YA-NewsWatcher and Filtering

A major improvement offered by YA-NW over original NewsWatcher is the ability to apply filters to articles listed in subject windows. With the appropriate filters, you can label posts on subjects or by authors you find especially interesting or instructive, remove unwanted annoyance posts, and generally make your usenet reading faster and more efficient.

YA-NW classifies filters into three categories:

1) The global filter, with a filter group name ".". Any filters in this group are applied first. The global filter group matches every newsgroup. This is the best place to create common-interest filters (all posts by yourself or the almost as notable John Norstad) or common-disinterest filters (to remove all MAKE.MONEY.FAST posts for example).

2) Regional hierarchical filters, with filter group names of the form "name1.name2.". These filter groups match any newsgroup whose name starts with the filter group name, excluding the final period. For example, consider a regional filter group named "comp.sys.". This filter group will match the newsgroups comp.sys, comp.sys.mac, comp.sys.mac.comm, and comp.sys.next. These filters are applied in order of length. For example, suppose there are filters defined for the filter groups "comp.sys." and "comp.sys.mac.". When filtering the newsgroup comp.sys.mac.comm, first the "comp.sys." filters would be applied, then the "comp.sys.mac." filters.

3) Local filters, with names identical to a newsgroup name. These apply only to the newsgroup named, and are applied last, so that any filtering they apply can supersede that performed by the earlier, more general purpose global or regional filters.

All three kinds of filters are stored in a single YA-NW settings file, and define a set of filters that can be applied to articles. All settings files are stored in a special folder named "YA-NewsWatcher Settings", which must be located in the same folder as the active prefs file.

As a further refinement, you can choose to have one set of filters apply to all the user group list windows you open, or you can have individual filter sets for each user group window. The settings file that will be used by all subject windows (including those opened from a search window of the full group list) that don't have their own individual filters is named "YA-NewsWatcher Filters." Settings files that apply to individual user group windows must have the same name as the corresponding user group file. For example, the settings file for a user group file named "My Sub" would also be named "My Sub," and would have to be located in the "YA-NewsWatcher Settings" folder.
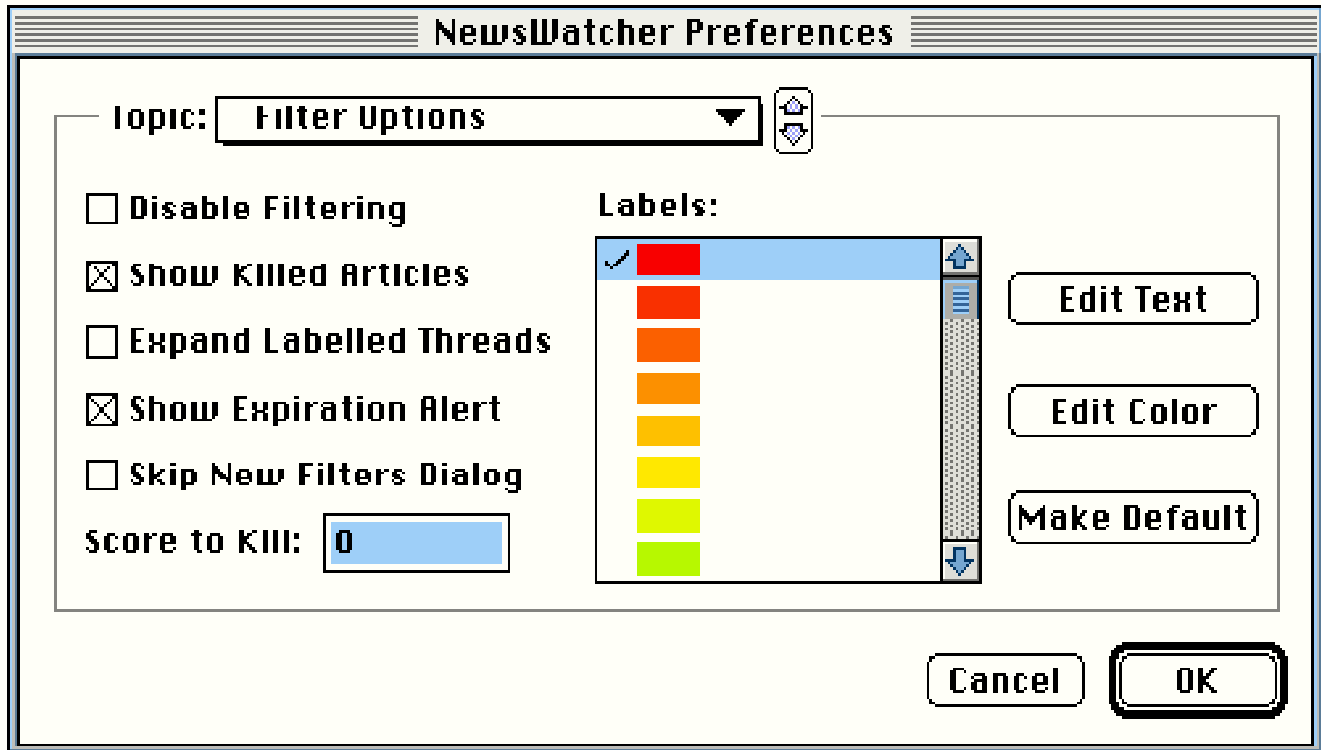
There is a new section of the Preferences dialog that deals with filters.

