

The Translator's Binoculars, Part II: Desktop Search Tools and How They Can be Used to Search Reference Texts



By Naomi J. Sutcliffe de Moraes

In Part I of this article in the August issue, I described several corpus tools that can be used by linguists to search monolingual reference texts. This month, I will discuss various desktop search tools that were not created specifically for linguists, but can be used to search reference texts with differing degrees of ease.

Desktop search tools usually create an index of the contents of all the words in files on your computer. Generally, only unformatted text is stored in the index to speed access when searching. LogiTerm, mentioned in Part I of this article, also indexes files in this way.

Archivarius 3000

I downloaded Archivarius for the first time in order to review it, and now I cannot imagine how I lived without it. With eight years of files on my hard drives, finding an old file used to be impossible, and both Archivarius and Copernic, discussed below, are great at finding files containing specific words.

A search on “evapotranspiration” in my monolingual reference files on my hard drive is shown in Figure 1 on page 33. The left panel shows several

occurrences of the search term with a few words of context to each side. Clicking on the “Switch List” button at the bottom shows the same list, but with full path information for each file instead of context. The right panel, a preview pane, shows much more context for the occurrence selected on the left. Note that the context is shown as

your hard drive and give the indices names. This is very good for people who share their computer or use it for both work and other purposes. Note, however, that Archivarius has an index size limit of 1.5G that may not let you create one index of your entire hard drive, though you can create many indices and search them all

Desktop search tools usually create an index of the contents of all the words in files on your computer.

simple text with no formatting. To open the original file, click on the “Open file” button in the lower left. One nice feature is the ability to search within the preview pane without having to open the original file (note the navigation buttons at the top of the right pane). You can even search for a different word inside the text version of the file in the right pane.

You can index different parts of

simultaneously. Note that 1.5G is the size limit on the index, which contains only text without formatting or images, so the size of the indexed files would be much greater.

Archivarius includes morphological searches for many languages and indexes mailboxes for many e-mail formats besides Outlook, including PocoMail, Barca, Netscape, and Thunderbird. ➔

