PC TAP
Consumer Report

Text Retrieval Software

Open Forum
PostScript Cartridge for HP Printers
New Hewlett-Packard LaserJet III Printer

Report #7
May 1990

PC Technology Assessment Program
EPA National Data Processing Division
Information Centers Branch - RIC II, MD-35
Research Triangle Park, NC 27711
Telephone: (919) 541-0568 (FTS) 629-0568

PC TAP CONSUMER REPORTS

From the Editor's Desk

This *Consumer Report* focuses on text-retrieval, or *bibliographic* software: packages that allow you to search one or more files for records containing a specific keyword or text string. As you will see, some of these products have other capabilities as well, like file and disk management. We've had a number of requests from users for information about this family of software products, and there has been active user involvement in the testing and evaluation of the products we obtained for this project. We hope you will find something useful to you in this issue, and that the user assessments of the group of software products we examined will help you select the one that's right for your purposes.

In the *Open Forum*, beginning on page 18, is a report on the new PacificPage PostScript cartridge for the HP LaserJet II printer. The cartridge, which only recently reached the marketplace, allows the non-PostScript LaserJet II to emulate a PostScript printer, providing most of the features and capabilities of PostScript devices at a fraction of the cost of a PostScript printer.

Another article about printers follows *Open Forum* in a new section, *This & That*, in which we plan to include items of general interest or updates on topics dealt with in previous *Consumer Reports*. In this issue, it contains a report of a new HP printer, the LaserJet III. This device boasts many of the features previously available only on true PostScript devices (which the LaserJet III is not), while it lists for $300 *less* than the LaserJet II. *This & That* appears on page 19.

In the coming months, we will be looking at some of the products on the Interim PC contract, and we'll be bringing you information about them in future reports. We'll also tell you about an informal PC TAP network we're organizing, with representatives at EPA facilities around the country, to help find people with the interest and expertise to assist in future assessment projects. Thanks for your continued and growing support for the PC Technology Assessment Program.

David A. Taylor
PC TAP Coordinator

Text-Retrieval Software

Introduction

Few of us have entered the "Information Age" without encountering an information management problem. As the rate and volume of publishing increases, so too does the size of our personal text information files. What may have begun as a small collection of text files has probably grown into a file collection of epic proportions! Imagine what a task it would be to manually search those files for a specific document or keyword. In this scenario what we need is the capability to quickly search our files for relevant information and to retrieve that information as fast as we can. With all the microcomputer text-retrieval software products now on the market, this has become possible.

PC TAP recently obtained seven text-retrieval software products. These products allow you to locate information based on specified search criteria. They're useful for textual analysis and for managing personal information files. For example, scientists and researchers at the EPA, who retain large amounts of text-based information, can use a product like AskSam to organize, retrieve, and maintain their text files. In order to learn more about these time-saving products, we've conducted an evaluation of the following text-retrieval programs to see how well they meet our users' needs:

- AskSam
- Lotus Magellan
- Pro-Cite
- Reference Manager
- Sonar Professional
- WordCruncher
- ZyINDEX

Although this isn't an all-inclusive list, these products comprise a good representative sample of what's available on the subject. The text-retrieval software field is extremely dynamic; new packages appear on the market and others disappear with great frequency. Each of these programs is PC based, with the exception of Sonar Professional, a powerful Macintosh program.

The study evolved over a period of time, and as you'll see, some products were better received than others. Several study participants with a strong interest in text-retrieval software offered evaluations of additional products not included in our list: *Guide* and *Hyperties*. We've included these for comparison, and hope you'll find the additional information helpful.

During the course of our study we came across a number of different terms that essentially deal with the same category of software. Some examples: text-based management systems software, file management software, bibliographic formatters, and text filing and retrieval systems. Furthermore, text-retrieval software has been called disk hunting software, full text search and retrieval software, indexing software, search software, and search and squirt software. Don't be confused about the terms; they're often used interchangeably. "Text-retrieval" software appears to be the most appropriate term, since it covers bibliographic and full-text retrieval of stored information.

In the pages that follow we'll explore the capabilities of text-retrieval software and discuss some innovative ways these programs are being used by our participants here at the EPA. Let's first begin with a review of the general characteristics of text-retrieval software.