"Disable Filtering" turns the filter option off, so that all groups read after this is checked will not have articles labeled by whatever filters you have defined.

"Show Killed Articles" is intended to let you still see killed (unwanted) articles in message windows (where they will be marked as read). Normally they are removed from subject windows.

"Expand Labeled Threads" is pretty much self explanatory,. If a thread contains an article that has been labeled, the thread will be automatically expanded when displayed in a subject window.

Filters can be set to expire a number of days after they have been created. Expired filters are removed from the settings file. If "Show Expiration Alert" is checked, the Status dialog will show at program startup if any filters have been expired.

If "Skip New Filters Dialog" is checked, when you create a new user group window you will not be asked if you want to also create an individual settings file for this window. The window will then use only those filters defined in the special "YA-NewsWatcher Filters" settings file.

YA-NW filters support the concept of assigning an importance weighting factor, or score, to filters and articles. As filtering progresses, each article's score is increased by the score value assigned to every filter that matches the article. For example, suppose that John Norstad posts an article titled "New NewsWatcher Version." You have defined three filters. One filter labels all articles whose From: header includes the word Norstad. The score for this filter is 500, because you are very interested in what John has to say. A second filter labels all articles whose Subject: line includes the word NewsWatcher. The score for this filter is 100. A third filter labels all articles whose Subject: line contains the word money. The score for this is set to -500, because you hate to see MAKE.MONEY.FAST posts all the time.

So, when filtering articles, YA-NW tries all three filters. The first two match John's post, assigning a score of 500 + 100 = 600 to the article. A MAKE.MONEY.FAST post gets a score of -500. If you set the "Score to Kill" entry in the Filter Options dialog to 0, the MAKE.MONEY.FAST post will be killed (removed from the subject window) because its score is less than the value you set.

The remainder of the items in the dialog are used to customize the labels that are applied to filtered articles. There are a total of 34 possible labels. They are displayed in order of importance from top to bottom. Threads with articles assigned labels from the top of the list will sort to positions higher in subject windows that threads containing lower rank labels, assuming that the sort by label option has been enabled. The top 32 labels can be dragged to establish a different order of importance. The color of these labels can be edited by selecting the item in the list and clicking on the "Edit Color" button. Just like Finder labels, a label can have both a color and a name associated with it. The name can be edited by selecting the label in the list and clicking on the "Edit Text" button.

The bottom two labels are special. They can't be edited or reordered. (Unlabeled) is used to mark an article with a filter without otherwise labeling it. This would usually be used when the "Articles Not Matched Are Deleted" option for a filter group. (Killed) is used to remove annoyance posts from subject windows.

