AI-related techniques, rather than through human-constructed ontologies, templates, or other form of knowledge engineering. The techniques we use include natural language tools for feature selection, adaptive clustering and classification of text documents, and collaborative filtering and swarm intelligence methods to extract implicit user feedback. We will discuss these as they might come into play during an interaction with the knowledgebase.

An illustration of the first end-user view of a typical knowledge base is shown in Figure 2. By default, this page lists a configurable number of knowledge base Answers, sorted in order of their Solved Count (rightmost column of the display). The latter quantity is a measure of how helpful the answer is likely to be, based on the analysis of previous user activity, as will be described shortly.

If the title of an Answer looks promising to an end-user, a click on it brings up the full text (along with graphics or any other additional information that can be provided on an HTML page). If the information there does not completely answer the user's question, he may return to the original list, or he may elect to follow one of a ranked set of Related Answer links attached to the Answer page. The relatedness ranking is derived from two sources: a simple document similarity measure based on word co-occurrence (with stopword removal and stemming) and accumulated implicit recommendations of previous users

To capture user perceptions of usefulness and relatedness of Answers, we use both explicit and implicit feedback in a manner inspired by collaborative filtering [Levy & Weld 2000] and swarm intelligence [Dorigo, Di Caro, & Gambardella 1999] algorithms. Associated with each Answer is a usefulness counter (solved count) which is increased each time the Answer is viewed, and can also be increased (or decreased) by an explicit rating which the