About Text-Retrieval Software

Text-retrieval software allows the user to retrieve text by using simple English language commands. For example, to locate all documents referring to John Doe, initiate a search query and the system will respond with all occurrences of that phrase. The user can then search the retrieved records for all correspondence to Suzy Smith between January 1984 and 1989. Keying in the specified search criteria for *date* will satisfy that information request.

Applications for text-retrieval software vary widely, as you'll see later in our report. However, there are some basic evaluation criteria we can apply over a wide range of programs. Useful features include a "no frills" text editor, Boolean search capabilities, compatibility with word processing files, a variety of input methods, sort methods, and automatic indexing. Many of the packages we examined differ in method of file preparation, indexing, query format, search result displays, user interface, and support. Some of the products have special features, like Lotus Magellan's DOS-Shell capabilities and AskSam's retrieval of graphics files.

How They Work

Index Creation. There are two types of text-retrieval software, those that create indexes and those that don't. For those text-retrieval packages that create indexes, a "concordance" is created for every *significant word* in the stored document. What is deemed *insignificant?* Generally, articles like "a" "an" or "the," in addition to other words the program has on its **"Noise Word List"** or **"Stoplist."** Because of their lack of specificity, these words can not be used as search terms, so the program ignores them or treats them as superfluous; hence, the term "noise." For example, take a look at the list for ZyINDEX; you'll see it's quite comprehensive (Table 1 p. 4). Many programs allow you to edit these lists by adding or deleting words.

Indexes serve an important purpose. They enable the program to rapidly identify text segments based on query input. The program searches the index in a matter of seconds and displays the pertinent information on screen. Most of the programs automatically create indexes, and in ZyINDEX and WordCruncher, for example, files must be indexed before they can be searched. Generally, programs that search indexes are considerably faster than programs in which every record is searched.

Downloading. Text-retrieval packages are useful for downloading citations and manipulating bibliographic information. Users of external information services, such as online databases like Dialog or BRS, may require text-retrieval programs similar to those outlined above. Pro-Cite and Reference Manager are two programs we examined that allow you to download bibliographic information, manipulate it, and format it according to pre-defined styles (i.e., MLA or APA) or to your own format style. Bibliographic references can be retrieved, sorted, and merged with other textual information. Not all text-retrieval programs we examined make provisions for bibliographic output. However, Pro-Cite and Reference Manager are flexible programs that make such provisions; they perform text manipulation and retrieval.

3

Table 1: ZyINDEX NOISE WORD LIST

a	g	n	there
about	get	never	these
after	got	not	they
all	h	now	this
also	ha	o	those
an	had	of	through
and	he	on	to
another	have	only	too
any	her	or	u
are	here	other	under
at	him	our	up
b	himself	out	v
be	his	over	very
because	how	p	w
been	i	q	was
before	if	r	way
being	in	s	we
between	into	said	well
both	it	same	were
but	j	see	what
by	just	she	when
c	k	should	where
came	l	since	which
can	like	some	while
come	m	still	who
could	made	such	with
d	many	t	x
did	me	take	would
do	might	than	y
e	more	that	you
each	most	the	your
f	much	their	z
for	must	them	
from	my	then	

Boolean Operators. Information is retrieved by using simple Boolean operators or more complex search strings. For example, the Boolean operators "And-Or-Not" work differently to retrieve information. The search request:

- "red or white" retrieves documents containing either of the search terms
- "ethics and law" retrieves documents containing both terms
- "cars not used" retrieves documents containing *cars*, but not *used cars*

Advanced boolean searches involve nesting search terms within parenthesis for more complex searches, or specifying that two words are within a certain number of words of each other. For example, in ZyINDEX this would be entered as:

- (lion or tiger) and (circus w/2 trainer)

The system is told to search for all documents that contain either the word lion or tiger **and** circus within two words of trainer. Text-retrieval programs have a variety of user interfaces to deal with Boolean operators. Some programs, like Sonar Professional, take a "dialog box" approach. All you have to do is "click" on the desired operator parameter; it's not necessary for you to remember the many complex commands and their input format.

Evaluation Method

In keeping with PC TAP practice, our users were heavily involved in this study. To advertise our project, we published a short article in the *"RIC RAGG"* (our Information Center newsletter) that identified the study and requested that those interested in participating in the software evaluation contact us. A number of interested individuals responded, and each one received an evaluation questionnaire and a software package or several packages to evaluate (depending on the participant's enthusiasm). No limit was set on how many programs each participant could evaluate. We gave each person one month to complete the evaluation.

