

Webmail: An Automated Web Publishing System

David Bell

NOAO¹, P.O. Box 26732, Tucson, AZ 85726

Abstract. A system for publishing frequently updated information to the World Wide Web will be described. Many documents now hosted by the NOAO Web server require timely posting and frequent updates, but need only minor changes in markup or are in a standard format requiring only conversion to HTML. These include information from outside the organization, such as electronic bulletins, and a number of internal reports, both human and machine generated. Webmail uses procmail and Perl scripts to process incoming email messages in a variety of ways. This processing may include wrapping or conversion to HTML, posting to the Web or internal newsgroups, updating search indices or links on related pages, and sending email notification of the new pages to interested parties. The Webmail system has been in use at NOAO since early 1997 and has steadily grown to include fourteen recipes that together handle about fifty messages per week.

1. Introduction

Publishing a document on the World Wide Web often involves several steps, which may be performed by multiple parties:

- Authoring – The document's content is created.
- Editing – Textual markup is added, and document may be split up.
- Posting – The document is copied to a Web server.
- Referencing – Links to the document are added or updated on other pages.
- Advertising – The document's presence is announced.

In many cases most or all of these steps can be completed in an automated manner. By reducing the need for direct human involvement, such documents can be posted and updated quickly and accurately, even by individuals without any experience with HTML.

2. Description

The Webmail software is installed in a dedicated account on the Web server which has been given write access to any directories or files it must update. On

¹National Optical Astronomy Observatories, operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation

the author's end, posting consists merely of sending an electronic mail message to this account. In some cases, a pre-agreed subject line is used to identify the document. For messages originating from interactive programs, this step is performed with a simple "Submit" button.

Procmail first filters the message by running a MIME-decoder, if needed. This has been particularly useful for handling messages sent from some PC-based mailers which unnecessarily encode plain text attachments, often without the sender's knowledge. Procmail then checks against a series of recipes to identify the message type. These include a list of addresses permitted to post or update the document, and usually information about the expected subject line or contents of the message. Procmail logs and archives all submissions and passes the message contents and type to a Perl script for further processing. Unidentified messages are forwarded for human review.

The Perl script includes recipes for each message type. Standard methods include Web posting, posting to newsgroups, archiving previous versions, and sending email announcements to distribution lists.

Specialized recipes are used for the more sophisticated documents. There is not always a one-to-one correlation between email messages and Web documents – a single message might be split into many pages or used to update just small sections of existing pages. Links may be added or updated on several related pages. For documents that are part of a searchable Web database, index files are also updated.

3. Examples

Webmail was launched in early 1997 for NOAO's internal posting of IAU Circulars, but has since grown to include fourteen recipes that together process about fifty messages per week. Recipes range from simple HTML-wrappers to those performing sophisticated parsing and reformatting. Some examples are described below.

- IAU Circulars – Message are indexed for searching, posted to an internal newsgroup, and saved as plain text. A CGI script on the NOAO Intranet is used to search circulars and add HTML wrappers and links to related circulars. A useful feature of this approach is that the links may point to circulars that arrived at a later time with updated information.
- Telescope Schedules – CTIO and KPNO schedules are sent from a *MS Access* database in a column-formatted table encompassing six months of dates for all telescopes. Individual months are split out of the message, formatted into HTML tables and cross-linked. For each scheduled project, links are provided to a page with detailed information including a full list of investigators and the project's title and scientific abstract. Minor schedule changes may occur several times each semester, so any changed months are announced by email to a distribution list. Example schedules may be found at http://www.noao.edu/kpno/forms/tel_sched/ or http://www.noao.edu/ctio/forms/tel_sched/.

- Conference Room Schedules – These are maintained on a PC running *Schedule+* and are output as HTML and mailed. Webmail fixes some of the markup and splits and recombines weekly schedules into more useful formats for posting on the NOAO Intranet. A single message updates over twenty Web pages.
- ACM Colloquium page – This is one of the simplest recipes Webmail processes, but may be of interest since the author is from outside NOAO. A pre-formatted HTML document is simply posted to the NOAO server at <http://www.noao.edu/noao/acm.html>.