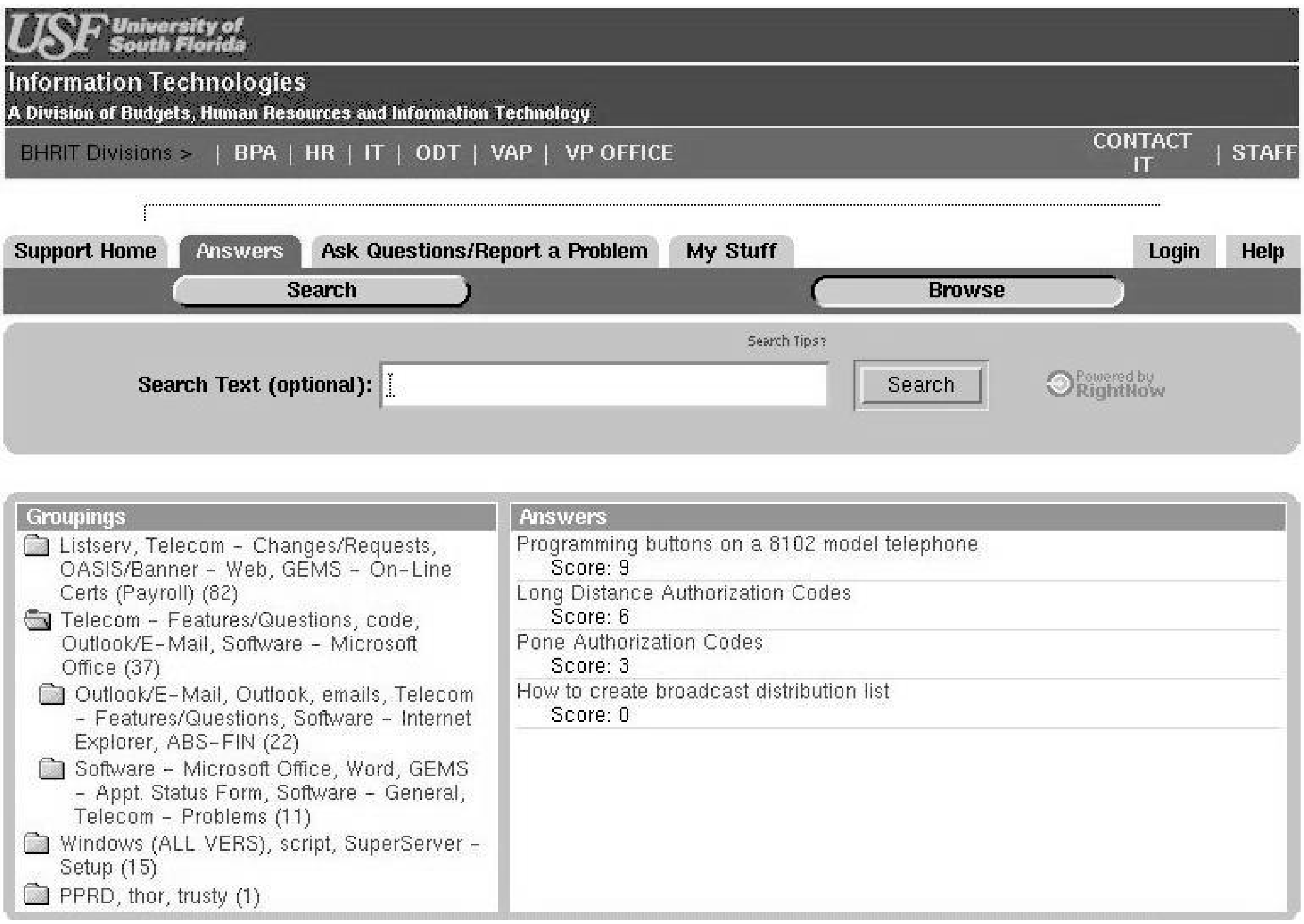

Figure 3. Web browser display from the Browse view of the eSC support page of the University of South Florida Information Technology division. This page displays a hierarchical set of folders and subfolders, where a given folder (like a typical computer file system) may contain both subfolders and Answer documents. The selected folder is the second one at the highest level of the hierarchy (leftmost in the display). Searching can be carried out within a selected browse folder.

user submits by clicking one of a set of rating buttons displayed with the Answer. In addition, a sparse link matrix structure is maintained, the corresponding element of which is incremented each time an end-user navigates from one Answer to another, presumably related one. Because a new knowledge base has no user-derived links, these are initially supplied according to statistical text similarity alone. In a way analogous to pheromone evaporation in ant navigation, both usefulness and link values are periodically reduced in strength when not reinforced. This "aging" keeps the knowledge base responsive by emphasizing recent usage patterns.

Of course, this links matrix contains noise in the sense that not every transition is necessarily made by users only on the basis of perceived relatedness. Nonetheless, when averaged over many users who each tend to be searching for information related to a specific need, we have found that the strong links indicate useful relationships. The potential tendency for highly ranked Answers to be overly reinforced due to their position in the list is mitigated by several factors. A user is unlikely to select an Answer if it does not appear related to her information need (as in any information or web page design, titles are important). If a selection turns out to be mistaken, its usefulness can be downgraded directly via the explicit rating mechanism, and indirectly relative to later Answers that satisfy the user's needs via the implicit mechanism. Also, the aging process decreases each Answer's usefulness (solved count) by a constant multiplicative factor, which reduces higher solved counts by greater amounts. For a fuller discussion of these collaborative and swarm intelligence methods, see [Warner, Richter, Durbin, & Banerjee 2001].

Users with specific information demands, especially if they are less common, may locate information most quickly by searching the knowledge base. Queries entered in the search box allow for a variety of search modes, including natural language input and similar phrase searching (which carries out spelling correction and synonym expansion). A search may be restricted to a given product and/or category, and returned Answers can be ordered by match weight or historical usefulness. The frequency with which terms are searched for constitutes one report that is useful to system managers. If some commonly entered search terms happen not to appear in the Answer documents, these terms can be added either to the synonym list, or to Answer-specific lists of keywords.

End-users may or may not come to a support web site seeking specific information, but in either case they may find it convenient to browse the knowledge base from a higher-level point of view, gaining a broad perspective on the available information. As shown in Figure 3, our system offers a browse mode of access where categories of documents are displayed as folders, which are labeled with the key terms most descriptive of their contents. Clicking on a folder opens it to display documents and sub-folders corresponding to more specific categories. Merely glancing at the labels on the folders at the highest level gives an outline summary of the contents of the knowledge base. Because the user can navigate by selecting subfolders and individual documents without needing to type search terms, this browse mode is especially helpful when the user is unfamiliar with the terminology used in the Answers and hence would have difficulty forming a productive search query. Thus we enlist the user's tacit knowledge, his or her ability to recognize more easily than articulate.

Supporting a browse function without a human-defined ontology requires clustering and categorization of the text items in the knowledge base. For this we employ a heavily modified version of the fast, hierarchical clustering algorithm BIRCH [Zhang, Ramakrishnan, & Livny 1996], which is run repeatedly while varying the threshold parameter. The best result, according to a clustering figure of merit, is used to learn RIPPER-style classification rules [Cohen 1995]. The final topic hierarchy is created by classifying knowledge base items according to the rules, allowing each item to potentially be classified in multiple places. Multiple classification recognizes the inherent multiplicity and subjectivity of similarity relationships. It makes searching via the browse interface much more convenient, as the end-user can locate an item along various paths without backtracking, and does not have to guess what rigid classification might control the listing.

