Enabling Dissemination of Meta Information in the Usenet Framework

Christopher Lueg
Department of Information Systems, University of Technology Sydney,
PO Box 123, Broadway NSW 2007, Australia
Email: lueg@it.uts.edu.au

Abstract

The paper discusses a transparent and flexible way to disseminate meta information within the global conferencing system. Examples of such information are ratings for Usenet articles or information about the behavior of other users. In this paper, the Usenet "overview" mechanism was modified to disseminate meta information to off-the-shelf news clients. The modified overview mechanism is discussed by example of two fully working prototype implementations.

1 Introduction

The global conferencing system Usenet news is by far the largest online system of its kind, serving the information needs of users every day. In May 2001, Usenet users were contributing about a million articles every day, producing net 100 Gbytes per day.

Despite its enormous growth since the early 1980s, Usenet has reached an outstanding level of reliability and fault tolerance. Without this reliability, it is difficult to capture the global conferencing system in a stable and complete manner. Examples are the standard format of Usenet articles [Horton and Adams 1986] and the Network News Transfer Protocol (NNTP) [Kantor and Lapsley 1986]. Account independent standards, news servers and news clients running on a variety of operating systems are able to cooper with each others' technology.

While efficiency and reliability of news servers have improved significantly over the years, moderate progress in the field of Usenet client interfaces has been observed. State-of-the-art Usenet clients have reached a high level of reliability but most client interfaces do not provide information about the behavior of other users. This is a rather primitive visualization of discussion threads.

This lack of progress in news client technology is astonishing because the global conferencing system has been used extensively for tasks such as information filtering projects (e.g. [Morita and Shinoda 1994, Lang 1995, Mock 1996, Terve 1994]) and investigations of Usenet communication traffic (e.g. [Whittaker et al. 1998, Smith 1999, Sack 2000, Smith and Terve 1998]). Research projects investigated Usenet's visible "output'' in terms of articles but they rarely fed results back to Usenet. Research results were used to develop the global conferencing system.

Some interesting work in human-computer interaction has also been done in the context of Usenet newsgroups, such as virtual newsgroups [Fischer and Stevens 1991] and groundbreaking work in collaborative filtering [Resnick et al. 1997]. Filtering agents [Sarwar et al. 1998] and social navigation support [Lueg 2000]. See [Lueg and Fisher 2002] for an overview.

Despite limited direct impact, experiences with collaborative filtering, interactive filtering and social navigation support suggest that such approaches indeed provide benefit to Usenet users. Moreover, Usenet has shown to be a huge information space providing an exciting environment for research in information filtering and social navigation.
indicate that it may be difficult to incorporate novel technologies into the Usenet framework.

From a research point of view, the question is how the incorporation of novel approaches into the Usenet framework relates, more general question is how the global conferencing system could become more accessible to research.

This paper addresses the problem that the Usenet framework does not provide support for the dissemination of meta information to Usenet’s own communication traffic. Recent approaches in human-computer interaction, such as social navigation or collaborative filtering, depend on the dissemination of meta information, such as information about the behavior of other users or ratings in the case of collaborative filtering.

From a technical perspective, the problem is that Usenet is a large infrastructure consisting of many servers and components that need to be changed according to fixed protocols. Such an infrastructure is difficult to change because all components need to be changed otherwise, old components may break down if they are provided input they are not able to handle. As it is hardly possible to change the whole system as a whole, we looked at Usenet’s technical foundations to find a way to disseminate meta information within the Usenet framework.

First the paper discusses some of the technical issues involved in disseminating meta information within Usenet. Usenet’s “overview” mechanism was modified so that it can be used to disseminate meta information. Two prototypes have been developed. The modified overview mechanism to disseminate collaborative filtering ratings and social navigation information are presented.

Next concludes with a discussion of some open issues and future research directions.

2 Specific Characteristics of the Usenet Framework

In its current form, the Usenet framework does not explicitly support the distribution of meta information, such as collaborative filtering ratings and social navigation information, in addition to Usenet’s own information. Accordingly, new ways had to be found to disseminate meta information. We found that two of the “standard” approaches that work well in other domains, such as the Web, can be applied to Usenet’s characteristics of the global conferencing system. The two standard approaches are introducing additional meta channels and adding meta information to documents that are already being disseminated.

