

News Control in Messaging

By Jacob Palme¹, e-mail jpalme@dsv.su.se, at the research group for CMC² (Computer Mediated Communication), which is a part of the K2Lab laboratory³ at the DSV⁴ University Department of Stockholm University⁵ and KTH Technical University⁶.

First version: 13 July 2002.

Abstract: News control is a central function in many Internet application. By news control is meant that the computer stores information about what a person has seen and not seen, and this information can be used by a user, for example to scan only new, unseen, information items.

Definitions

The word group is used below to denote a group of information items, such as messages sent via a certain mailing list, articles in a certain Usenet News newsgroup or contributions to a certain forum.

Introduction

News control is a central function in many Internet applications. By news control is meant that the computer stores information about what a person has seen and not seen, and this information can be used by a user, for example to scan only new, unseen, information items.

The ideal news control facility should know exactly what a user has seen and not seen anywhere on the Internet. I know of no

facility which completely fulfills this requirement.

A common restriction in news control is that if the same message is sent to more than one group (=forum, mailing list) and a user is a member of both groups, then the message is marked as new only within a group, and the user will thus be shown the same message as new in each group. (In a few cases, this is what the user wants, but usually not.)

News Control in Web Browsers

Most web browser stores a history list of web pages, which a user has recently seen. Usually, this list is set to contain about a hundred pages or pages views in for example the last 30 days.

When a user views a web page with links, links to pages on this history list have a

¹ <http://www.dsv.su.se/~jpalme>

² <http://www.dsv.su.se/~jpalme/DSV-CMC-Home-page.html>

³ <http://www.dsv.su.se/research/K2LAB/K2LAB.html>

⁴ <http://www.dsv.su.se/>

⁵ <http://www.su.se/>

⁶ <http://www.kth.se/>

different color than other links. This is useful when a user clicks around on pages in a web site, to avoid seeing the same page more than once.

This function has, however, usually several restrictions:

1. It usually does not recognize when the content of a page with the same URL has changed.
2. News information is only visible when a user views a page with links. The user does thus not get a complete overview of what is new, the user is only informed of a new item if the user visits a page with links to it.

News Control in E-mail and Forum Systems

News control in e-mail and forum systems is usually based on some way or marking which message a user has seen. This mark is often only valid within one forum/ mailing list, and a message sent to more than one forum/ mailing list is thus shown at both places.

Global or Local Identifiers

The ideal way of storing news information is to have a data base, with globally unique identifiers of each item, and which stores for each item whether a particular user has seen it or not. Examples of such globally unique identifiers are URLs for web pages and Message-IDs for mail messages and Usenet news articles.

Many systems, however, instead use an identifier which is unique only within a particular group. For example, most Usenet

News implementations uses the number of an article within a particular newsgroup.

The identifier used in Usenet News is local to each news server. The same article in the same newsgroups can thus have different identifiers on different news server. Since, however, most users always use the same news server, this does not normally reduce the quality of the news control provided.

Central or Local Storage

The news control information can either be stored in the personal computer of the user, or on a central server used by several users.

Storing the information in the personal computer saves load on the server, and has the advantage that news control is also available when the user is not connected to the central server.

Storing the information in a central server has the advantage, for some users, that they can use different personal computers and still get compatible news information.

Also note that storing the news control information in the local computer of the user requires a client on the personal computer. Thus, when the user uses a general web browser as client, news control has to be done on the server, except when the limited news control built into web browsers is enough.

(Theoretically, news control information could be stored in cookies, which would allow it to be stored on the local computer without a special client. However, cookies are usually too limited in size to allow for this.)

Usenet News stores all news control information in the personal computer of the user. News servers usually do not even contain any directory of users.

E-mail using POP stores news control information in the client. E-mail using IMAP can store news control information in either client or server.

Compressed Storage

If the storage of news information is not compressed, one bit is required for every item and every user, to store whether this user has seen this item or not, plus the full text of the identifier for each item.

Some systems use identifiers which are sequential numbers, for example of articles in a newsgroup. Since there are often long sequences of articles which a user has seen or not seen, Usenet News uses a compressed format. Instead of storing:

| Newsgroup: alt.foo | |
|--------------------|---------|
| Article No. | Is seen |
| 1 | True |
| 2 | True |
| 3 | True |
| 4 | False |
| 5 | True |
| 6 | False |
| 7 | True |

Usenet News stores the following information (in a file named newsrc):

| |
|---------------------------|
| alt.foo: 1-3, 5, 7 |
|---------------------------|