A number of people requested the same product. For example, we found Lotus Magellan and ZyINDEX enjoying more popularity than the others. We circulated the products and tried to give everyone as much time as possible to complete their evaluation forms.

The Evaluation Questionnaire

Our evaluation questionnaire was distributed to this opportunity sample of EPA employees in order to determine the full range of software features, and to elicit their personalized responses to the software. On the evaluation form, participants had an opportunity to provide examples of how the product has been applied to their specific work environment. Thus, each participant was able, if he or she chose to do so, to take a "case-study" approach to the product and its applicability. We asked for examples of how text-retrieval software will be used in the work place, and what uses have been made of such software in the past. We're pleased that a number of respondents made additional comments on the forms.

- Following is a brief description of each product's main features. The products are described in terms of ease of use, index methods, compatibility with various software packages, search capabilities, output display, user interface, and sort features. Afterwards, we'll see how well they held up to our evaluators' expectations.

AskSam

AskSam v.4.2 is a powerful package with a number of interesting features. The program comes with excellent documentation, a tutorial, and online help. AskSam's high points include a hypertext feature that allows global scanning for searching across multiple files, and a dialog display box allowing you to construct your own programming options. This program requires extensive file preparation. AskSam directs itself to free-form entry, so it differs from other programs we examined, like ZyINDEX and Lotus Magellan, that are designed around importing pre-formatted text files. It's only directly compatible with ASCII and dBase files. Importing a WordPerfect file will work, but bear in mind that all formatting codes will be stripped. Unlike most of the programs we examined, AskSam does not create an index.

AskSam has a set of query commands for manipulating information. It supports a variety of search statements, such as Boolean and proximity searching. Its major drawback with Boolean searching is that it requires curly brace delimiter characters for nesting search statements. This is really cumbersome because the shift characters are difficult to type quickly, especially for a complex search.

It's important to understand how the program is structured. AskSam is organized around three different field types: explicit, implicit, and contextual. *Explicit* fields are similar to a database style named field, like a word (eg: "name" or "company") followed by a left bracket. *Implicit* fields are data flags such as $ in $99.00. *Contextual* fields can be searched by a specific characteristic. For example, to search a database of names with occupational listings, the occupation "librarian" can be used as a contextual field marker.

AskSam's sort feature can be used on either a simple or more complex level. For example, in a series of records consisting of names and companies, you can sort by entering either NAME or COMPANY. More advanced sorts and report formatting features require multi-parameter commands.

The on-screen output is easy to follow and quite user-friendly. After a search, what shows on screen for each retrieved record is twenty lines of text at a time. When a Boolean search is specified on a command line, the entire screen within which the search string appears is displayed. Like Magellan, related records can be linked, so you can search or print as though the related records constituted a single item.

One of the new features in AskSam version 4.2 is its ability to retrieve graphic images. Aside from Sonar Professional, none of the programs we examined had this capability. AskSam allows the user to associate any graphics file with an AskSam record, making it possible to retrieve and display scanned photos, captured video, and other digitized graphics information. Combined with its hypertext capabilities, the graphics support will let you build and design databases that combine text and graphics into your own personal retrieval system.

Lotus Magellan

Lotus Magellan is a high-powered package that will index an entire hard disk of applications and perform searches across applications. Our first use for Magellan, organizing a 30-meg hard disk, was relatively effortless and was accomplished with great speed. Magellan is easy to learn to use, especially on the basic level. Of all the packages, it has the most user friendly interface, is most intuitive, and has easy-to-follow documentation.

Magellan requires virtually no file preparation. It has a list of 120 stop words the user can edit by adding or deleting words. The terms on the stop list are non-searchable. Unlike AskSam, Magellan is compatible with a variety of packages, including database, spreadsheet, and word processing products. It's very

6

convenient to use, and the program "zooms" right in on specific text strings. Magellan is organized in an easy-to-use format. It categorizes information by file name and file content, lets you view the information within your files, and searches for text and data in multiple files across your hard disk. Remember that before you use Magellan, you must have it index the files on your hard disk. Once this is done, you can issue commands to index all new and updated files.