The "Make Default" button is used to make one label the default for newly created filters. The default label is shown with a checkmark beside it in the labels list. It does not need to be the topmost label.
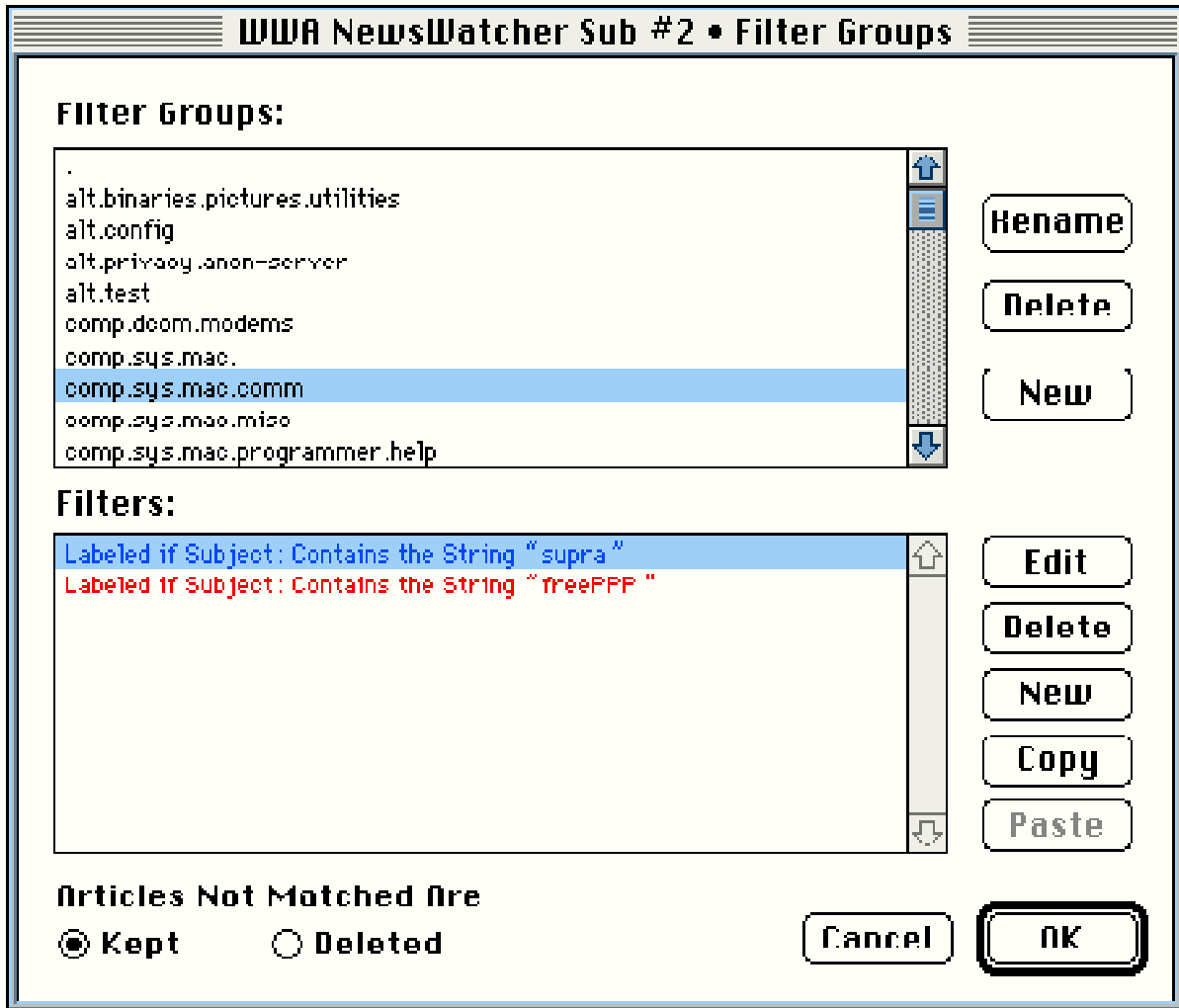
A Filters menu has been added to the program. This menu is broken up into two sections. The topmost is used to view and edit filters.

"Edit Filters…" brings up a two-paned moveable modal dialog showing the filters groups within the associated settings file in the upper list, and the filters defined for the selected filter group in the lower list.

The three buttons to the left of the Filter Groups list are used to edit the list. Clicking the "Rename" button lets you change the name of a filter group, effectively assigning all the filters defined for that group to a different newsgroup. The "Delete" button is used to remove a filter group and all the filters it contains. The "New" button lets you create a new, initially empty filter group.

The five buttons to the right of the Filters list are used to edit the list. Clicking on the "Edit" button, or double-clicking a filter entry, causes an article filter dialog for the selected filter to be displayed (see below). "Delete" removes the selected filters from the list, while "New" allows a new filter to be created. "Copy" and Paste" do what you'd expect, allowing you to copy and paste filters from one filter group to another.

There are two radio buttons in the Group Filter dialog which determine which articles are killed. If "Kept" is selected, then all articles not matching any kill filters are shown, and the ones marked for labeling are so displayed. If "Deleted" is selected, only articles matching the various keep/label filters will be displayed. All other articles are killed. Use this option when you only want to see a few specific articles within a newsgroup.

The Filter menu command "New Filter…" lets you add a filter to the appropriate settings file.