The features on which the clustering is based are obtained from the document texts by shallow natural language processing involving part-of-speech tagging with a transformation-based tagger [Brill 1994]. Noun phrases are identified and receive the highest weight as features, but selected other words are also used. In addition, customer-supplied keywords and product or category names provide highly weighted features. These features are increased in weight if they are frequently searched for by users.

Extraction of the classification rules allows new knowledge base items, as they are created, to simply be inserted into the hierarchy in the same way as previous Answers. However, after a predetermined amount of change in the knowledge base, due to modification, addition, or removal of documents, a re-clustering is performed so that the browse hierarchy reflects the current state of the knowledge base, rather than a fixed hierarchy.

User Experience with eService Center

The system we describe has been used, through several versions, by a wide variety of commercial, educational, and governmental organizations. Drawing from their accumulated experience, we have gathered both aggregate statistics and numerous case studies demonstrating the dramatic reduction of time and effort for knowledge base creation and maintenance, and the increase in satisfaction of knowledge base users. This holds across the spectrum of organizations and applications, including those outside the area of conventional customer service.

The ease of installation is such that it has been accomplished in as little as a day, if initial seed Answers are available and major customization is not needed. As a demonstration that is part of our sales process, companies can set up pilot installations in 2-5 days. Once set up, the knowledge base can grow rapidly. For example, the United States Social Security Administration started with 284 items in their initial knowledge base, and over 200 new items based on user-submitted questions were added within two weeks. Now, after two years, the number has stabilized at about 600.

The ability of a web self-service system to handle dynamic fluctuations in usage can be very important. As one example, the January 2001 announcement of a rate hike by the U.S. Postal Service led to a short term increase in visitors to the support site of Pitney-Bowes, which provides mailing services, of nearly 1000% over that for the previous rate hike. Attempting to handle such volume via telephone or e-mail would have resulted in huge backlogs.

A quantitative measure of end-user success in finding information, as well as of cost reductions to a company, is the self-service index, defined as the percentage of end-users who are able to find their own answers online, rather than sending a message to a CSR. Table 1 is excerpted from a Doculabs study [Watson, Donnelly, & Shehab 2001] in which it was found that, depending on the type of organization, the self-service index using eSC ranged from 75% to almost 99%, averaging 87%. According to anecdotal statements from customers, these benefits are largely attributable to the key elements of the self-learning knowledge base as described above [Case studies]. We believe that credit is also due to the knowledge acquisition processes facilitated by eSC: the experts' time and energy is used much more effectively [Durbin, Warner, Richter, &Gedeon 2002].

In addition to standard customer service, eService Center is flexible enough to be used in other knowledge management settings. A number of organizations use it internally to provide information to their members, from general interest news to specific areas like personnel forms and procedures. Within our company, RightNow Technologies, it is also used as a shared information resource between quality assurance and development teams. In this use, quality assurance testers submit bug reports (analogous to customer questions), while developers respond to them. A single bug history may contain a number of transactions involving several people on each team. This system not only facilitates the communication between the two workgroups, but provides a valuable organizational memory for future reference.

Industry	Visits	Escalations	Self-Service Index
General Equipment	342,728	4,144	98.79%
Manufacturing	22,784	489	97.85%
Education	8,400	317	96.23%
Entertainment/Media	113,047	4,622	95.91%
Financial Services	40,574	1,972	95.14%
Contract Manufacturers	77,838	4,203	94.60%
Utility/Energy	19,035	1,122	94.11%
ISP/Hosting	147,671	8,771	94.06%
IT Solution Providers	53,804	3,277	93.91%
Computer Software	449,402	27,412	93.90%
Dot Coms	267,346	20,309	92.40%
Medical Products/Resources	17,892	1,451	91.89%
Professional Services	24,862	2,142	91.38%
Insurance	40,921	3,537	91.36%
Automotive	3,801	373	90.19%
Retail/Catalog	44,145	6,150	86.07%
Consumer Products	1,044,199	162,219	84.46%
Computer Hardware	101,209	15,759	84.43%
Government	108,955	17,347	84.08%
Travel/Hospitality	27,099	4,610	82.99%
Association/Nonprofit	14,620	2,772	81.04%
Telecommunications	809,320	202,158	75.02%
Overall Total	**3,779,652**	**495,156**	**86.90%**

Table 1. Self-service index for various types of organizations using RightNow eService Center. The self-service index is the fraction of end-users that find needed information in the Answer knowledge base, rather than initiating contact with a support person (escalating) via e-mail or online chat.

