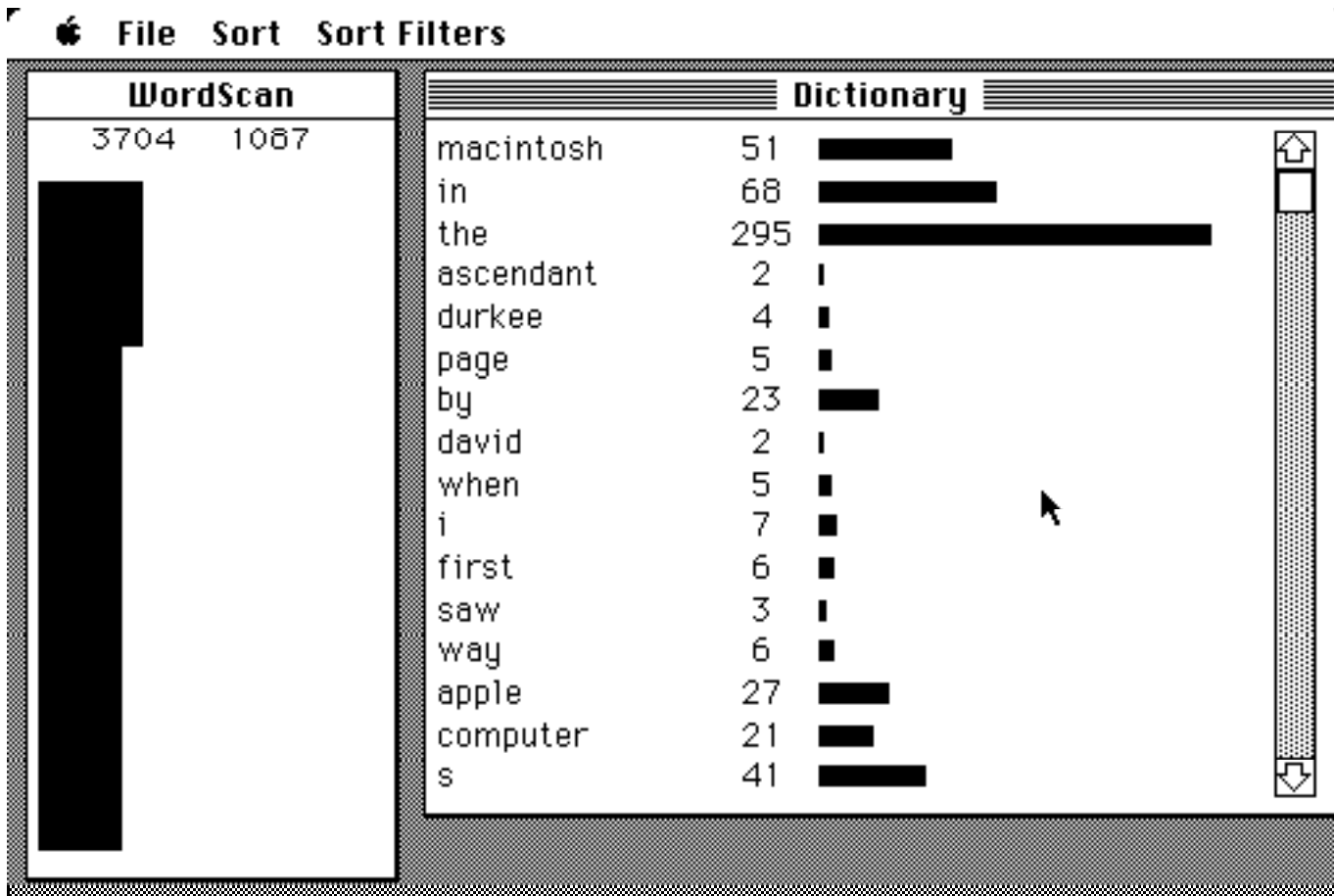


WordSCAN™

WordSCAN is a specialized text analysis program we (Online Publishing Systems, before we went Underground) had written for a special project when we first got our Macintosh. Its purpose is to look at the vocabulary content of text files: to show what words are being used and how frequently. It can give a gross breakdown of word use, showing all the words used in the document, or it can filter the word list with a user-created exception list. It can be told either to ignore all the words in the exception list or to sort and display only the words that appear in the exception list. And it sorts in any of three different ways.