The program's output display is one of its best features. The screen is divided into two windows: a file list display is on the left, and a view of highlighted files within a larger widow is on the right. The file list can include all the files that meet pre-defined criteria, like all WordPerfect files or all spreadsheet files. As you scroll through the list of files, the window on the right changes to the exact view you'd have if you were actually viewing the document from within its application.

Magellan performs a variety of search tasks. The EXPLORE function is probably the first one you'll become familiar with. To make use of this function, enter a search phrase and Magellan locates all of the files that contain the search phrase and presents the filenames in ranked order in the LIST window (on the left side of the screen). You can easily generate a list of files containing a specific word. Magellan performs boolean and wild card searching. You can select groups of files by application, using criteria such as filename, file extension, or any combination of these. Once the file list is displayed it can be sorted. You can also tag files for manipulations such as erasing or copying.

Here's a short illustration of how Magellan operates. Suppose you wrote a paper on *"Information Technology Use in the National Library of Greece,"* and you want to locate the file containing the paper and append it to another document entitled *"Current Information Technology Trends: Libraries in Greece."* You could use Magellan to search for all occurrences of the words "technology" or "Greece" within all files in your word processing applications. You could even use Magellan to gather elements from these files and consolidate them into a new document. For example, you can copy text from multiple files, regardless of the applications, and combine the text into other files. In other words, Magellan gives you the power to create new documents from old ones. In addition, you can view your files in their original directory using the tree display.

One of Magellan's most useful features is its DOS shell, which can serve as a DOS front end. In some ways, Magellan resembles Automaxx on its more basic level.

Pro-Cite

Pro-Cite, by Personal Bibliographic Software, Inc., is a menu-driven package for managing personal bibliographic databases. The package consists of Pro-Cite, Pro-Search (a front-end for online searching of BRS and Dialog), and a range of Biblio-Links programs that assist with converting downloaded data into Pro-Cite compatible format. While the package is available for both IBM and Macintosh microcomputers, in our report we'll only examine the IBM version. It is important to stress that in order to use Biblio-Links your records must have tagged fields, like the tagged fields from DIALOG citations.

The program assumes bibliographic information, and it comes with twenty pre-defined data structures, or *Workforms*, covering a range of print and non-print materials. These authority lists can be called up to assist in data input. In addition, you may create up to six "User Defined Workforms" to your own specifications. This is where the program gets its flexibility. Being able to create your own workforms allows you to adapt the program to other kinds of textual information. Therefore, you can rename the existing fields to suit the type of information you have.

Pro-Cite gives you a variety of indexing options, such as word, author, title, or index. Index indexing allows you to create an index from the controlled vocabulary field, which is normally used to provide descriptors to give you additional search terms. In indexing, a term may be a single word or a phrase. Terms can be used separately or in combination to produce a dictionary list of terms. For example, author indexing would be used to produce a list with the author's name highlighted above the full citation. This list could be created to form an authority list.

Two search options are available: Search and Quick Search. The Search option allows for free text, or field-defined, full boolean searching with left or right truncation. The search process is sequential; inverted indexes are not used. Quick Search is used for searches confined to the author, title, or date field, and is quick because only the appropriate field in each record is searched. Authority lists may be called up again at the search stage, and terms (authors, publishers, journals, index terms, etc.) may be inserted directly into the Search Box from the Authority List box.

Since Pro-Cite is so well suited to managing bibliographic information, it gives you a range of output formats from which to choose. Some of these are Turabian, ANSI, MLA, APA, University of Chicago, UMI, and Science Journal. As with the Workforms, you can also create additional Punctuation Files (these are the style formats), using the Style command.

Pro-Cite comes with documentation and a tutorial. Compared to the other programs surveyed, this is the least intuitive and has the most unwieldy and cumbersome manuals. A UNIX version of Pro-Cite is scheduled for release later this year.

Reference Manager

Reference Manager has been referred to as a "bibliography builder and manager" that facilitates downloading of citation information from online services. In this way it's similar to Pro-Cite. It's a menu-driven program with pre-defined fields set up for book and journal items. The book entry has thirteen pre-defined fields and the journal entry format has ten. Included are fields for author, article, journal title, volume number, and year.

