NRT (News Retrieval Tool) A User's Guide

Section Page

1 Introduction

1.1 What is NRT ?

NRT the News Retrieval Tool is the result of a six month research project between Glasgow University and the Financial Times. The aim of the project was to improve the retrieval methods and the user interface of the existing Profile retrieval service.

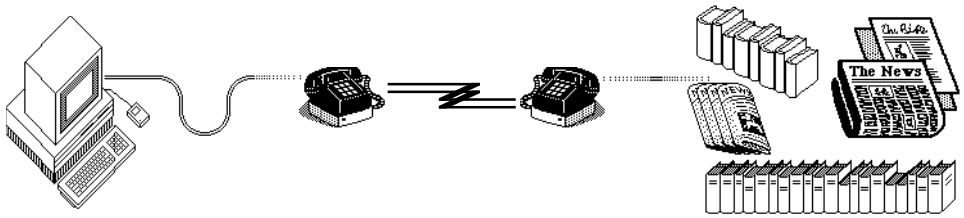

fig 1

Figure 1 shows a diagram of a test system that has been built by the project. The basic set up shows two computers, on the left a Macintosh and on the right the Profile search engine, communicating over a phone line. The Profile computer, using a new set of search algorithms developed by the project, searches a store of the 1989 articles of the Financial Times. The Macintosh runs a graphical user interface that presents an easy to use front end to the remote Profile search engine.

2 First things

When you start up NRT, the first thing you will see is it connecting to the Profile computer (see fig 2).

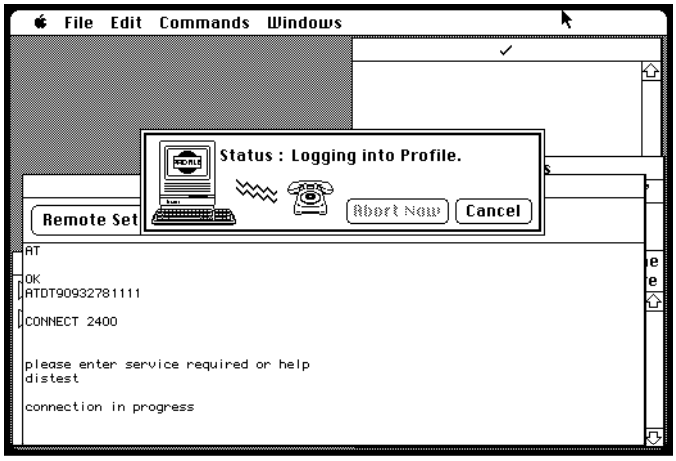

fig 2

If it seems to be failing to connect, read section 6.2, Remote connection.

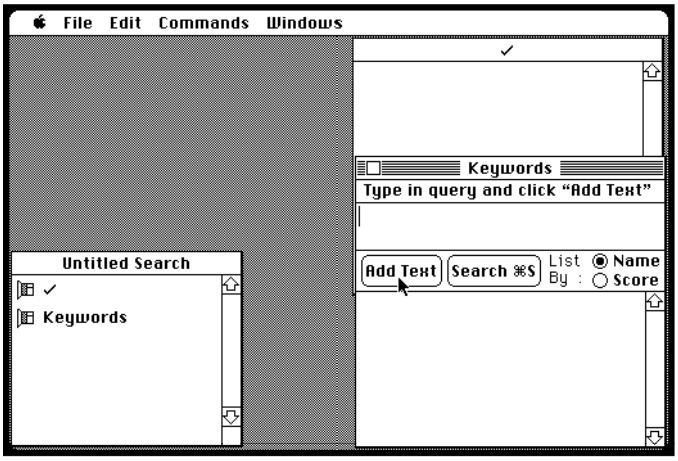

fig 3

If it successfully connects, the "Remote Connection" window will disappear and the screen will look like fig 3. You will see that there are three windows.

• "Keywords" window (bottom right) displays your query to the Profile computer.

• "✓" (tick) window (top right) will be covered later.

• "Untitled Search" (bottom left) is the root of the search. It holds a list of all the windows belonging to the search. If you wish to name your search then this will be shown in the window's title.

2.1 Starting a query

To initiate a query, first type a textual query into the text box in the "Keywords" window (see fig 4).