Introducing additional distribution channels, such as additional augmented protocols (e.g. [Maltz 1994]), involves changes to access modified information sources must be capable of using the additional capabilities. Experiences made in the context of Usenet filtering of Usenet articles [Resnick et al. 1994] suggest that it is hard to introduce modified versions of existing meta channels. Experiences made in the context of groupware also indicate that it is unlikely that many users are willing to adopt new tools if the benefit of using them is unclear (e.g. [Grudin 1988]).

The second approach, i.e., adding meta information to documents that are already being disseminated, is discussed in more detail. While modifying documents works fine in uni-directional domains like the Web (e.g. [Barrett et al., 1994]), consistency problems in the multi-directional Usenet environment. It is a specific characteristic of the global conferencing system. Articles are being disseminated among news servers on a “flood-fill” basis. Flood fill means that all servers offer all other news servers they are exchanging articles with. In order to avoid inconsistencies, strict rules exist as to which news servers are allowed to add certain information, such as Path:, Message-ID:, and Date: headers, to an article. Three rules prevent adding new meta information would cause inconsistencies as different articles with the same unique Message-ID: identifier.

Any approach to disseminate meta information in the Usenet framework should be in line with current Usenet standards. Meta information should be available even to off-the-shelf clients which means that being able to access the information is always an option.

3 Details of Information Dissemination in the Usenet Framework

The global conferencing system Usenet is based on a client-server model in which one or more news clients require access to a news server (Figure 1). News servers are central repositories where news articles are collected and stored. To distribute articles among news servers, news articles are being disseminated among news servers on a “flood-fill” basis. Flood fill means that all servers exchange articles with each other. In order to avoid inconsistencies, strict rules exist as to which news servers are allowed to add certain information, such as Path:, Message-ID:, and Date: headers, to an article. Three rules prevent adding new meta information would cause inconsistencies as different articles with the same unique Message-ID: identifier.

Any approach to disseminate meta information in the Usenet framework should be in line with current Usenet standards. Meta information should be available even to off-the-shelf clients which means that being able to access the information is always an option.
clients have to submit articles to a news server that then disseminates the articles to peer news servers.

![Usenet's client-server model involving one or more news clients requesting articles from a news server.](image)

Figure 1. Usenet's client-server model involving one or more news clients requesting articles from a news server. The server exchanges articles with other news servers that are not included.

When a user wants to read a particular newsgroup, his or her news client has to connect to a news server and request articles available in that newsgroup. Once received, the list of articles is presented to the user (see Figure 2 for an example). The user can then choose to read a single article which is in turn requested by the client from the server. Finally, the article is displayed to the user.

![News client Knews visualizes the discussions available in a particular newsgroup.](image)

Figure 2. The news client Knews visualizes the discussions available in a particular newsgroup (Knews is courtesy of Dominik Mäurer).

Most current news servers maintain a dedicated “overview” database in which only selected article properties, such as Subject: headers, author identifications (From: headers), and unique article identifiers (Message-ID), are stored. When a client requests the list of articles from a news server, the server uses its overview database to provide the information requested. The benefit of the size of an article’s overview data is much smaller than the size of the full article. This means that although the overview database implies additional processing overhead for news servers, it is still a quick and resource-friendly way to obtain an overview of a newsgroup. Neither the news server nor the news client have to scan the full articles in order to...
required.

Our research indicates that the "overview data" concept is flexible enough to be used for the dissemination of Usenet's regular information. Moreover, meta information transported as part of the overview data should be aligned with current Usenet standards (see below for experimental verification).

4 Details of Overview Data Communication

When a user starts to read a particular newsgroup, roughly the following communication between the user's news client and server takes place (Figure 3):

1. The client requests overview data describing the articles available in the currently selected newsgroup by issuing a `newsgroup` command to the server.
2. The server returns the overview data as requested.

![Diagram of overview data communication](image)

Figure 3. Most clients request "overview data" from a news server in order to provide users with an overview of the newsgroup. Full articles are only requested on the user's selection.