In a number of ways, Reference Manager is similar to Pro-Cite. It has provisions for downloading files from online vendors like Medline/BRS; at the menu screen, just highlight the online service you're downloading from. Like Pro-Cite, the program includes provisions for maintaining authority control over the terms used in citations. At data entry time every author, editor, keyword, and journal name that is new to the current Reference Manager file is marked with a flashing asterisk. This signals the user to consider possible corrections or to check these items against lists of existing names. For example, if the author's last name, "Jameson," is entered in the record, the asterisk will flash, prompting you to hit F3 to get an alphabetical list of names and check this entry against others.

More than 100 pre-defined journal formats are available. However, one of the strengths of this package is the ease with which new formats can be created. To create a bibliography for a particular journal, Reference Manager extracts the desired citations from the current file, formats them in accordance with the format of the selected journal, then places them in a new file in either sequential or alphabetical order.

Sonar Professional

Sonar Professional is a user-friendly Macintosh-based software product. It's a straightforward text storage and retrieval program that can search through documents created by a variety of Macintosh programs, like

8

Microsoft Word, Microsoft Writer, Microsoft Works, WriteNow, WordPerfect, and Full Write Professional. It also supports some desktop publishing packages like PageMaker and Ready-Set-Go! The program offers a number of useful retrieval capabilities, and can also accept ASCII files. In its high-speed searching mode, Sonar Professional purports to search over *10,000* pages per second!

Automatic indexing is supported. The program creates an ASCII text file containing a list of all the words and phrases to be indexed. The list includes one word or phrase per line, and you can print it out for easy reference. The index terms are in alphabetical order, and the occurrence of each term in the document is listed for added information. Sonar Professional is multi-finder friendly, so you can switch among programs while the index setup is running.

The program is effective in its user interface design. It has the standard Macintosh pull-down menus, mouse operations, and dialog boxes. The documentation is easy to follow, and after going through the demo disk you have enough of a background to begin using the program.

One of Sonar Professional's best features is its easy-to-use searching capabilities. Searching can be limited to specific documents or to groups of documents. The program supports exact word and phrase retrieval, embedded character string matching, right truncation for root word searches, and wildcard searching. "And-Or-Not" Boolean search terms can be used to create complex search specifications by linking several search terms together. Proximity commands can locate search terms within a specified number of words of one another. Sonar Professional is easy to use because you don't have to remember complex search terms. You complete the search specifications by simply filling in the blanks in the dialog box and selecting the appropriate operators. Search terms can be typed or cut and pasted from displayed documents using the Macintosh clipboard.

Sonar Professional's output display is by paragraph. It displays the paragraphs, with the desired search terms, in separate windows with a title bar that indicates their page location within the documents. Search terms are highlighted for quick identification. Each paragraph is annotated with its document name and page number. The user can then select items from the displayed paragraphs and append them to the Macintosh clipboard for later editing, printing, or other manipulation. Like AskSam, Sonar Professional retrieves graphics found in documents; graphics appear in a separate window.

WordCruncher

WordCruncher resembles AskSam in file preparation, but differs from AskSam since it's not designed to search across files. It's much better suited for searching one large file. WordCruncher is not an intuitive program, and it lacks the user-friendly atmosphere of Lotus Magellan and Sonar Professional.

The program is best suited for textual and literary analysis; originally, WordCruncher was conceived at Bringham Young University for these purposes. WordCruncher can analyze patterns buried within a textbase for comparative analysis. For example, you could find out if two papers were written by the same author by analyzing word patterns and keyword occurrences.

WordCruncher is structured differently from the other programs reviewed in our study. First of all, the program contains two modules, WC Index and WC View. WC Index indexes your files, and WC View searches them. WC Index is used to import text files and to add multi-level reference codes within the text. It recognizes three levels of reference codes that identify three levels from the standard format of an outline. You can supply the program with lists of words that shouldn't be indexed (like ZyINDEX's list of noise words). Groups of files can be indexed separately, and large files can be divided into segments and then joined into massive indexes. So WC Index really creates a "roadmap" of the file. Once the files are

indexed, they are added to the "bookshelf" and receive a .BYF extension. Think of this as an "electronic library," similar to arranging books manually on your bookshelf. Once you select a file and add it to the bookshelf, you're ready to begin searching.

WC View, the search part of the module, allows you to locate specific references, words, phrases, lists of words, substrings, and contextually defined groups of words. The .BYF files contain frequency counts of word occurrence, and the total number of words in the document. To search your document, you have two choices: you can either input the term, or scroll through the index and select a highlighted term. Or you can start typing anywhere at the index, and the cursor will be placed as close to the typed term as possible. Once you select a term, WordCruncher displays all instances of the word when it occurs, and you can then browse the hits "in context."