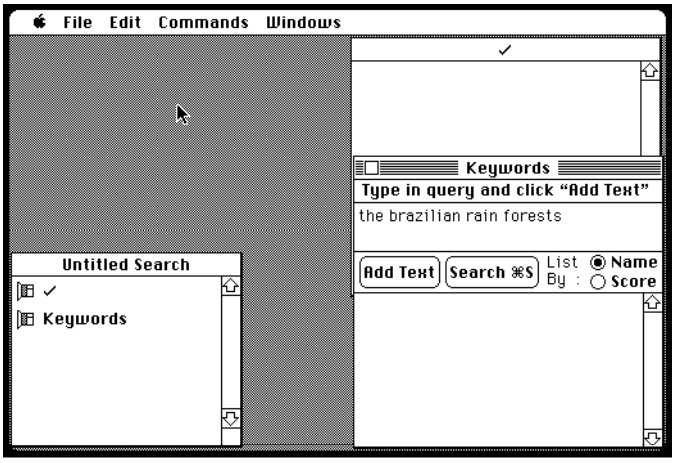

fig 4

When you have finished typing, click the "Add Text" button to add the text to the query (pressing the return key on the keyboard has the same effect as clicking "Add Text") (see fig 5).

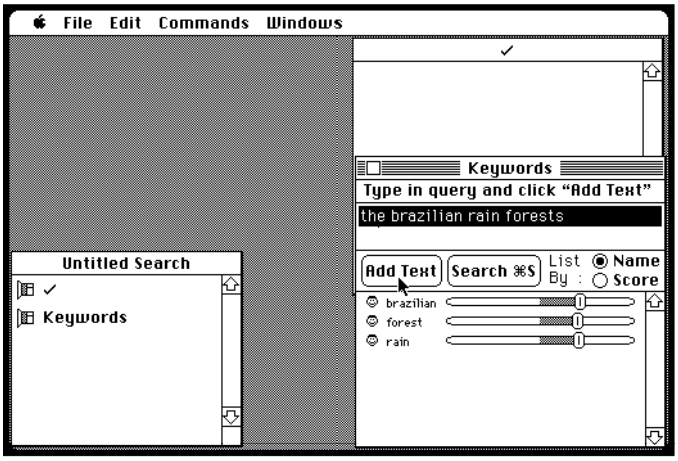

fig 5

You will now find that a list of words has appeared in the window. These words make up your query. You may see that not all the words typed in are in the list. This maybe because the word doesn't appear in any document in the store or because the word appears in too many documents to be useful (e.g. words like "the", "and", "it", "but" etc.). You may also notice that some of the words have been shortened. This is due to an automatic process called conflation. Conflation strips common suffixes (e.g. "s", "ies", "ing", "tion", "ed" etc.) from a word. This is done to make a query match to a greater number of documents (e.g. "forest" will match to "forests", "foresting", "foresters" etc.).

Each word in the list is assigned a weight which indicates the importance of the word. A word that appears in many documents has a low weight where as a word that appears in few documents has a high weight. The weight of a word is shown as a slider on the right hand side of the list. The further right the knob of the slider is, the higher the word's weight.

Notice that next to each word is a small smiling face icon (☺). A face by a word indicates that the word has been entered by the user. Later on you will see other ways in which words can be added to your query.

At anytime during a search, you can add more text to the "Keywords" window.

2.2 Performing a search

To start a search, you click the "Search" button. This makes NRT send a search request to the remote Profile computer. It replies with a list of documents. These are displayed in a new window called "Retrieval No. 1" (see fig 6). The documents are listed in order of relevance with the most relevant at the top of the list.

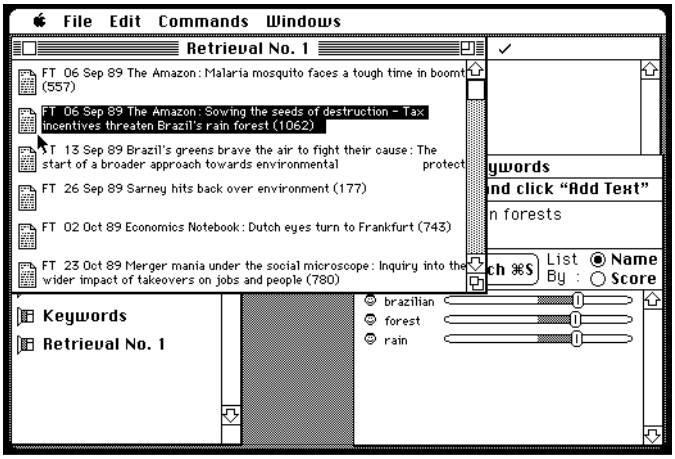

fig 6

The relevance of a document is a statistical calculation based on the number of query words appearing in a document. The more query words appearing in a document, the higher it's relevance.