Clients can request a description of the overview data format by issuing a `list overview.fmt` command following listing of a telnet session shows the format description as returned by the news server `news.ifi.unizh.ch` requesting the data is emulated by manually connecting to the server's NNTP port, which is the standard port to news clients.

```
<lueg@padma><514 ist>telnet news.ifi.unizh.ch NNTP
Trying 130.60.48.84...
Connected to sifnos.ifi.unizh.ch.
Escape character is '^]'.
200 news.ifi.unizh.ch InterNetNews NNRP server INN 2.2.1 25-Aug-1999 ready (posting complete)
215 Order of fields in overview database.
Subject:
From:
Date:
Message-ID:
References:
Bytes:
Lines:
Xref:full
quit
205 .
```
Internally, the format is kept in a regular file:

```
<news@sifnos><45> cat overview.fmt
## $Revision: 1.3 $
## overview.fmt - format of news overview database
## Format
##  <header>
##  <header>:full
##  header is a news article header, known by innd. If ":full" appears,
##  then header name will be prepended. Order of lines is important!
Subject:
From:
Date:
Message-ID:
References:
Bytes:
Lines:
Xref:full
<news@sifnos><46>
```

The description of the overview format lists the elements of the overview data as well as their particular sequence. It is important as only the Xref header is sent along with an unique identifier (indicated by the additional full key) format description, clients know about the structure and the semantics of the data that will be returned when requesting newsgroup.

The following listing of a telnet session shows overview data returned by the news server news.ifi.unizh group ifi.general was used to enter the newsgroup ifi.general and the command xover 1246 overview data describing article 1246. As described by the overview data format, overview data returned after a notification in line 8 consists of the internal number, the Subject: header ([Brownbag] ... (Verena Vanessa Hafner ...), the Date: header (Tue, 10 Apr 2001 ...), the Mess:
(Pine ... @ludwig.ifi.unizh.ch), and some more information (C: denotes what the client rate received).

```
C: telnet news.ifi.unizh.ch NNTP
S: Connected to sifnos.ifi.unizh.ch.
S: Escape character is '^]'.
S: 200 news.ifi.unizh.ch InterNetNews NNRP server INN 2.2.1 25-Aug-1999 ready (posti
C: group ifi.general
S: 211 4 1244 1247 ifi.general
C: xover 1246
S: 224 data follows
S: 1246    [Brownbag] Thursday, April 12, 12:30, Minoru Asada      Verena
S: Vanessa Hafner <vhafner@ifi.unizh.ch>    Tue, 10 Apr 2001 13:59:32 +0200
S: <Pine.SOL.4.33.0104101255420.5496-100000@ludwig.ifi.unizh.ch>
S: 2218    39      Xref: news.ifi.unizh.ch ifi.general:1246 ethz.general:333
S: unizh.general:149 de.sci.informatik.ki:5043
S: .
C: quit
S: 205 .
S: Connection closed by foreign host.
```
Based on the overview data received, a news client can generate a list of articles currently available in the news example. Before presenting the list to the user, the client may apply user-defined actions, such as highlighting. In addition, most clients consider the articles already read by the user so that only new articles are listed.

The next step is that the news client waits for the user to select a specific article for further reading. Only upon does the client fetch the full article:

3. The client issues an **article** command along with some other information to the server to
4. The server returns the full article as requested.

Finally, the client presents the full article to the user. The process of selecting and fetching articles continues until change to another newsgroup or to quit the news reading session.

5 Issues in Implementing Extended Overview Data

A commonly used news server was modified to implement extended overview data, i.e. overview data that may include additional meta information. First, the modification of the overview mechanism itself is described. Then, comparing the overview implementation to the standard overview mechanism is discussed. Finally, the paper describes how the overview data has been extended to be able to retrieve meta information from external information sources, such as a collaborative ratings system for Usenet articles.

5.1 Implementation of the Extended Overview Data Mechanism

The regular overview mechanism along with its separation of data and data format description is flexible enough to accommodate additional data fields. The use of specific headers allows news clients to distinguish regular overview data from other application contexts. For example, an advanced news client may look for **Navigation:** headers providing useful for social navigation purposes while **Recommendation:** headers providing collaborative filtering ratings for Usenet articles.