This "in context" display, normally shown on screen, contains several lines from a maximum of five (this size is adjustable) occurrences. To select or view any occurrence, just cursor to that section and hit return to expand the text to full screen view. You can then proceed to scroll the text just like you would in a word processor. The "hit" terms are highlighted for easy reference.

WordCruncher has two limitations. We mentioned above that it requires more file preparation than many of the other programs because of the three-tiered system of reference codes that provides for multi-level searches. Furthermore, WordCruncher only uses DOS text files and is not compatible with any of the software applications we normally use.

This is a great program for searching large files of textual information. We've got WordCruncher installed in the RIC II Information Center to run with PC SIG Library on compact disk. All you have to do is select a term or software name from the index list (or input your own term), and WordCruncher will search the entire compact disk of software program information, and then display those programs that match the search request.

ZyINDEX

ZyINDEX is an easy-to-use text retrieval package that indexes ASCII files and word processing files from popular word processors, and performs Boolean searches on those files, including proximity and nesting operations. It's a lot like Lotus Magellan, since it's easy to use and can function as a hard disk indexer. However, unlike Magellan, ZyINDEX is directly compatible only with word processors, and it supports most of the popular ones, including WordPerfect, Microsoft Word, MultiMate, and WordStar. Like Pro-Cite and Reference Manager, it makes provisions for downloaded information from online information services.

The first thing you'll notice when starting to use the package is that, like WordCruncher, it works in two separate steps: indexing and searching. In order to index the files within a directory, just enter ZYI at the DOS prompt and specify the drive, directory, and file names for indexing. ZyINDEX indexes entries automatically, and allows you to create user-editable stoplists. Additionally, groups or types of files can be excluded from indexing; all you need to do is highlight the files you want indexed. To search the indexed files, enter ZYS at the DOS prompt, and you're immediately prompted to enter the search criteria from the menu.

ZyINDEX supports exact word searches, right truncation (wildcard) for root word searches, left truncation for suffix matches, and retrieval specifications based on single and multi-character wildcard symbols. Boolean nested search (parenthetical) specifications containing up to 60 characters in length are allowed. ZyINDEX can locate adjacent words in specified sequences with the "within" operator, thereby permitting phrase searches.

The user interface is clear: some of the functions are menu driven, like display and mark; other operations are command oriented. After a search is executed, the output is displayed on screen with search terms highlighted in displayed documents. The page and line number are displayed with the text, and you can easily browse through the displayed documents by scrolling the screen.

Although ZyINDEX provides facilities for field-oriented data, it's primarily useful for documents, document abstracts, and other unstructured text segments. The program is well suited for complex applications because it places few significant limitations on document length.

Evaluation Summaries

We've just discussed seven powerful packages, and although they basically carry out the same functions, each product has its own particular strengths and "quirks." From speaking with our users, we found that even a high-powered package with one annoying characteristic could result in someone completely disregarding that package. Let's take a closer look at the products and find out how well they satisfied our users' information management needs.

AskSam

Although AskSam is a powerful and popular program, it has one of those annoying features that automatically excludes it from some users' lists of programs. One of our evaluators, an experienced PC user, refused to look at AskSam because of its troublesome curly braces delimiter characters { } used for nesting operations. Any time you make an error typing, you have to go back and enter the characters again. But don't get the wrong impression, AskSam has many strengths. Let's take a look at how two of our evaluators put the program to work.

As we mentioned in the beginning of our report, text-retrieval software has numerous applications in a data processing environment. One user, an experienced Senior Computer Operator, needed a program that would store and retrieve error codes, phone lists, and information from operations manuals. After considering WordCruncher and ZyINDEX, AskSam was selected because it required less file preparation than the other two programs and had more comprehensive documentation. It was chosen above the others for one more reason: its hypertext feature. The user wanted to customize the screen for this particular operation, and AskSam allowed him to construct his own programming options. In addition, the program supports an amazing array of search features, including boolean, proximity, truncation, nesting, numeric, and multiple searches.