2.3 Browsing the document list

You can move through the document list by using the scroll bars on the right hand side of the window. Click on the up and down arrows to move up and down through the list.

2.4 Opening a document

If you see a document that is interesting and you want to see its contents, double click the mouse (click the mouse twice in quick succession) on the document's heading. This makes NRT send a request to the remote Profile computer for the contents of the document to be sent. The contents are displayed in a new window (see fig 7). Note that the icon (the small picture to the left of the document header in the "Retrieval No. 1" window) of the opened document, in the retrieval result window, has changed to indicate that the document is open.

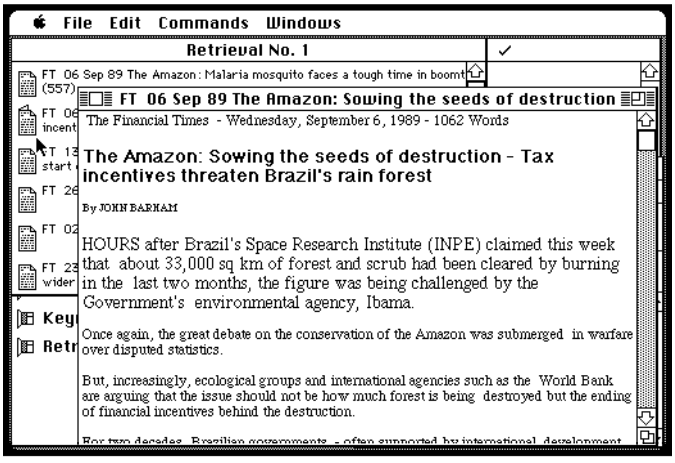

fig 7

2.5 Browsing a document

You can move through the document by using the scroll bars on the right hand side of the text window. Click on the up and down arrows to move up and down through the document. To clear the window from the screen, click the mouse in the window's go away box (the small

white box in the top left hand corner of the window).

2.6 Relevant documents

Up until now, you have specified your query using words that you have thought of. However, thinking up words that will specify your query is a hard task to do especially if you aren't very knowledgeable of the subject you are searching. It is quite possible that you aren't even sure of exactly what information you want !

In order to help with this problem, NRT allows you to mark retrieved documents, that you think cover your information need, as relevant. When documents are marked relevant, NRT analyses them, picking out words which are statistically significant to the documents. These picked out words are then added to the "Keywords" window. This process of marking documents as relevant is known as relevance feedback. It is a very powerful and useful utility as it enriches your query with words from the documents that you won't have thought of (e.g. marking the Brazilian rain forest articles as relevant might add words like "ecological", "Amazon", "jungle", "greenhouse" etc.). It has been found that relevance feedback broadens a query with subsequent searches bringing in documents on related topics.

Because the picking out of significant words is a statistical process, the accuracy is improved the bigger the sample there is. So the more documents that are marked relevant, the better NRT will be at picking out relevant words. Note the size of individual documents being marked has no bearing on the accuracy.

2.7 Marking a document as relevant

To mark a document as relevant, it has to be moved to the tick (relevant) window.

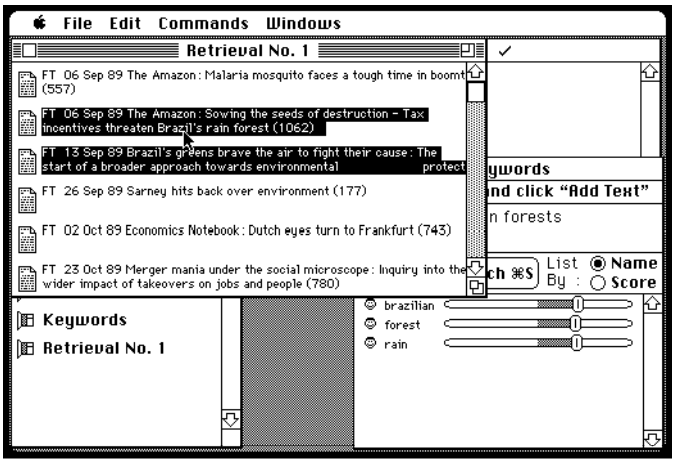

fig 8

To do this, press down the mouse button over the header of the relevant document (figure 8 shows two documents about to be marked), drag the document to the tick window (in the top right of the screen) and release the mouse button. A copy of the document will appear in the tick window (see fig 9), notice also that the document icon will change to a tick☑.