The following example shows how a news server informs clients about additional information available when a `list overview.fmt` format is requested. As in previous examples, communication between client and server is emulated by using `telnet`:

```plaintext
C: telnet news.ifi.unizh.ch NNTP
S: Trying 130.60.48.84...
S: Connected to sifnos.ifi.unizh.ch.
S: Escape character is '^]'.
S: 200 news.ifi.unizh.ch InterNetNews NNRP server INN 2.2.1 25-Aug-1999 ready (posti
C: list overview.fmt
S: 215 Order of fields in overview database.
S: Subject:
S: From:
S: Date:
S: Message-ID:
S: References:
S: Bytes:
S: Lines:
S: Xref:full
S: SELECT-Rating:full
S: .
C: quit
S: 205 .
S: Connection closed by foreign host.
```
The following example of actual overview data shows specific rating information that is stored in the `SELECT-Rating:` header.

```
C: telnet news.ifi.unizh.ch NNTP
S: Connected to sifnos.ifi.unizh.ch.
S: Escape character is '^]'.
S: 200 news.ifi.unizh.ch InterNetNews NNRP server INN 2.2.1 25-Aug-1999 ready (posti
C: group ifi.general
S: 211 5 1233 1247 ifi.general
C: xover 1246
S: 224 data follows
S: 1246    [Brownbag] Thursday, April 12, 12:30, Minoru Asada      Verena
S: Vanessa Hafner <vhafner@ifi.unizh.ch>    Tue, 10 Apr 2001 13:59:32 +0200
S: <Pine.SOL.4.33.0104101255420.5496-100000@ludwig.ifi.unizh.ch>
S: 2218    39      Xref: news.ifi.unizh.ch ifi.general:1246 ethz.general:333
S: unizh.general:149 de.sci.informatik.ki:504   SELECT-Rating: high
S: .
C: quit
S: 205 .
S: Connection closed by foreign host.
```

News clients looking for `SELECT-Rating:` meta information may use the additional information provided. A `Rating:`-aware client may filter all articles that received low ratings while highlighting all articles that received high rating. Other clients would simply ignore the information.

Extended overview data should not require modifications of clients to cope with the additional information as meta format is in line with the current NNTP standard. Additional information should not cause problems with clients that do not support the extended overview database. In general, clients should simply ignore any fields they do not know.

To implement "extended overview data" an InterNetNews (INN) news server, probably the most widely used news server, was responsible for communication with clients, was therefore modified in such a way that the extended overview database was transparently "inserted" when sending clients. Figure 4 shows at what point the header is inserted into the communication between news clients and the news server.
Figure 4. Most clients request "overview data" from a news server in order to provide users with a brief overview available in a newsgroup. We use overview data to disseminate meta information to news clients.

5.2 Compatibility the Extended Overview Data Mechanism

When investigating compatibility of "extended overview" to existing installations, two cases were distinguished. Whether currently available news clients are able to handle extended overview at all (e.g. whether they crash when sent overview data instead of regular overview data). Tests with a variety of news clients, such as Knews and the mozilla Messenger, suggest that sending additional meta information as part of the overview data does not present problems.

Second, we wanted to know to what extent current clients are able to make use of additional information provided by an extended overview headers. Providing additional overview headers is in line with current Usenet standards but we were unsure how many clients would notice the additional information. Tests suggest that few clients consider the format of a server's overview data format is in line with the current NNTP standard. Most clients do not retrieve the overview.fmt file in order to use the overview format. The reason is probably that overview data has not yet been used to distribute application-specific meta information. Many news clients actually provide mechanisms to deal with overview data, such as article filtering based on the server's overall performance. That being able to handle extended overview data is just a question of becoming aware of the information.

5.3 Implementation of the Interface to External Information Sources

To make use of information sources, such as databases or recommender systems, external to the news server, we developed a Perl-based interface. The interface is accessed whenever dedicated clients request overview data. Whether a client considers the format of a server's overview data format is in line with the current NNTP standard. Most clients do not retrieve the overview.fmt file in order to use the overview format. The reason is probably that overview data has not yet been used to distribute application-specific meta information. Many news clients actually provide mechanisms to deal with overview data, such as article filtering based on the server's overall performance. That being able to handle extended overview data is just a question of becoming aware of the information.

6 Practical Applications of Extended Overview Data

Two fully implemented systems that use "extended overview data" to disseminate social navigation information are discussed below.
6.1 CollabNews: Extended Overview Data and Social Navigation

The idea of social navigation is roughly to support users in navigating information spaces by providing information users. Behaving in a particular way influenced by the activity of others would then be an instance of social navigation. Concrete examples of social navigation are selecting a specific Web page or a particular Usenet article for similar selections. A broad overview of the social navigation field can be found in [Munro et al. 1999].