When editing a filter, via clicking on the filter in the "Filter Groups" dialog box, or via a "New Filter…" menu commands; you are presented with the "Article Filter." The dialog is intended to be read from top to bottom left to right as an English sentence.

WWA NewsWatcher Sub #2 • Article Filter

For `comp.sys.mac.comm` ▼

Subject ▼

Contains the String ▼   ☒ Ignore case

`freePPP`

Label: [red] ▼

Weighting factor: `0`

Expire in `0` days

[Cancel] [[OK]]

The topmost type-in popup menu is used to define the filter group for the filter. The initial value corresponds to the current newsgroup when a new filter is being created. Below this is a popup menu used to select which header line will be used for filtering. Most headers support filtering by matching text in the header line. For example, you can choose "From" as the header and then in the text box below the popup menu enter the text "John Norstad" to filter posts by the author of the original NewsWatcher program. Just above the textbox is a third popup menu that is used to select what kind of text matching will be performed. Most often the "Contains the String" item is the best choice, but you can also create filters that look for specific words, or phrases that start or end with a specific string. By checking the "Ignore case" checkbox, string matching will treat upper and lower case characters as the same. Much more powerful (but also slower and more difficult to use) matching can be done by selecting the "Contains the Reg. Exp." item to perform regular expression matching. The syntax for regular expressions understood by YA-NewsWatcher is described in the file "Regular Expressions."

YA-NW does its best to insert text from all the available headers into the textboxes shown for all the possible settings of the header popup menu. If the filter dialog is opened when an article window is topmost, all the headers from that article will be placed in the appropriate textboxes. If the dialog is opened from a subject window with a selected article, text from the available headers (those displayed in the subject window) will be placed in the textboxes. If no article is selected, the textboxes will all be empty.

Note that filters for the Date and Line headers do not use string matching. For dates you instead choose to filter articles more than a given number of days old, or less than a given number of days old. YA-NW is smart enough to know that if you enter a number of 2 for "less than," and 4 for "more than," that you want to filter articles that are less than 2 OR more than 4 days old; while having the numbers reversed means you want to filter articles that are less than 4 AND more than two days old. Filtering on lines works in a similar manner. With dates there is an option to filter or not filter articles with missing or invalid date headers. For lines there is an option to filter or not filter articles with missing or zero line counts.

The label popup menu is used to set what kind of a filter is being defined. The uppermost menu items are used to mark posts with a specific label color and name, similar to Finder labels. The colors, labels, and order of these items can be edited in the "Filter Options" portion of the Preferences dialog. The bottom two entries are special. (Unlabeled) is used to mark a post as of interest without assigning a label color or name. The final entry, (Killed), is usually used to remove unwanted annoyance posts like MAKE.MONEY.FAST. The order of the items in the label popup menu defines their priority, with the topmost being the most important. When sorting by label, threads containing articles labeled with the topmost label will be displayed at the top of subject windows, followed by any threads labeled with the second from the top label, etc.

A filter can be set to expire a specified number of days after it is created. This will be useful when a filter is used to select articles of passing interest. Without this option the filters file would continue to grow as new filters are added, and the time it takes to filter subject windows would also continue to increase. Setting the expiration time to 0 makes the filter perpetual.

Sundry notes:

Filter priority. The rank of a label (its order in the list of labels) is used to determine whether a later filter will override a previous filter's labeling of an article. The rules are:

1) A kill label always overrides any previous label. A kill label may itself be overridden by any subsequent filter. Thus it is possible to unkill an article with a later filter.
2) Non-kill labels will only override a previous label of equal or lower rank.

Killing articles by score. After all filters have been applied to the articles in a subject window, a final pass through the posts is made to mark as killed all articles with scores less than the value designated in the Filter Options dialog. This permits somewhat more selective killing than would be the case if articles were killed as soon as their score dipped below the kill threshold. For example, you might want to kill all posts with "money" in the subject lines unless you're reading alt.make.money.fast. You would then set up a global filter to score at -500 all articles with money in he subject line, and a local filter in alt.make.money.fast to score these articles at +500. Then the posts will be killed everywhere except in alt.make.money.fast (assuming the default kill threshold score of 0). Also, when filtering using scores, the score is applied for every filter match, even if the filter is a (killed) or (unlabeled) one

A filter icon has been added to message windows. When checked, a global filter will be created for the message you create. The filter is added all the open settings files. The filter is configured to use the topmost (highest priority) label, filter on the Subject: header, and expire in 10 days. The topmost rather than the default label is employed in case the user has set the default to (killed).