Another evaluator, a systems analyst, was investigating the use of text-retrieval software for the management of metadata (data about other databases). He wanted to design and develop a text-retrieval system that would contain information about environmental databases. AskSam is a flexible package, and what the evaluator liked most was that the user can leave text unmodified and do straight searches for text, or perform searches and sorts on specific categories. AskSam can be complex, but at the same time it can be executed without delving into its complexities. Although AskSam needs less file preparation than WordCruncher, it can demand a lot of file preparation if the application requires many categories of information for purposes of sorting, retrieval, and directed searches.

Two points to remember if you're considering AskSam: (1) it has a "primitive" editor, and (2) it doesn't facilitate bibliographic output. Our evaluator combatted the editing problem by using a WordPerfect macro to insert all special characters into the text for defining all fields (categories) for AskSam input. However, those of you with bibliographic information will undoubtedly want to consider another package. AskSam makes no provision for bibliographic references and treats them as regular text.

Generally speaking, AskSam is a great package for an experienced PC user. It can appear overly complex to a novice, although with a little training you could run a "canned" AskSam application or use AskSam at the "bare bones" level. Actually, you can use it at the basic level very quickly. For an application using no fixed fields and straight text searches, you're better off going with a more straightforward software package. Keep this program in mind for fixed applications, such as online manuals, online help facilities, or online forms with text information. Our evaluator discourages those who want to use the program to manage a wide variety of textual information.

Lotus Magellan

Lotus Magellan is a powerful program that's designed for hard drive users. During installation, Magellan indexes the files on your hard drive and creates an index file. Afterwards, updating the index file is easy. However, in some cases, users may not have a hard drive and will need to use Magellan with floppies.

The difficulty in using Magellan with floppies, as our evaluator discovered, is in the indexing procedure. For example, when you use Magellan to index your floppy disk for the first time, it looks first for the index file, doesn't see it, and stops. Our user found a way around this by copying the index file onto all of the floppies. Magellan was used to copy the index file onto the floppies, then to update the floppy before initial use. Once this is done, the program is easy to use. If you are using floppies with Magellan, bear in mind that the documentation is written for hard drive users and will be of little use.

Using Magellan with floppies presents another inconvenience: slower processing time. Aside from that, our evaluator liked the fact that searches and sorts use only one keystroke. Furthermore, creating subdirectories and moving files is a logical and intuitive procedure in Magellan. Aside from the inconvenience when using the program with floppy disks, Magellan comes highly recommended. And for use with hard drives, it's an excellent program.

Pro-Cite

Pro-Cite is a powerful program, but the documentation is unbelievably complex and convoluted! Pro-Cite is great for organizing and formatting bibliographic information. Researchers, scholars, librarians, and other information professionals who need to maintain bibliographic information and produce bibliographies will find this program indispensable. In each Pro-Cite database, users can store up to 32,000 references to books, journal articles, dissertations, or any other media source.

Pro-Cite is flexible in its input methods. References can be entered into a Pro-Cite database manually, converted and transferred from online database searches using Biblio-Links, or imported from other database programs. You can then manipulate the references in a variety of ways: they can be sorted, indexed, searched, and formatted automatically into bibliographies using any punctuation style. Pro-Cite eliminates problems with creating bibliographies because it automatically creates accurate bibliographic records. The user doesn't need detailed knowledge about the majority of popular styles.

Our evaluator, an experienced PC user, previously used Ref-11, (another text-retrieval product), for a private reprint library, and has used MASS 11 for text searches. Although we found the documentation extremely cumbersome, our evaluator rated it "good" and found the package easy to install and easy to use on a basic level. However, Pro-Cite couldn't incorporate data from Ref-11, so Pro-Cite does not come recommended for those who have a system already implemented.

Reference Manager

Reference Manager, by Research Information Systems, Inc., was well received by both our evaluators. Reference Manager's "Capture Module" allows you to import citations from other systems. Although Pro-Cite and Reference Manager are similar, why would someone choose one over the other? Well, our first evaluator, who previously worked with Pro-Cite and found it a "cumbersome and difficult program to use," calls Reference Manager a "far superior" product. Keep in mind that this is an experienced PC user who's constantly looking for ways to use the PC. He reported that Reference Manager contains a number of journal formats, has a more appealing user-interface, and can be used on a basic level more quickly than Pro-Cite. This may be a result of more readable documentation.

