Mr.Web: An Automated Interactive Webmaster

Andrea Lockerd, Huy Pham, Taly Sharon, Ted Selker
MIT Media Lab
20 Ames St. E15-322
Cambridge, MA 02139
+1 617 253 0219
<alockerd, huy, taly, selker>@media.mit.edu

ABSTRACT

This paper describes a system, Mr.Web, designed to interact with users over email to create and update Web pages. Our goal is that users interact with Mr.Web as if it were a human Webmaster. We collected 325 examples of people writing email requests to a Webmaster, and used this to generate the semantics of Mr.Web's email parser. The results of the survey indicate that the limited context of a Webmaster gives us a reasonable subset of the natural language processing (NLP) problem. This paper explains the system design, user study results, and plans for future work.

Keywords

Human-Computer Interaction, Natural Language Processing

INTRODUCTION

The Mr.Web system is designed for individuals and groups that don't have a Webmaster being paid to manage their Web presence. In these cases, creating a Web page and keeping it updated is a time and resource intensive task that takes away from someone's primary job. Thus, a number of pages out there in cyberspace are out-of-date.

How can we get people to update their Web pages? The premise of the Mr.Web project is that these very same people constantly send email to update people about their various projects. If updating one's Web page were as easy as sending email to a colleague…problem solved! Our project explores this avenue by having Mr.Web, an automated Webmaster, react directly to email requests, making updating and correcting Web pages easier and less time consuming.

Related works include: Majordomo [2], a system that automates the management of mailing lists; and Website management tools such as Strudel [1]. The key contribution of Mr.Web is the use of email as an interface for Web maintenance.

MOTIVATION

A project member, Sally, realizes her name is misspelled on the Web page. Sally sends email to Mr.Web with the correct information. Mr.Web checks its email, adds it to the list of things to do, makes the changes, and recreates the page. Sally receives a confirmation email from Mr.Web and a link to the fixed page. Sally follows the link to see her name is now correct. By using an automated Webmaster, this can all be achieved with a delay of less than a minute.

USER STUDY

In order to discover patterns and regularities in the email requests people send to a Webmaster when they need to initiate a Web page change, we ran a user study. The results of which were then used to design Mr.Web's email parser.

Procedure

By soliciting participation over email, we had 65 subjects (a strong majority were students). Each subject performed the task of composing an email request to the Webmaster five times, resulting in 325 example requests [3].

Each task consisted of writing an email requesting a particular change to a given Web page. Of the three basic types of change requests (add, delete, and update), this survey covered both delete and update. Subjects were presented with "before" and "after" pictures to show what
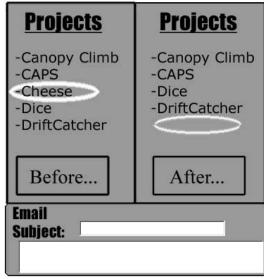

Figure 1: Survey Task: write an email to the Webmaster to initiate the above change.

change they were supposed to initiate. We chose not to use words to describe the problem so as not to have the language of the problem descriptions influence the language the subjects naturally choose to use. An example of a representative task, removing a project listing from the project page, is shown in Figure 1.

The subjects were told that they did not know the Webmaster and were not informed of the project goal: an automated Webmaster.

Results

The resulting examples were analyzed by hand for semantic regularities in three categories: change-type, where-to-change, and what-to-change.

The survey data found there to be consistency in words used to describe the delete and update change-types; about 85% of the set showed a noticeable semantic pattern. For example, we found that users wishing to initiate an update often said 'change', 'correction', 'update', 'replace', 'should be', 'wrong', or 'needs to be'. Where-to-change was the most straightforward category. In about 85% of the survey examples, people gave the page name and/or URL of where they wanted a change to take place. The what-to-change category was the most varied of the three. The data exhibited a semantic pattern in only 70% of the examples.

These results inform us of the language and semantics people use in the context of changing Web pages, and are the basis of Mr.Web's email parser.

DESIGN

The goal of Mr.Web is to allow a user to communicate with an automated Webmaster about Web pages as easily as with a person.

Natural Language Processing

The design centers around two factors, Limited Context and Fail-Soft Interactivity, to minimize the difficulty of the natural language processing (NLP) problem:

1) Limited Context: since Mr.Web is only expected to communicate about changing Web pages, this limits what Mr.Web can expect to find in an email interaction with a user.

2) Fail-Soft Interactivity: the parser built and presented in this paper shows promising initial results, but is still not able to fully understand every change requested. A crucial element of the design is what it should do in the unsure cases. In order for the system to gain the users' trust, it is important that Mr.Web has a fail-soft solution; therefore, when Mr.Web is not able to completely parse an email request, he will send a follow-up email asking for further clarification.

System Architecture

The Mr.Web system has three main elements: Content Management, Email Communication, and System Administration.

Content Management

Content of a website managed by Mr.Web is kept in a database thereby reducing the problem of content management to the task of keeping the database up to date. Scripts are used to generate static pages from this database.

Email Communication

The backbone of the Mr.Web system is the email parser that implements the upkeep of the content database. This parser is based on the semantic tendencies people exhibited in our user study, and by using semantic-transition trees it translates English questions and commands into database query commands [4]. The resulting parser is able to fully understand the change-type, where-to-change, and what-to-change in 65% of the email requests represented in the survey set. The remaining 35% fall into the category mentioned previously that would be further clarified through the Fail-Soft Interactivity mechanism.

General System Administration

Some tasks one would expect of all good Webmasters are also implemented. Mr.Web generates Web page statistics, and notifies users if a page has not been updated recently. Mr.Web also notifies system administrators if the Web server is down and logs errors to assist in repair.

FUTURE WORK

There are two main points of future work. First, we plan to further verify the reliability of the email parser by running a second version of the user study and also by deploying the system to manage a small group's Website. Second, we are in the process of implementing the Fail-Soft Interactivity portion of the system that will allow Mr.Web to "double check" and ask for clarification on questionable requests.

CONCLUSION

This paper demonstrates an area where limited context simplifies the problem of NLP, allowing a computer to act competently in the stead of a person. We ran a survey to study the language and semantics of Webmaster requests. Our survey demonstrated that people naturally use a constrained language when communicating with a Webmaster over email. Mr.Web's email parser was based on the results, and is able to correctly decide what to do with 65% of our sample set of email requests. In the coming months we will be developing the interactivity portion of the system as well as deploying the system to gain a better understanding of its reliability.

REFERENCES

1. Fernandez, et. al., *Overview of Strudel - A Web-Site Management System*, Networking and Information Systems 1(1): 115-140, 1998.

2. Majordomo, http://www.greatcircle.com/majordomo.

3. User study: http://cac.media.mit.edu:8080/contextweb/huy/survey.htm.

4. P. H. Winston, Artificial Intelligence, 3rd Ed., Ch. 29, Adison-Wesley, 1992.