We used **WordSCAN** to analyze newspaper stories downloaded from UPI for sociological trends. It's really useful for that sort of work: if you are looking for sexist trends in writing, for example, you can search a large body of text against an exception list of gender oriented words. But we see possibilities for the program beyond that. A computer magazine geared toward the naive user could survey their articles for jargon content. A writing teacher could use it to look at vocabulary and diction in student essays. And there are probably a lot of uses we haven't considered.

When you first run the program from the Finder, you get three menus and two empty windows. The first thing you would want to do is give it a text file to scan. Do this by selecting **Scan File** in the **File** menu. You will get a standard open file dialog box with listings of text-only files. The program can't read formatted MacWrite files, but it can read straight text files, which can be created with MacWrite, Word, MockWrite, Edit, File, *et cetera ad infinitum*, or which can be downloaded from nearly any network or BBS.



Screen 1: A text file scanned.

While the file is being read into memory, a black bar will grow in the **WordScan** window. Each time the bar is incremented means one *unique* word has been read: a word that hasn't previously been found in the file. When the scan is done, two numbers at the top of that window will appear, indicating the total number of words in the file and the number of unique words found.

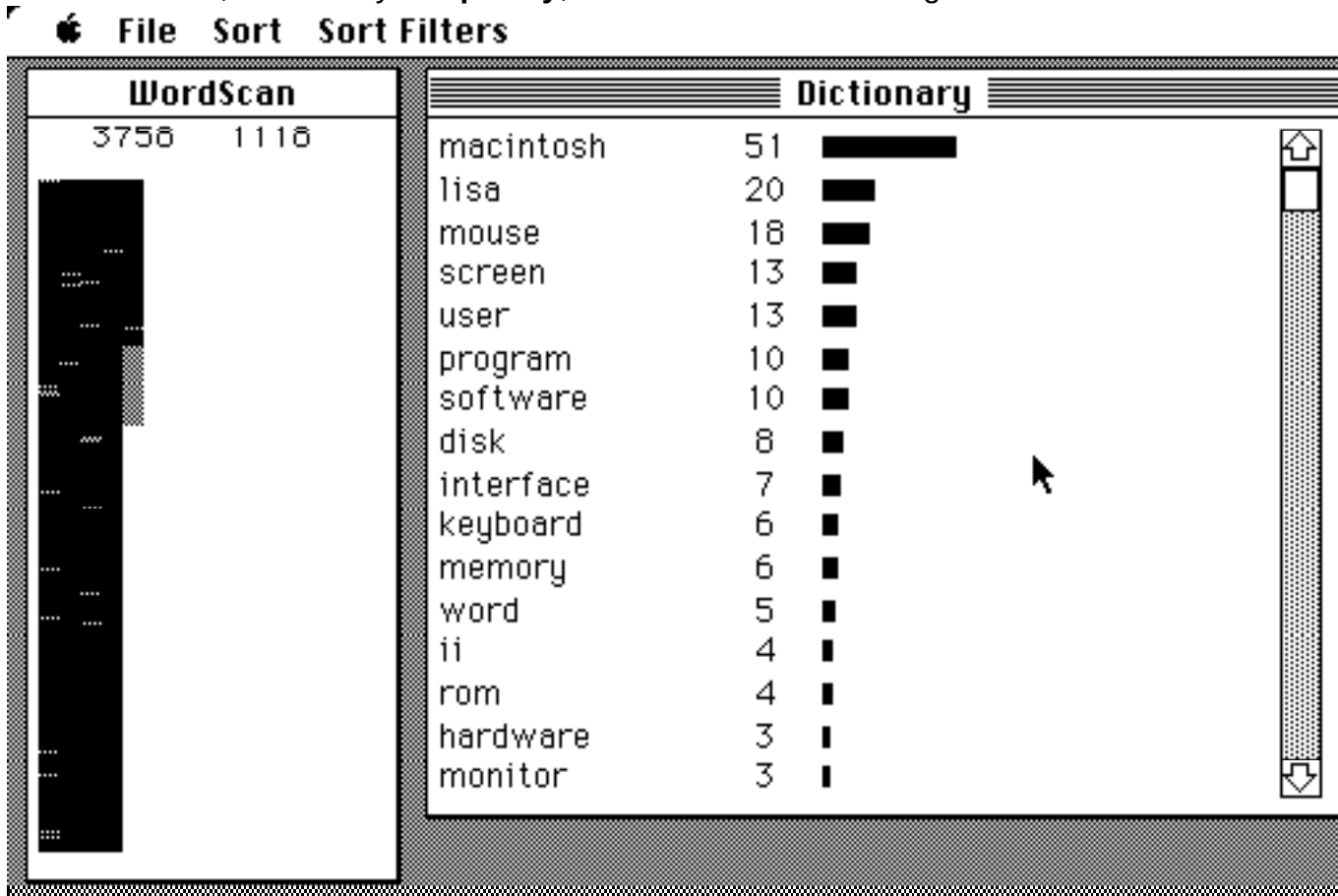
Then, a bar chart will appear in the **Dictionary** window. This is a complete list of all the unique words in the document, with each word followed by the number of times it appeared and a bar showing that number graphically.

