The Private Filtering Agent
A Short Description

Fredrik Kilander
fk@dsv.su.se

April 28, 1995

1 Introduction

This document describes work in progress. Parts and components are still under development. As a result, the text may be terse or speculative in places.

The Private Filtering Agent (PFA) is a system which classifies messages[1] directed at a person (the user), but which have not yet been seen or read. The concept of a private agent underscores that each PFA is individually tailored to the requirements of each user. This adaptation happens as a result of the user's explicit commands, implicit directives in the form training examples and possibly also as a consequense of the user's general behaviour.

Messages can accumulate either in a user's mailbox (electronic mail) or in conference systems such as Usenet News. The rate of accumulation depends on how often and fast the user consumes any outstanding messages and the number of sources the user subscribes to.

The PFA runs as a background process in the user's computing environment. At regular intervals it examines the locations where messages are known to accumulate (the mailbox or the server for Usenet News) and *classifies* each new message according to the user's interests as the PFA knows them. The classifications are stored with message identifiers in the user's private database.

When the user decides to read new messages, the reading program is equipped to take advantage of the classification database. Using the classification affixed to each message by the agent, the reader has the ability to rearrange the presentation sequence to allow the most interesting messages (or message threads) to be shown first.

To make the system complete a device is needed to communicate with the classification agent. The user must be able to submit or change outstanding examples of (dis-) interesting texts and to examine the agent's motivations for a particular classification.

[1] Currently the message type is restricted to Internet Messages (RFC 822).

1.1 System Description

The PFA consists of three active components and a number of passive, or secondary components. The active components are the agent, the reader and the combined trainer and editor. These interact through the user's `.newsrc` file, the classification database and the trainer database.

2 The Agent

The agent is a stand-alone process, running independently and out of sight from the user's regular working procedures. At regular intervals or predetermined times it wakes up to do its work.

There are two possible reasons for the agent to classify a text. The most common is that the text is not previously classified and has recently appeared at the user's regular server for Usenet News. For each pertinent newsgroup, the agent compares the messages available in the server with the information in the user's `.newsrc`[2] file.

The other reason is that the text *already has* a classification, but updates to the trainer database requires a re-classification.

For each message to be classified, the agent acquires the whole message, performs its analysis and records the classification in the classification database. The message is then discarded.

3 The Reader

In order to be useful the reader must be able read the information in the classification database. As a general assumption, each session with the news systems visits a number of newgroups. Within each group the user encounters a sequence of discussion threads. Since it would be awkward to present messages in a sequence determined by their classification alone, the classifications assigned by the agent are summarized for each thread, and the sequence of threads is the one manipulated. Likewise, the combined classification weights for the unseen messages in each group are used to alter the sequence in which groups are visited.

In summary, the task of the reader is to extract the classification of each message (if available) and perform the appropriate sorting. The selectivity among the facets in each classification enables the user to make short-term adaptations of the message filtering behaviour. For instance, in moments when time is short, preference could be given to short and important threads, while other occasions may favour more massive threads.

4 The Classification Database

The classification database is a regular file. It is updated by the agent and read by the reader. For each message id (in Internet Message format) it has

[2] The `.newsrc` file stores a reckoning of the texts read at a particular newsserver.

an entry of the classification performed by the agent. Each classification may consists of a final judgement, a number of smaller values and real and computed properties of the message.

The database is not intended to supersede the role of the `.newsrc` file in the role of flagging seen and unseen messages, but rather to communicate the agents opinions to the reader and trainer.

5 The Trainer and the Trainer Database

The trainer implements the user interface to the agent. It may be an integrated part of the reader or a standalone, interactive program. It enables the user to do two things: manage a set of training examples and examine the agents motivations for any particular classification.

The set of training examples is simple a collection of concepts, named by the user with a suitable label and a binary judgement in form of *YES* or *NO*. Each concept is examplified by a collection of examples in the form of texts. These texts may be snippets from a few lines to several pages, or whole articles. Each example is also labelled as being a positive example (an instance of the concept) or a counter-example (not an instance of the concept).

The user maintains the texts by adding, updating or deleting texts, concepts and labels. Whenever this is done, the results are compiled into an intermediate form which can be efficiently used by the agent.