

AC 2009-1812: ASK NJIT LIBRARY: A NATURAL-LANGUAGE KNOWLEDGE BASE SELF-SERVICE SOLUTION

Haymwantee Singh, New Jersey Institute of Technology

Haymwantee P. Singh, Technical Reference Librarian, Robert W. Van Houten Library, New Jersey Institute of Technology, singhh@njit.edu

Richard Sweeney, New Jersey Institute of Technology

Richard T. Sweeney, University Librarian, Robert W. Van Houten Library, New Jersey Institute of Technology, richard.sweeney@njit.edu

ASK NJIT LIBRARY: A Natural Language Knowledge Base Self-Service Solution

Abstract

As library websites proliferated over the years, librarians have developed ways of enhancing users' experiences as they navigate a vast ocean of valuable library resources. Librarians have also been creative in handling high volume inquiries via online help desk service that include FAQ, chat service and email reference that have some 24/7 elements. These services are necessary but still not sufficient for today's impatient library users looking for instant gratification. This paper addresses a 24/7 website self-service experience that makes it easy for website visitors to type a question in natural language and get one right answer. It will be demonstrated how ready reference questions can now be answered through a natural language knowledgebase powered by IntelliResponse that can be updated every time a user asks a new but important question. There are many benefits of this empowering program such as reduction of email volume, phone calls and chat sessions and reduced navigation linked with website Q&A or FAQ solutions. The quality of end user experience is measured objectively via system generated reports. Potential exists for institutions to collaborate and build such knowledgebases more efficiently and for development of self-learning tools.

Introduction

Ask the library, ask the registrar, ask the IT help desk, ask the calculus professor, ask the XXXXXX, and receive an immediate, consistent and accurate online answer to your question 24x7, 365 days per year. Imagine that each time a librarian is asked a question, s/he may not have to answer the same or similar question in the future. Imagine that students can ask questions online and are able to receive their professor's answers even while that professor is off doing his or her research, on sabbatical, or just on down time. A natural language knowledge management system could be the solution. An integral component of knowledge management systems, a knowledge base, is used to optimize information collection, organization, and retrieval for an organization, or for the general public^[1]. Functions of a natural language knowledge base make it possible to answer specific questions that are likely to be asked repeatedly by other users but perhaps in a slightly different manner. Natural language knowledge bases currently used by some academic departments such as libraries, registrars, and information technology may also be exploited by individual faculty and academic degree programs. They too must answer the same questions that are asked repeatedly by different students in different sections or even over different semesters.

S.R. Ranganathan (1892-1972), renowned international librarian, proposed five laws of library science. His fourth law states "*save the time of the reader*"^[2]. On Ranganathan's fourth law, Noruzi explains that in order to save the time of the user, web sites need to be designed with an efficiency factor that will enable users to find what they are looking for quickly and accurately, as well as to explore the vast collection of information available that could potentially be useful.^[3] According to Steckel, this law has both a front-end component (make sure people quickly find what they are looking for) and a back-end component (make sure our data is structured in a way that information can be retrieved quickly)^[4]. In short, the best way to save

the user's time is to make sure that s/he can always obtain immediate answers (access) anytime and anywhere, and that the answer delivered is always accurate and consistent. Reusable information is the hallmark of and the future of the web as well as the reason for knowledge bases.

Knowledge Bases

Higher education is a competitive business. It is imperative that the users' learning styles and experiences, based on their exposure to the internet and other learning technologies, be considered in customer relations management applications, in order to better serve customer needs and enhance their satisfaction. As web savvy users browse college websites, they expect to find answers that are accurate and delivered efficiently. College websites have begun to introduce website self service programs that users have grown accustomed to in the business world. Users no longer wish to browse the menu options of a website and go through layers of web pages and hope to get to the one accurate answer for which they were searching. It is now possible to ask questions in natural language about tuition, admission requirements, SAT and GRE scores, and so on and receive an accurate answer without ever speaking to a live person.