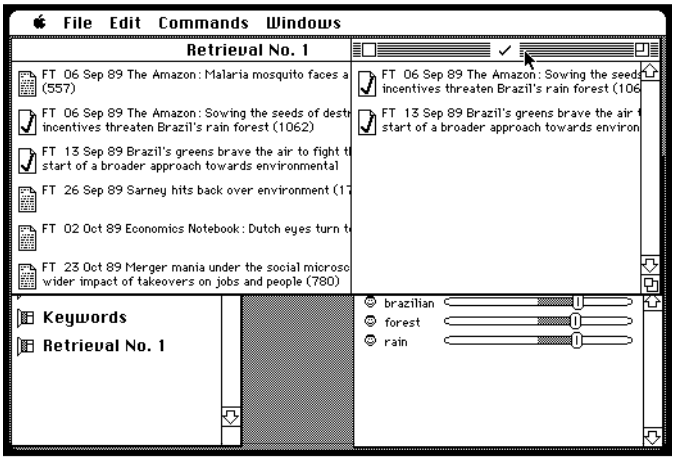

fig 9

2.8 Performing a second search

Now that the documents have been marked as relevant, a second search should be performed (by clicking the mouse on the "Search" button).

Before the search takes place, NRT checks to see if the contents of the tick window have been changed since the last search. In this case it has, so NRT first analyses the documents in the tick window and adds news words to the Keywords window using the relevance feedback techniques mentioned above. Notice that the new words added to the Keywords window have a small document icon (🗋) next to them. This indicates that the words have been added by NRT using relevance feedback.

After the new words are added, the search is performed. Another window with a new set of documents will appear. The documents marked as relevant will appear near the top of the list.

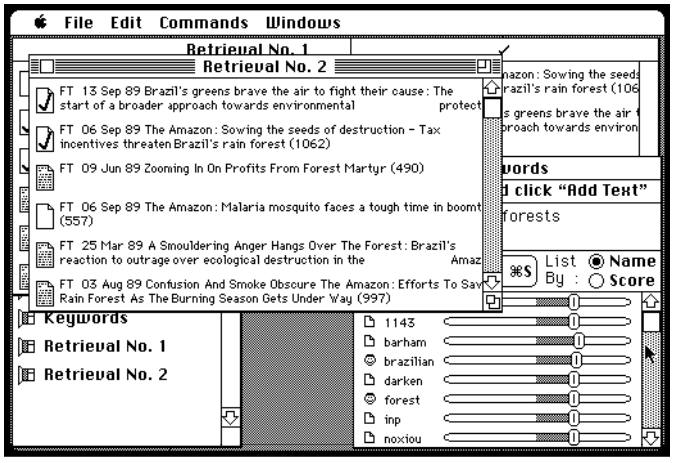

fig 10

You will notice that some of the documents have blank icons 🗋. This indicates that the documents have already been retrieved and can be found in at least one of the other retrieval windows. A document with a filled, text icon 🗎 indicates that this is the first time the document has been retrieved in this search (see fig 10).

If subsequently, you change the contents of the tick window and perform another search, all the document icon words will be deleted and a set of words reflecting the new contents of the tick window will be automatically added.

3 Controlling your search

3.1 Adding text to your query

As has been said already, you can add text to the “Keywords” window at any time. So far you know how to add text by typing. It is also possible to add text directly from the contents of a document. Suppose you are interested in the “Space Research Institute (INPE)” and you want to add that text to the “Keywords” window.