Figure 1: Archivarius Search Results Window

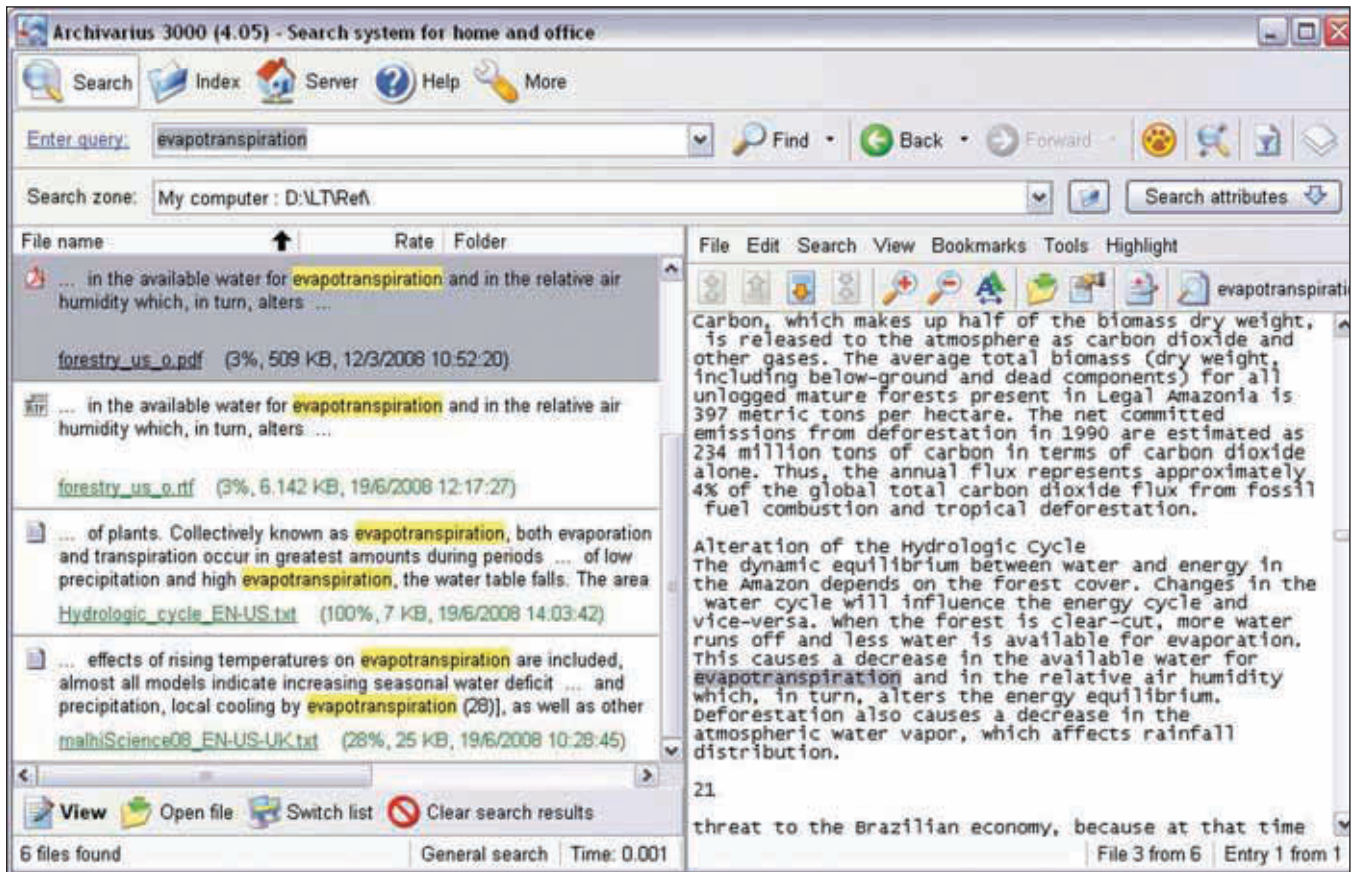


Figure 2: Copernic Desktop Search 2 Search Results Window

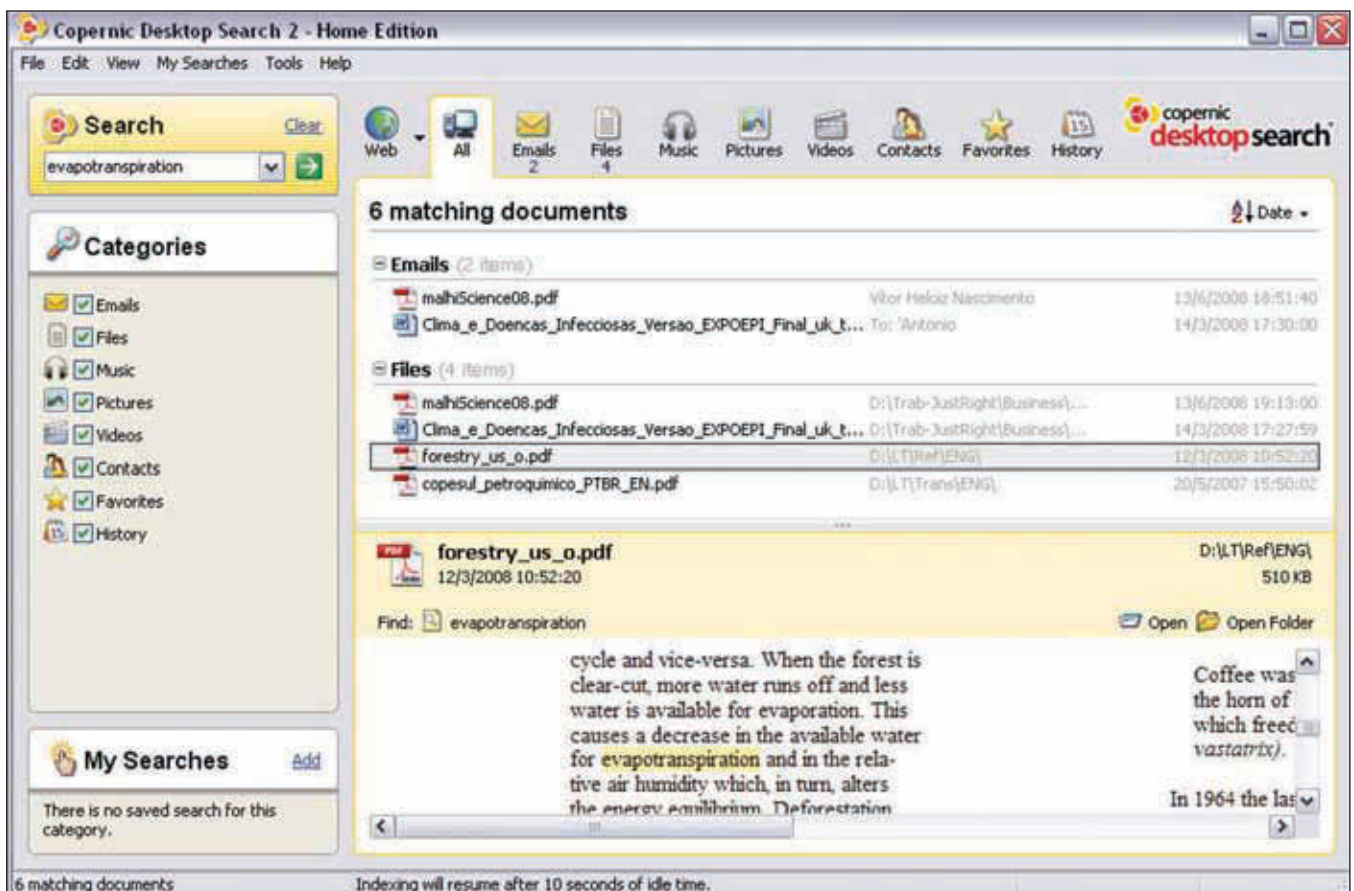


Figure 3: Google Desktop Search Results Window



Copernic Desktop Search 2

Copernic is a free desktop search program that many readers may already use. It classifies files by types and lets you search just in e-mail (but only Microsoft Outlook) or just in documents, for example. Its interface is very clean and intuitive. Unlike Archivarius, you cannot create more than one index. The program allows some special folders and drives to be excluded from the index, but is inflexible. I back up files on my D drive to my C drive nightly, and I must exclude my entire C drive if I do not want two copies of every file to appear in search results.

Copernic's search results window is similar to that of Archivarius, but it uses a viewer to open the original file in the preview pane and highlights the search term. (See Figure 2 on page 33.) Archivarius is better for the linguist because it gives you a bit of context in the left pane so you can quickly scroll through the results, whereas Copernic provides no context except in the preview pane. In Copernic, you cannot search for another word in the preview pane, though you can scroll through the file. The original file can be opened by clicking on the Open button.

Unfortunately, on my computer, Copernic hangs (sometimes for several minutes) when trying to preview large Excel files and certain other file types. There is no way to open a file through the program without previewing it first, and context can only be seen in the preview window.

Copernic does have a Corporate Edition that appears to allow customization of the directories to be searched, in addition to other features like company network searching. I did not have time to try it, but the information on the website does not lead me to believe that its interface is different from that of the free version.

Google Desktop

Google desktop is the least flexible of the most common desktop search tools and has almost no settings. It

indexes and searches all hard drives, and as a result, it often shows me several versions of a file: the original and several zipped backup copies. It shows both full path information and a bit of context around the search term. (See Figure 3.) Clicking on the entry opens the original file. It also seems to index only Outlook mail messages and contacts, and not those of other e-mail programs. Google desktop puts a search field on your Windows Start bar for easy access.