NJIT (New Jersey Institute of Technology) has instituted such a natural language knowledge management system powered by IntelliResponse (www.intelliresponse.com), sometimes called an online instant answer agent. This self-service program assists users to find concise and consistent answers 24/7 to frequently asked questions, anytime anywhere. IntelliResponse was first implemented by the NJIT IT Helpdesk, Admissions/Registrar and Continuing Professional Education. When NJIT librarians learned about this natural language knowledge base concept, they embraced it immediately.

Ask NJIT Library

Like many library homepages, NJIT library homepage is menu driven (library.njit.edu). Each menu item links to a large hierarchy of associated pages with useful information. A menu item is expected to convey accurate information about the contents of the underlying submenus or links. In the case of library.njit.edu, a question such as "How do I place a book on hold?" can be answered via a menu option. But it is not obvious which one of the menu choices should be selected in order to obtain an accurate answer to this question. Links to the answer are at two locations. If one selects the "Policies" menu option and then navigates to the "Circulation Policy" via the link "Borrowing of materials", a partial answer can be found. The FAQ site at library.njit.edu also has an answer and this site can be found under the menu item "Research & Instruction". On the other hand, the average library user may not know the library jargon "hold" and may not know what term or phrase to look for. Looking through a FAQ list may prove more fruitful if the question happens to be on the list. As such a library user may be in a situation where valuable time is spent browsing layers of websites and even going back and forth between pages, often a frustrating experience.

The NJIT Library now has a search box (i.e. query box) in the center of the main library homepage. A searcher may type a question in natural language and submit it at "Ask the Library" and, receive an accurate response/answer. Users do not have to navigate many layers of

web pages associated with website Q&A or browse a lengthy list of frequently asked questions. Using the IntelliResponse technology, the library developed a knowledge base of questions, associated responses and criteria (natural language parsing rules) and this resides in a knowledge management system. Criteria allow the knowledge base administrator to create searchable terms including truncated synonyms that increase the likelihood of the searcher obtaining the correct answer.

System reports allow the librarian or administrator to recognize and add a new but important question that is not already in the knowledge base. The next time a user asks a similar question, there will be a consistent response. And, the user does not necessarily have to ask the identical question in a complete sentence format but can use natural language terms such as “place a book on hold” or “recall a book” or “request a book that is out” or simply “holds”. Another example about the library hours, “What time do you close?”, “What are the library hours?”, or “library schedule”, “library calendar”, will provide the same common answer and will point the user to the same web page for more detail. The criteria for this example may include the synonyms: close, clos*, hours, hors, schedule, and calendar.

IntelliResponse has a feature that allows each user to give immediate feedback to the knowledge base administration about the answer given to a query. Each user can rate the level of satisfaction of an answer on a scale of poor to excellent. The program also provides leads to other possible questions that may be related to the user’s query. In addition, a user has the option to escalate feedback to the library via email. Even before one attempts to type in the query box, there is the option to view the top ten (Frequently Asked Questions) questions.

Libraries are continuously exploring ways to improve users’ library experiences including empowering them to help themselves. Libraries have had open stacks for easy access to books and journals, all catalogued and systematically arranged. More recently, over the past decade, libraries have permitted self-service electronic access to databases, online journals and other valuable resources. Reference services were enhanced through virtual reference management Q&A tools 24/7/365. QuestionPoint, OCLC’s virtual reference management tool is a well known chat based reference medium mediated by librarians. Chat sessions between users and librarians are recorded. Follow up questions and answers from an online chat session are sometimes done via phone and email conversations. This follow up information is often added to a chat session to make a complete transaction that could be added to a knowledge base that allow for future retrieval. Tam Dalrymple of OCLC describes this knowledge base as the QuestionPoint repository^[5]. Unfortunately, the democratization of the QuestionPoint tool is not enough to empower all library end users to query the knowledge base and retrieve information. It can only be queried by librarians.