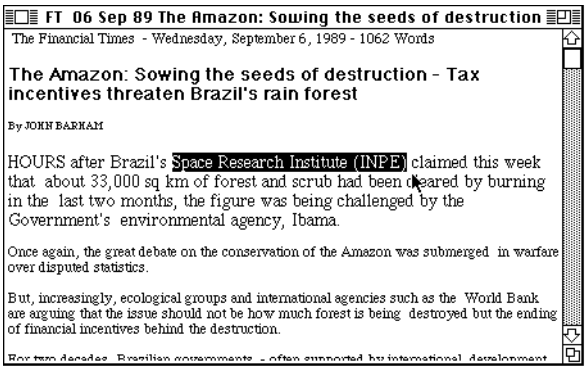

fig 11

First select the desired area of text (see fig 11), then choose the menu command “Add To Query” (under the “Edit” menu).

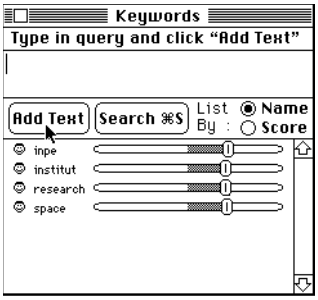

fig 12

Figure 12 shows the selected text added to “Keywords”.

3.2 Changing the number of documents to be retrieved

The standard number of documents retrieved from a search is 15. To change this number, choose the “Num of Docs” command under the “Commands” menu (see fig 13).

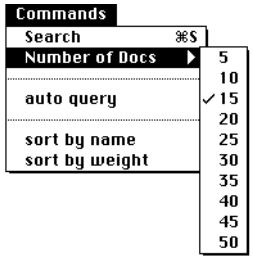

fig 13

A submenu appears showing a menu of numbers with a ✓ (tick) by 15 (the current number). Choosing a new number from the submenu changes the number of documents to be retrieved

in the next search.

3.3 Lists

The majority of actions you perform using NRT are on lists. What follows are the operations available to you for list manipulation. They apply to all the lists in NRT (e.g. “Keywords”, “Retrieval No. *”, “✓” and “Untitled Search”) unless otherwise stated.

• The menu command “Select All” (under the “Edit” menu) selects all the items in a list. Note, “Select All” will also select all the text in a document window.

• The menu command “Clear” (under the “Edit” menu) removes the selected items from a list. You can use this to remove documents from the “✓” (tick) window. It is also possible to do this by dragging a ticked document out of the tick window. This command doesn’t work in the “Retrieval No. *” and “Untitled Search” windows.

• It is also possible to select more than one item from a list.

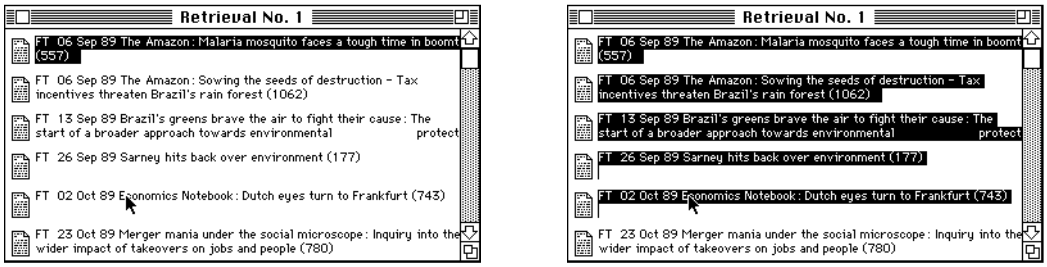

fig 14

You can extend the selection of a list by holding down the shift key on the keyboard while clicking the mouse. Fig 14 shows the effect of a “shift click” on a list.

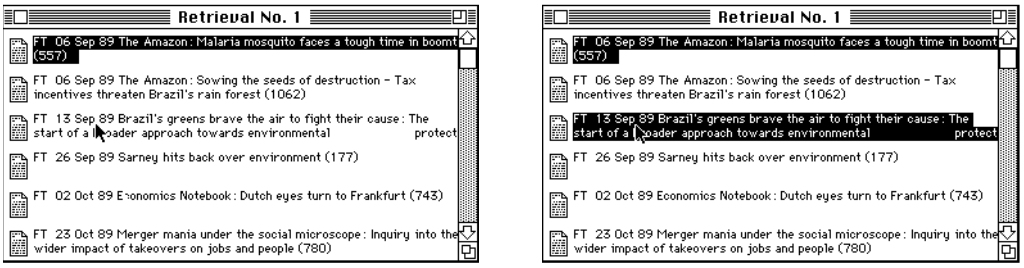

fig 15

You select individual items in a list by holding down the “⌘” (command) key on the keyboard while clicking the mouse. Fig 15 shows the effect of a “command click” on a list.