Grep

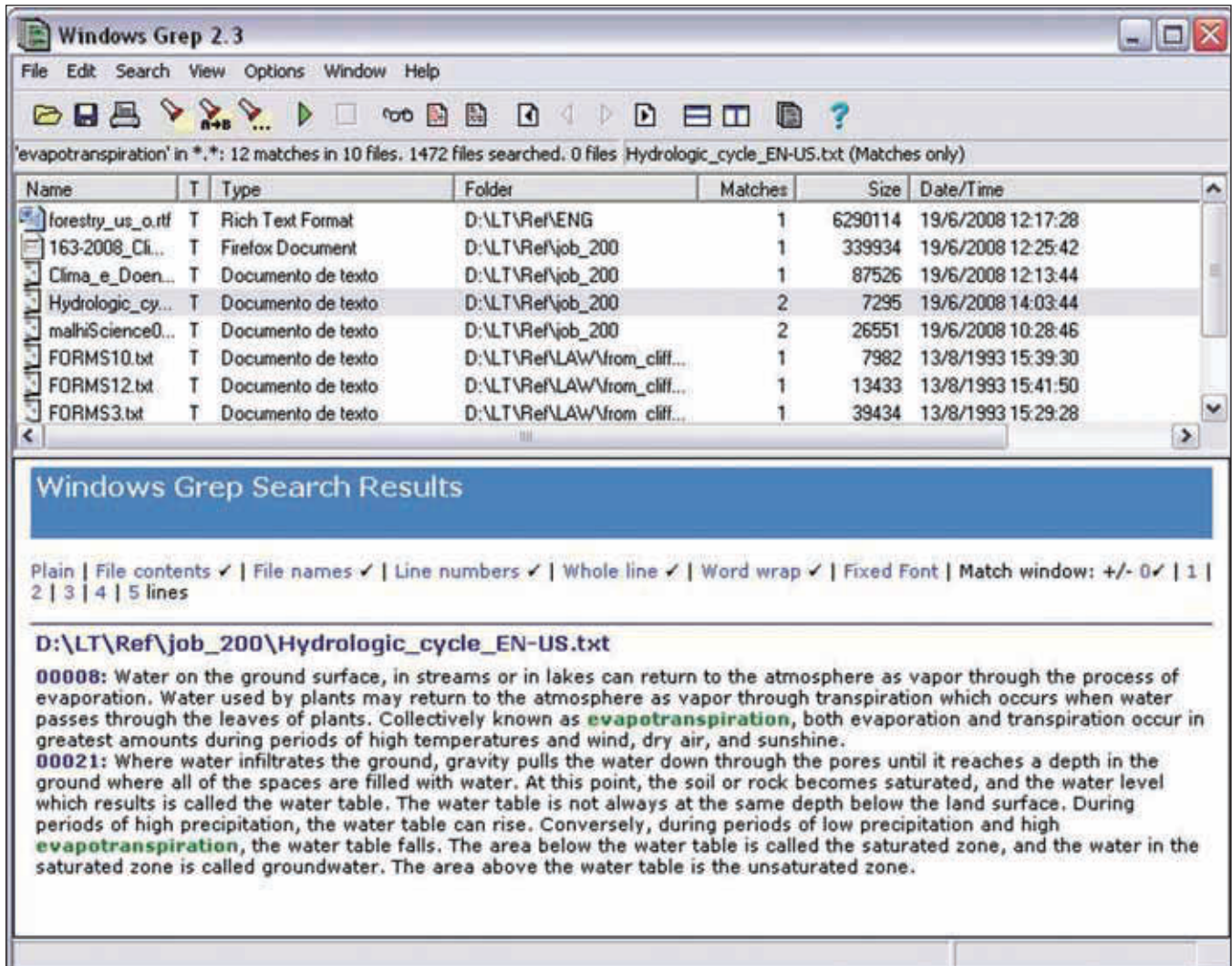
Grep is a Unix operating system utility which is built-in on Linux systems. Unlike the other tools described above, it does not index files in advance, but rather searches through them on the fly when a search is performed. The command "grep" is typed on the command line followed by the search term and the files and/or directories to be searched. Figure 4 shows an example where the file names in which the search term was found are in purple and the search term itself is in red. Note that when a word appears more than once in a file each occurrence is shown with its context on a separate line. This makes it more like AntConc and TextSTAT, which I described in the first part of this article, than the desktop search tools described above. Grep can only search text and text-like files.