Discussion

Despite our emphasis on the successes of eSC, there is certainly room to do better. Some improvements, such as making clustering more adaptive to differing knowledge bases, are fairly straightforward. More difficult is the problem of automatically producing good summary labels for the clusters; our current heuristics work well in some cases, and less well in others. The area of multi-document summarization is one of active current research (see e.g. Mani & Maybury 1999) and one of our priorities is to improve this aspect of eSC in future releases.

More qualitative enhancements can be obtained from applying AI techniques to a greater number of functions. Incident routing, text categorization, and natural language processing are all areas we are working on.

As knowledge bases inevitably become larger and more complex, the need for a system like eSC increases. The knowledge bases with which eSC has been used so far have not been extremely large, very seldom reaching more than a few thousands of documents (though many more items are usually in the Incidents database). Algorithmic changes may become necessary to scale some of the behavior to much larger databases, especially for processing that is done while an end-user is waiting.

Another trend affecting many Internet-based applications is that toward greater personalization of user interfaces. Care must be exercised to ensure such customization facilitates and enhances rather than constrains and obscures. In an information-finding task, one doesn't want to miss something because of an agent's faulty assumption. The extent to which significant personalization is feasible and desirable for frequent or for one-time users is still being investigated.

Conclusions

We have described the web-based customer service application RightNow eService Center, which relies on a number of AI techniques to facilitate construction, maintenance, and navigation of a knowledge base of answers to frequently asked questions. These techniques include collaborative filtering, swarm intelligence, fuzzy logic, shallow natural language processing, text clustering, and classification rule learning. Many of these individual techniques have been employed for similar purposes in other commercial applications, but we know of no other system that combines all of them. Customers using eSC report dramatic decreases in support costs and increases in customer satisfaction due to the ease of use provided by the "self-learning" features of the knowledge base.

The principles and methods embodied in eSC are also applicable in other settings. For example, the relationship between a government agency and concerned citizens is closely analogous to that between a business and its customers. In fact, organizations and associated constituencies with information needs are ubiquitous in our modern society. In the conception and development of eSC, we have emphasized the generalizable features of dynamic focus on current information needs, ease of updating and maintenance, and facilitated access to the knowledge base.

References

Brill, E. 1994. Some advances in transformation-based part of speech tagging. In Proceedings of the Twelfth National Conference on Artificial Intelligence. AAAI Press.

Case studies are available at the following web page: http://www.rightnow.com/resource/casestudies.php.

Cohen, W. H. 1995. Fast effective rule induction. In Machine Learning: Proceedings of the Twelfth International Conference. Morgan Kaufmann.

Dorigo, M., Di Caro, G. & Gambardella, L. M. 1999. Ant algorithms for discrete optimization. Artificial Life, 5(2), 137-172.

Durbin, S. D., Warner, D., Richter, J. N., & Gedeon, Z. 2002. Organic knowledge management for web-based customer service. In Nemati, H. & Barko, C, (Eds.), Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance, in press.

Levy, A. Y. & Weld, D. S. 2000. Intelligent Internet systems. Artificial Intelligence, 118, 1-14.

Voorhees, E. M., & Harman, D. K., (Eds.). 2001. NIST Special Publication 500-249: The Ninth Text REtrieval Conference. U. S. Department of Commerce, National Institute of Standards and Technology.

Mani, I., & Maybury, M. T. (Eds.). 1999. Advances in Automatic Text Summarization. Cambridge: MIT Press.

Warner, D., Richter, J. N., Durbin, S. D., & Banerjee, B. 2001. Mining user session data to facilitate user interaction with a customer service knowledge base in RightNow Web. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 467-472. Association for Computing Machinery.

Watson, J., Donnelly, G., & Shehab, J. 2001. The Self-Service Index Report: Why Web-Based Self-Service is the ROI Sweet-Spot of CRM. http://www.doculabs.com

Zhang, T., Ramakrishnan, R. & Livny, M. 1996. BIRCH: an efficient data clustering method for very large databases. In Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, pp. 103-114. Association for Computing Machinery