We developed support for social navigation in Usenet newsgroups. The social navigation tool CollabNews [Lueg 2002] interesting discussions by indicating which discussions were read by other newsgroup participants. Usually this in reading articles, as opposed to posting new articles, is a hidden activity in the realm of newsgroups. Visualization provide valuable insights into the social dynamics within newsgroups as well as into the particular interests of individuals willing to share information about their news reading behavior with other newsgroup users.

Figure 5 shows how social navigation information is provided to clients by propagating the information as part of particular implementation, the news server queries a trace database where information about the news reading collected. At the end of every single news reading session, the news server submits collected reading data to the request, the database returns the information (typically, if a specific article was read by a particular user) and the server inserts the information into the overall news reading data. At the end of every single news reading session, the news server submits collected reading data to the request, the database returns the information (typically, if a specific article was read by a particular user) and the server inserts the information into the overall news reading data. Clients can then use the information according to their built-in filtering and visualization example, articles could be highlighted in different colors according to who reads the articles.

![CollabNews Diagram](image)

Figure 5. Providing social navigation information to regular news clients by injecting the information into the overview data. In this example, the news server uses the Perl interface to access an external database containing social data.

6.2 SelectNews: Extended Overview Data and Collaborative Filtering

Collaborative filtering [Goldberg et al. 1992] aims to provide users with specific information that helps them choose alternatives. The idea is that users help each other to distinguish between high quality and low quality items by they have investigated. These ratings are collected and can then be used by others to focus on those items collected (rated acceptable). Collaborative filtering has been implemented for various domains ranging from Usenet articles [Resnick and Varian 1997] for a broad overview).

Figure 6 shows how rating information requested from a collaborative filtering system is provided to news clients information as part of the overview data. In this implementation, the news server connects to the SELECT collab [Alton-Scheidt et al. 1999] and sends a list of message-IDs generated from overview data describing the articles in SELECT server generates ratings for the articles and returns a list of ratings for the message-IDs. This information is inserted into the overview data sent to clients, as shown in Figure 4. Clients can then use the information, for example, to highlight articles that received low ratings and to highlight articles that received reasonably good ratings.

To generate the collaborative filtering information, the SELECT server processes trace data describing the news...
users. The trace data is collected by the news server and is delivered to the SELECT server at the end of single n current implementation, the news server submits information indicating how much time the user spent reading s time has been identified as a good indicator of user interests [Morita and Shinoda 1994].

Figure 6. The news server uses the Perl interface to request rating information from a SELECT collaborative filter based on HTTP

7 Conclusions

The paper has presented a transparent and flexible means to provide off-the-shelf news clients with user-specific characteristic of the approach is that it builds on existing Usenet standards while offering further opportunities for extensions.

The examples discussed show that using overview data to distribute meta information to news clients is a promising area to go from a couple of prototypical implementations to a widely used method for distributing n clients from news servers. However, increasing interest social navigation and collaborative filtering can be expected an extension.

Clearly, a server-centered approach such as the one discussed has certain limitations. Users have to use a particular information from external sources are not propagated to peer news servers (for good reasons). This also means that social navigation setting have to access the same news server. We consider these limitations to be reasonable as remote servers independently of their physical location. The only design alternative would be to introduce a what the Better Bid Bureaus developed by [Resnick et al. 1994]. Then again, significant modifications of clients would what we wanted to avoid.

Future work includes investigating what kind of information is most useful when browsing Usenet newsgroups, an be encoded in such a way that most off-the-shelf clients can act on the information using built-in filtering capabilities. The benefit of providing templates for inserting user-specific information into the overview data is documented reference implementation.

Acknowledgements

The technical work reported in this paper was done while the author was with the AI-Lab, Department of Informatics of the University of Zurich, Switzerland. Funding was provided in part by the Swiss Federal Office for Education and Science 98.0065 (SELECT project in the EU Framework IV / Telematics Programme). The author is grateful to Rolf Pfeifer research environment in Zurich, to Jim Underwood for his ongoing support in Sydney, and to Jacob Palme, Igor H anonymous reviewers for helpful comments on draft versions of this paper. Some of the figures were originally provided at the Annual Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society and are reproduced with permission.

References

Alton-Scheidl, Alton-Scheidl, R., Ekhall, J., van Geloven, O., Kovacs, L., Micsik, A., Lueg, C., Messnarz, R., N