There is a version of Grep for

Figure 4: Grep Command and Sample Results on Linux System

```
prompt> grep Kalman *.tex mybib.bib --color
bertsekas-rhodes-TAC71.tex:A soluo para o primeiro caso   o filtro de Kalman habitual.
mybib.bib:keywords={ Kalman filters, asymptotic stability, control nonlinearities,
mybib.bib:title={On the identification of variances and adaptive Kalman filtering),
```

Figure 5: Windows Grep Search Results Window



Windows unimaginatively called Windows Grep. It is similar to the original Unix version, but with a Windows interface. Figure 5 shows the results of a search on “evapotranspiration” in the same files searched in Figure 1 on page 33. The first thing I noticed was how long the program took to search the files, since it does not create an index in advance. The second thing I noticed was the inclusion of some legal documents in the results list that do not actually contain the search term. Clicking on a file name shows the search term and con-

The best program for you depends on how and where you store data on your computer, and which operating system you use.

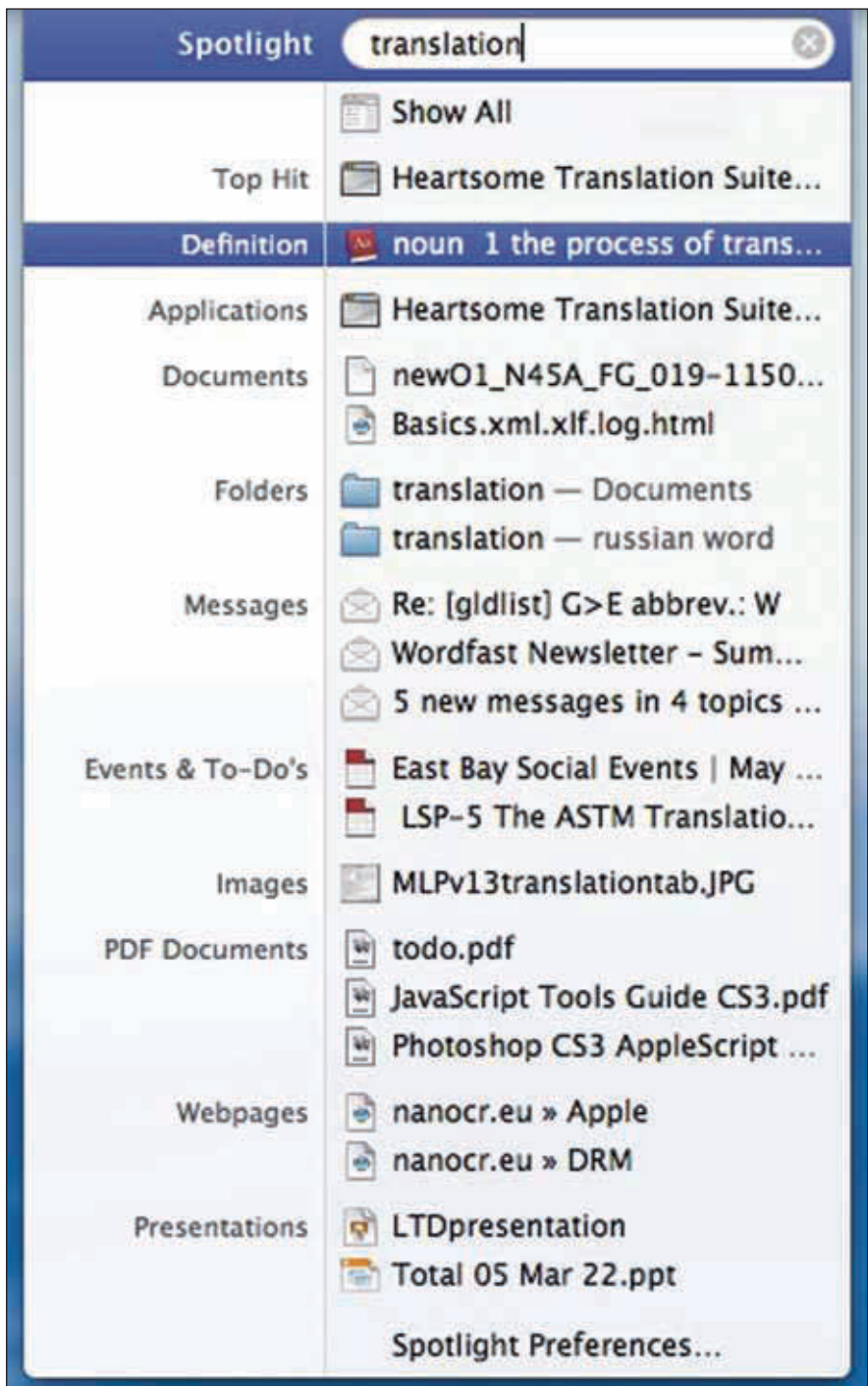
text in the lower pane. Only text and text-like files can be searched.