4. Conclusion

We have found the Webmail approach to be quite useful and powerful, yet perhaps so simple as to be easily overlooked. For pre-formatted documents needing little or no markup, new recipes can be set up with just a few lines of procmail and Perl code. The resulting process is more efficient than passing each document through a human Webmaster, and more secure than giving all authors direct access to the Web server. For documents requiring extensive parsing and markup, the benefits of an automated system are even greater.

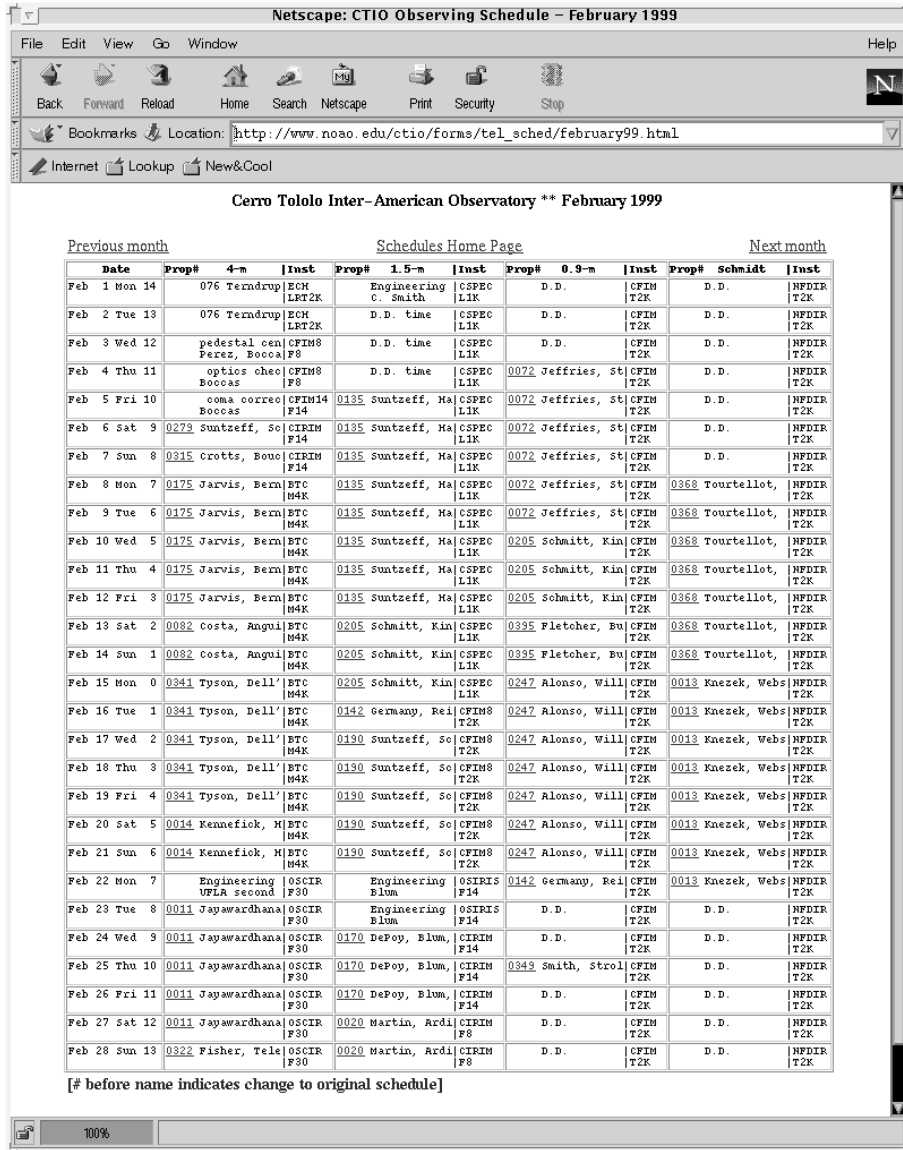


Figure 1. A telescope schedule page produced by Webmail.