The list is initially sorted by appearance (if you look in the **Sort** menu, you will find the first item, **Appearance**, checked). This means the first word shown was the first word in the file scanned, the next word in the list was the next unique word in the file, and so on. If you're looking for specific words (and you haven't made an exception list yet) try selecting **Alphabetical** under the **Sort** menu. Then you can scroll through the list and pretty easily find the words you want. (Be warned, though, that the scroll bar has a few bugs in it. They are nonfatal, but mildly annoying. When using the elevator, move it in small increments; the bottom of the dictionary may only be an inch down the scroll bar.)

Now try sorting by **Frequency**. This will tell you which words appear most often. You might use this feature to find overused words in your writing.

Now lets look at the real power of **WordSCAN**: the items under the **Sort Filters** menu. For this, you will need a text-only file of words separated by returns. (If you don't have a desk accessory that can create such files, like MockWrite, you will have to quit **WordSCAN**, create the list in MacWrite, save as text-only, come back to **WordSCAN**, and scan your original file again. Sorry.)

Now select **Load Exceptions** from the **File** menu and load the list you just made. You will see certain parts of the black bar in the **WordScan** window become gray as words from the exception list are matched to words in the dictionary (the scanned file). Now select **Watch List** from the **Sort Filters** menu. This tells **WordSCAN** to treat the exception list as a list of words to watch. We did this with an article about the Macintosh and a list of computer related words, sorted it by **Frequency**, and screen 2 is what we got.



Screen 2: Macintosh article filtered with a Watch List.

Exclusion List in the **Sort Filters** menu is just the opposite. It uses the exception list to tell the program which words in the scanned file to ignore. You might use this to look at a writing student's diction without cluttering the **Dictionary** window with common words like *a*, *the*, *some*, *of*, *for*, *yet*, *but*, and so on. Finally, use **No Filter** at any time to view the whole list.

The final feature of **WordSCAN** is the **File** menu's **Write Dictionary** command. This saves the contents of the dictionary window as they appear at the time (sorted and filtered) in a text-only file. You can load this file into a word processor, spreadsheet, or charting program to make reports and graphs with the data.