Spotlight

Spotlight is the built-in Macintosh

desktop search utility. Since I do not have a Macintosh, I cannot say much about it, but a screenshot provided by a colleague is shown in Figure 6 on page 36. Clicking on the file

Figure 6: Spotlight Search Results Window



name opens the original file. Unfortunately, it does not show the context of the search word, just the name of the file in which it occurs.

Windows Search 4.0

Windows Search 4.0 is made by Microsoft and can be downloaded free from the Microsoft site. Why in the world was it not built into the XP operating system!? This program or something similar to it is included in Microsoft Vista.

The program places a search field on the Windows Start bar, and the results window looks suspiciously like the Spotlight window shown in Figure 6. This window can be enlarged to show a preview pane, but the search term is not highlighted and is not necessarily visible in the pane. There are also no navigation tools to search in the preview pane, just a scrollbar.

Indexing

Some people do not like desktop search programs because they constantly index in the background and slow down your computer. Some programs only index when the computer is idle, while Archivarius allows you to schedule indexing. LogiTerm, mentioned in Part I of this article in the August issue, indexes only when you choose Index from the menu. Grep and Windows Grep do not index at all, and thus take longer to search files. As I write this, I have Archivarius, Copernic, Google Desktop, and Windows Search installed simultaneously, and I have not noticed my computer slowing down.

Parting Thoughts

The best program for you depends on how and where you store data on your computer, and which operating system you use. If you use Windows,

Summary of Programs Evaluated in Part II

Program	Cost	Platform
Archivarius 3000	EUR 29.95 (about \$47.00)	Windows
Copernic	Home Edition is free, Corporate Edition is \$59.95 (about EUR 38.00)	Windows
Google Desktop	Free	Windows, Linux, and Mac
Grep	Part of operating system	Linux
Spotlight	Part of operating system	Mac
Windows Grep	Free	Windows
Windows Search	Free	Windows

I would recommend Archivarius for desktop searching because it is the most flexible and provides the output format most similar to that of the corpus tools described in the first part of this article in August. I usually use AntConc and LogiTerm (see Part I) to search reference files in my reference file folders, but Archivarius allows me to search all files in all data folders.

References

Archivarius 3000

www.likasoft.com/index.shtml

Copernic Desktop Search 2 (Home Edition)

www.copernic.com

Google Desktop

<http://desktop.google.com/index.html>

Windows Grep 2.3

www.wingrep.com

Windows Search 4.0

www.microsoft.com/windows/products/winfamily/desktopsearch/default.mspx

ata

What's New on ATA's Website?

ATA 49th Annual Conference Program and Registration

www.atanet.org/conf/2008

ATA News and Activities

www.atanet.org/ata_activities/broadcasts.php

Business Smarts August 2008

www.atanet.org/business_practices/smarts_2008_august.php

Division News Updates

www.atanet.org/divisions/division_news.php

Translator Earnings Calculator

http://www.atanet.org/business_practices/earnings_calculator.php