• All words in the “Keywords” window are assigned an initial weight. These however can be adjusted. Increasing the weight of a word increases it’s importance in the search and decreasing the weight decreases it’s importance. If you decrease the weight of word to a negative value, documents containing that word will become less favoured.

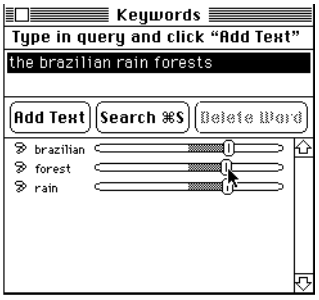

fig 16

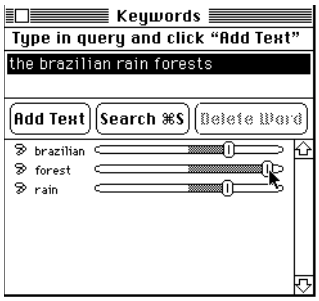

fig 17

You adjust the weight of a word by clicking the mouse on the knob, of the word's slider (see fig 16), and drag it to the new desired weight (see fig 17). The range a weight can have is -100 to +100 although the actual value of the weight isn't displayed.

3.4 Keyword list order

The words in the "Keywords" window are, by default, listed in alphabetical order. You can choose the order of the list by clicking the "Name" (for alphabetical order) or the "Score" (for weight order) radio buttons. Figure 18 shows the words listed by name and figure 19 shows then listed by score.

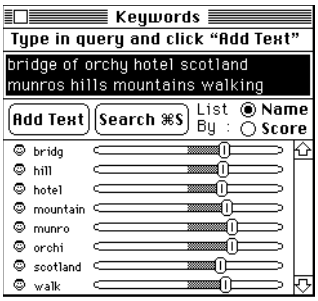

fig 18

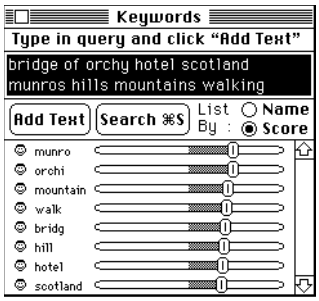

fig 19

Note if the words are being listed by score, and you change the score of a word so that the order of the list becomes invalid, NRT will reorder the list, the next time the list adjusted.

4 How to construct a good query

NRT is very different from a conventional Boolean query system, like Profile, with it's complex language full of ANDs, ORs and NOTs and retrieved documents listed in orders that aren't particularly useful. When specifying a query on these systems, the emphasis is on choosing 3 or 4 very important keywords that will whittle the list of documents down to a manageable size.

With NRT, it is completely the opposite. A query should have as many relevant words in it as possible. Because of this difference, users switching between systems will have to re-learn the way they specify their query. Below is a list of hints and ideas to keep in mind if your search seems to be getting nowhere.

• When you start a query, don't just type in the first two words that come to mind, think about what it is you want. Think of all the words that could be useful and add them to your query.

• If you see relevant documents, mark them as such and perform an auto query.

• If any of the documents you read contain words that are relevant to your query, add them.

• If none of the retrieved documents are relevant, increase the number of documents to be retrieved and repeat the search.

• Try increasing the weights of key words that are especially relevant.

• Try deleting or decreasing the weights of key words that you think may be spoiling the success of the search.

• Remember, it is possible that there just isn't a document covering the subject of your query.

5 Searches

5.1 Starting a new search

If you want to start a completely new search, then choose the "New Search" command (under the menu "File"). This will close the current search and open a new untitled search.

5.2 Saving a Search

The current state of a search can be saved to disk. This can be done in the following ways.

• The command "Save" (under the menu "File") saves the search to disk. If the search is untitled, you will be asked to supply a name for the search.

• The command "Save As…" (under the menu "File") saves the search but first asks you for a new name for the search.

• When the "New Search" command is chosen or NRT is quitted, you will be asked if you want to save the search (see fig 20).