In a 2008 article about the reference library service in the digital environment, Michael Buckland asks a very important question, “So how now, might we design the ideal reference library service to empower users in an increasingly digital environment?” He makes reference to students working in their dorm rooms on their assignments between 9 p.m. and 5 a.m. and asks how reference library service might be designed for them without limiting ourselves (or them) to call center solutions^[6]. NJIT’s library website self service solution for ready reference or frequently asked questions that accommodates natural language vocabulary can be considered a partial

response with regards to Buckland’s question as well as testimony to Ranganathan’s words of wisdom. Buckland also states that “A digital environment characterized by local computing power everywhere and the ability to engage with remote resources anywhere, leads naturally to user self-help and a desire for self-sufficiency. The economic and psychological grounds for promoting self-help are compelling and, if our service is going to be used, we need to minimize the time and effort, and the difficulty of use” [6].

Work on NJIT Libraries self-service (“Ask the Library”) program commenced in August 2007. The complete implementation of the program took approximately 200 hours over seven months. This time period included training of one professional librarian for about ten hours over two days, creation of a knowledge base of about seventy questions and focus group testing on February 12, 2008. The knowledge base creation was the most challenging undertaking. Librarians across all departments produced a list of over 100 questions. This list included questions stated in natural language and questions with similarities. Questions were also obtained from our library FAQ list and from other library websites. Many answers or responses to questions had to be integrated with existing websites and many new answers had to be created. The focus group consisted of 15 end users who were undergraduates, graduate students and faculty. The focus group feedback captured by IntelliResponse was analyzed over a two weeks period (average of two hours per day). This invaluable feedback helped to improve existing questions and uncover new content opportunities. The NJIT Libraries launched its natural language knowledge base and self-help program on March 18, 2008. By the end of December 2008, there were over 5,800 questions asked.

System generated reports provide information on the system’s performance, question analyses by frequency of responses, unanswered question listing, type of questions by time frames such as daily, weekly and monthly and much more. The system health or strength is defined as the ratio of total questions answered to total questions asked. Table 1 shows quarterly data on questions asked and the associated system health for each period. It is obvious that the system health continues to improve.

Quarter	Total Questions	System Health
Jan - Mar	202	62%
Apr - Jun	1483	77%
Jul - Sep	2046	74%
Oct - Dec	2031	83%
Total	5762	

There are many questions that come to mind. Is a response or answer to a query checked for accuracy? It is very important that the responses to questions be consistently checked for accuracy as a quality control measure to make sure that users are getting not only an answer but an accurate answer based on the term(s) or phrase used in the query. Are the user ratings on the scale of poor to excellent helpful? Unfortunately, our users have not been rating the answers enough in order to summarize this type of feedback. The library could conduct another focus group in the future to solicit ratings feedback. How much time is spent on the knowledge base

maintenance since the product was launched? In the first 30 days after the launch, 513 questions were asked of the system. The system health was then 65%. In the next 30 days, there were 1,073 questions and the system's health was 72%. The first two months required at least two hours maintenance daily in order to improve question and answer matches for existing responses, add new questions, add and revise criteria and answers. In the process of articulating accurate answers to a specific question, we have uncovered policy statements that needed revision and website pages that needed renovation. One may also have to study, research and record oral policies. Thereafter, at most one hour per day is necessary for maintenance. Of course there are peak periods and slow periods during the semester. Towards the end of 2008, two more professional librarians have been trained for backup and for further knowledge base development.

Benefits of the Technology

There are many benefits to this program. A response page to a question can also help to augment the user's knowledge on a topic since one response unit could possibly answer more than one question. For example, the answer to the question about "placing a book on hold", provides additional information such as "how long does a hold or recall take", steps to place a book on hold, a link to the library's circulation policy, and a link to the online catalog.

This answer associated with the term "book on hold" constitute a self-learning information literacy experience for the user. Below are some of our thoughts on the benefits and impact of this innovation on library users and the library community.

Benefits to Library Users:

- Accessibility anywhere/anytime via the web
- Increased accuracy and consistency to directional, policy and procedural questions
- Reduction of a user's time to email, to make a phone call, to engage in chat, to make personal contact with a staff person or to make a trip to the library or to navigate library website Q&A or FAQ solutions

Benefits to the Information Community:

- Introducing a model for library self-service that permits natural language query
- Introducing a model that can make it possible for libraries to collaborate and build and share questions and answers in areas such as information literacy as well as ready reference rather than try to build Q&A from scratch. For example, it might be possible to collaborate and build a folder of all the possible questions and answers with respect to the question "How do I cite a website". It would be possible to link both a web page as well as a short video explaining how to do this.