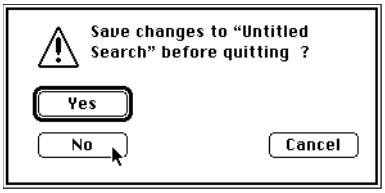

fig 20

6 Communications

6.1 Status window

When NRT is communicating with the Profile computer, the status window appears (see fig 21). The window has four sections.

fig 21

• The status message tells you what NRT is doing

• The lightning lines between the Macintosh and the phone flash to indicate activity on the communication line.

• The “Cancel” button cancels the operation that brought the status window up. Note, cancelling may take a short while to take effect (see fig 22).

fig 22

• The “Abort Now” button is greyed out and unusable until the “Cancel” button is clicked. Clicking it causes NRT to cancel immediately and it should be used only if you think something has gone wrong with the communications link or the Profile computer has gone down.
Warning, don’t use “Abort Now” as a quick way of cancelling as it is likely to cause problems.

6.2 Remote connection

The “Remote Connection” window is split into two sections. A top panel of buttons and a text window (see fig 23).

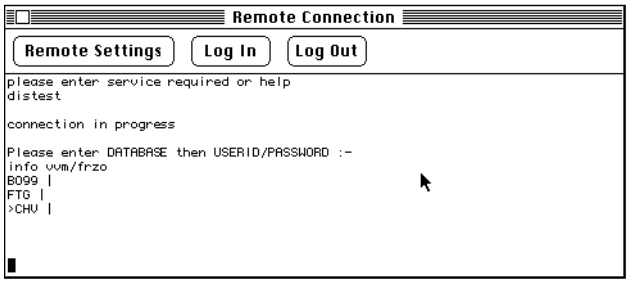

fig 23

The text window acts as a simple terminal emulator showing the activity on the communications line. If you wish, you can type into this window sending commands directly to the Profile computer.
The top panel consists of three buttons.

• Clicking the “Remote Settings” button shows a window which allows you to set the serial port settings (see fig 24).

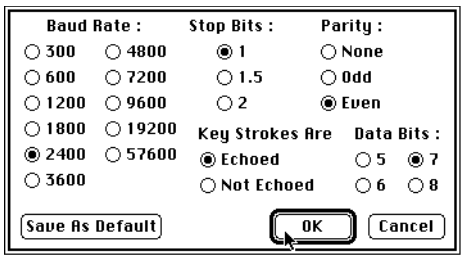

fig 24

The “Key Strokes Are” option tells NRT whether when a key is pressed down, it’s character is echoed to the screen or not (e.g. the Profile 300 baud service does not echo key strokes). The button “Save As Default” will save the settings shown in the window and make them the default settings of the program.

• Clicking the “Log In” button initiates a sequence that logs NRT onto the Profile Computer. It is possible that NRT will get into problems and you may have to manually log in. What follows is a script of the current log in sequence starting from a Hayes compatible modem. The bold text is what you type in.

```
AT

OK
ATDT90932781111

CONNECT 2400


please enter service required or help
distest

connection in progress

Please enter DATABASE then USERID/PASSWORD :-
info  vvm/frzo
B099 |
ftg  |
>CHV |
```

If you successfully type in the whole sequence, NRT is ready to communicate with the Profile computer.

• Clicking the “Log Out” button initiates a sequence that logs NRT off the Profile Computer. It is just possible that NRT will get into problems and you may have to manually log out. What follows is a script of the current log out sequence. The bold text is what you type in.

```
end  |
>END |
```

At this point you should hang up the phone line.

If “Log In” or “Log Out” successfully complete their sequence, the “Remote Connection” window is automatically closed. You can bring it back by choosing the command “Remote Connection” (under the “Windows” menu). Because having this window up risks you accidentally logging off or crashing the Profile computer a warning message appears (see fig 25) before the window opens.

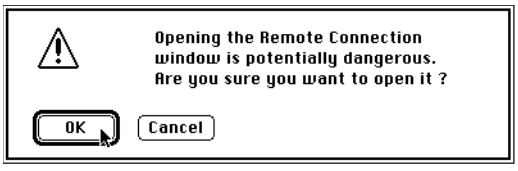

fig 25

Only have this window open if you know what you are doing.

7 Controlling the relevance feedback

Under the “Commands” menu, there is an option called “Relevance Setting…”. Choosing this causes a window (see fig 28) to appear. This allows you to change two options in NRT.