Impact on the Library:

- Provision of better quality and more consistent answers to questions
- Creation of new knowledge from captured unanswered questions can be used to create new knowledge
- Reduction of staff time to answer emails, to take phone calls, or to chat
- Reveals implied customer needs without surveys or focus groups

Impact on Perception of the Library:

- Library staff themselves find the knowledge base valuable that even some of them have begun to use it to obtain answers to frequently asked questions.
- Enhanced library service – anecdotal evidence favors the system

Cost may be of interest to the reader. IntelliResponse provides a fully hosted solution. According to IntelliResponse, pricing is based on the model that may be considered a best fit for an institution or agency. Some pricing criteria may also apply such as number of interfaces, number of responses, number of enrolled students, and others. The software price range for academia is \$35,000 to \$55,000, depending on the preferred license structure. In addition, there is an annual maintenance fee. Assuming a call center solution, Table 2 presents two employment scenarios of a student worker and a librarian whose time could be spent in answering 6,000 questions (ready reference/FAQ) in a period of twelve months. One must also take into consideration that 6,000 questions answered in 500 contact hours are not in steady state but the hours must be spread over the year. So it is obvious that the cost could be much more since the 6,000 questions are answered 24/7/365. It is assumed that it may take five minutes to answer one question.

Total questions per year	6000
Number of minutes per question	5
Total hours for 6000 questions	500
Annual cost at student rate of \$8.00 per hour	\$4,000.00
Annual cost at Librarian rate of \$30.00 per hour	\$15,000.00

During the month of February 2009, the average number of questions asked per day was 24. If this system requires at most one hour per day for maintenance by a professional librarian, then one can assume that there is a positive return on investment for this product.

Conclusion

While the NJIT libraries have used IntelliResponse software and currently are the only libraries in the US doing so, there are other natural language knowledge base question and answer options available. RealDialog from Astute Solutions, RightAnswers, RightNow and Talisma may be similar to IntelliResponse and may be implemented at other academic institutions. More importantly, in the case of the library, a natural language knowledge base should not be limited to answering ready reference type of questions but should be expanded to more sophisticated information literacy questions. Also, knowledge bases of this nature should not be limited to the mainstream departments in higher education such as IT helpdesks, registrar and admissions. There are many other departments in colleges and universities that support lengthy lists of web-based FAQs. Every course could have such a natural language knowledge base of every question and answer given to every student in every section over the years. Questions that had been asked previously in a particular course may, initially be mined from the learning management systems such as BlackBoard, WebCT, Sakai, Moodle, etc. Questions can be content based (How do I divide fractions?) or processed based (What percentage of my grade is for the final exam?). The answers given do not have to be just text and may incorporate short videos and other graphics.

The future of natural language knowledge bases could include peer-to-peer questions and answers as well as student-to-teacher questions and answers. If students ask questions and other students answer those questions, the faculty members could appoint teaching assistants to edit and add the responses without having to write answers individually to every question asked. The future will certainly include empowering end-users to answer questions for other end-users with or without review by authorized editors and having those answers included (as many are doing within wikis).

Ranganathan may say today, a successful library knowledge management system saves the time of the user, as well as, saves the time of the librarian.

References

- [1] J. May, "Definition of a Knowledge Base," March 5, 2007.
- [2] S. R. Ranganathan, *The Colon Classification* vol. IV. New Brunswick, NJ: Rutgers University Press, 1965.
- [3] A. Noruzi, "Application of Ranganathan's Laws to the Web," in *Webology Volume 1, Number 2*, December, 2004.
- [4] M. Steckel, "Ranganathan for information architects," in *Boxes and Arrows*. vol. 7, 2002.
- [5] T. Dalrymple, "'Just in Case' Answers: The Twenty-First-Century Vertical File," *Information Technology and Libraries*, vol. 27, pp. 25-28, Dec. 2008.
- [6] M. K. Buckland, "Reference library service in the digital environment," *Library and Information Science Research*, vol. 30, pp. 81-85, 2008.