• The size of the search query sent to Profile. At present, the Profile system can only accept a search query 230 characters long. If in the future this limit is changed, you can inform NRT of this change by typing in the new value. If NRT can’t fit a search into a query, it will create a search that has as many of the highest scoring words as possible.

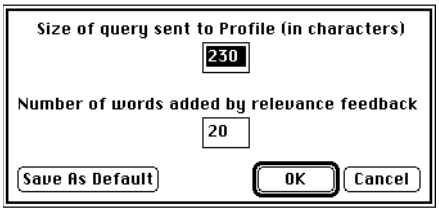

fig 28

• The second option allows you to change the number of new words relevance feedback adds to the keywords window. With the present search query limit, it is only worth setting this value to about 15. Note that there is a maximum limit of 256 for this number. If you try to type in a number bigger than the maximum, NRT will inform you of the error of your ways (see fig 29).

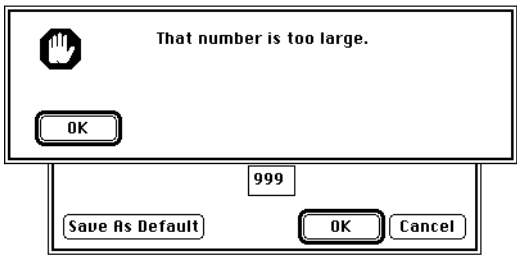

fig 29

Like the remote settings window, there are three buttons.
• “OK”, which changes the settings to the values shown in the window.
• “Cancel”, which leaves the settings as the values they were when the window was displayed.
• “Save As Default”, which will save the settings shown in the window and make them the default settings of the program.

8 Appendix of terms

Clicking

Clicking is the name given to the action of pressing and releasing the button on a mouse (see fig 30).

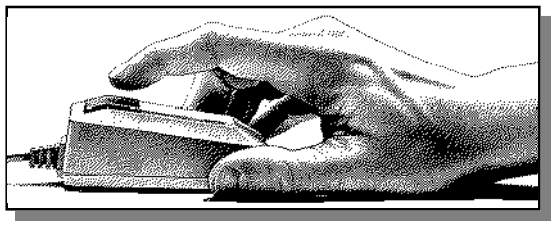
fig 30

Double Clicking

Double clicking is the name given to the action of clicking the mouse twice in quick succession.

Dragging

Dragging is the action that is performed on the Macintosh. It involves pressing down the mouse button over an object and then moving the mouse. An outline of the object follows the mouse. When the mouse button is released, the object moves to the location of the mouse.

Figure 31 shows the three stages of an object being dragged.

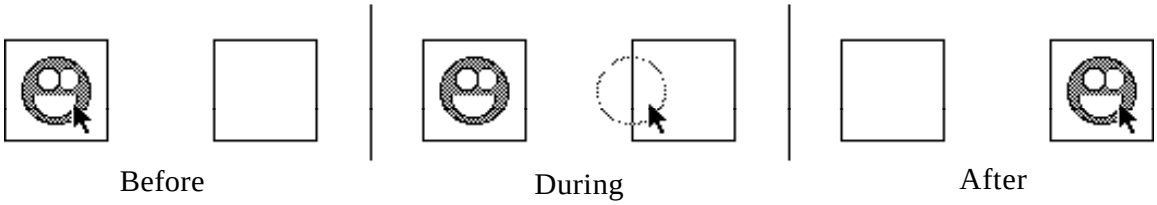

fig 31

Go Away Box

The go away box is a region of a window usually located in the top left hand corner. The user clicks in the box when he wants to close (remove) a window. Figure 32 shows a go away box being clicked in.

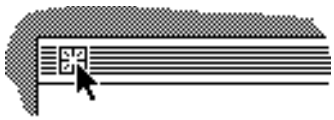

fig 32

Macintosh

The Macintosh (known as the mac) is a small graphics workstation produced by Apple. It is most notable for having a high quality graphical user interface, which has set the standard in the PC market.

Menu

Menus are the Macintoshes way of invoking commands. A menu consists of a title followed by a list of commands. The title indicates the general area that the commands in the menu cover (e.g. editing commands are in menu “Edit”). Figure 33 shows an example menu with the second command being chosen.

fig 33