Our second evaluator, an experienced PC user, uses text-retrieval products to prepare manuscripts of research results. This is the first text-retrieval software with bibliographic output that he's used, and consequently he really enjoyed it. However, the documentation received only a satisfactory rating. Our evaluator found the documentation lacking in the area of Data Base Utilities.

Sonar Professional

A Macintosh expert liked this product and rated the documentation "good." As with most Macintosh products, the easy installation procedure and the user interface were found to be appealing. Since our evaluator does not use text-retrieval software very often, these features were an added bonus.

The menu interface is an important criterion in text-retrieval software. We, too, were amazed at how quickly one can begin using Sonar Professional, and at how fast it indexed a document. A most impressive feature is its dialog box search strategy. Someone who has a basic knowledge of logical operators can begin constructing complex search strings in minutes. It's not even necessary to remember the parenthesis; all you have to do is click on the right item.

Sonar Professional is a fast program. Although we didn't search 10,000 pages of text to find out if it actually searched them in seconds, it appears to be one of the fastest programs on the market.

WordCruncher

WordCruncher is a difficult program to use, although it has many impressive capabilities. After examining the documentation and considering the extensive file preparation necessary to use the program, our evaluator opted for AskSam. Another reason our evaluator decided against using WordCruncher to index procedures manuals is that the reference code feature was very time consuming. These codes provide additional information about the text, such as page, chapter, and paragraph numbering. Inserting the reference codes is a tedious process, and you must remember to insert these codes in the original text before running WC Index. Our evaluator also rated the documentation "poor." Overall, WordCruncher is powerful, but difficult to learn, and this problem is compounded by lack of adequate documentation.

As we mentioned previously, WordCruncher is installed to run with PC SIG on CD-ROM. We had an evaluator use it to retrieve text on various software programs. The evaluator found the processing time exceptional, and was impressed that WordCruncher could search the entire PC SIG compact disk in a matter of seconds!

ZyINDEX

Although ZyINDEX is a powerful package, it's not as intuitive as Lotus Magellan and can not be used effectively without constantly consulting the manual. For example, in order to input search queries, very

little assistance is available to the user on screen. We had to refer to the manual for assistance in constructing appropriate search queries. This was the main feature our evaluator found fault with. In other words, the user who wants to input more complex searches needs to have a good understanding of search strategy formulation.

It took us some time to figure out how to use proximity operators to search adjacent terms. In ZyINDEX, the proximity operator "within" allows you to specify that two words occur within a certain number of words of each other. For example, in order to find *steeple* within two words of *church*, you'd enter: "steeple w/2 church" and execute the search. With ZyINDEX, it's not possible to limit a search to paragraphs or sentences. This is a significant limitation.

ZyINDEX is a useful product since it can, like Magellan, serve as a hard disk indexer, and it's certainly more user friendly than Pro-Cite.

Additional Text-Retrieval Products

Guide and Hyperties are two products one of our evaluators wishes to recommend to our readers. Hyperties, by Cognetics Corp., provides hypertext and full-text retrieval. Guide, by Owl International, is a "hypertext authoring" program that organizes and manages free-form information.

Both packages are good for online manuals. Hyperties is easier to use than Guide, but not as user friendly. The Windows atmosphere of Guide contributes to its more attractive user interface. The word search feature is easier to access than Guide's. Neither of the packages makes provisions for bibliographic references.

Hyperties differs from Guide in one major area. Hyperties accommodates multi-page articles and is easier to use when working with many articles. The program seems geared more towards multiple article searches while Guide is more geared towards searches within long documents.

What did our evaluator like least about Hyperties? First, there is no user-defined field capability, and second, no sort features are supported. With Guide, there are some limitations with word searches. Furthermore, no summary information is provided about the "hits," information many users often find helpful.

Looking Ahead

Two text-retrieval programs we evaluated in our report, Lotus Magellan and WordCruncher, have announced updates to their products. Let's take a look at some of their newest features.

Lotus Magellan

Lotus has recently announced the arrival of Magellan 2.0. The program promises many exciting changes. Version 2.0 has a "point-and-shoot" feature that lets you copy from one directory to another. You can adjust keyboard speed, turn the beep on and off, and use macros to control file-group operations. Included is an ASCII text editor, making it possible to edit ASCII text files from within the program.

We're excited with Magellan's new graphics capability. Viewing graphics files will be possible because two new viewers allow Magellan to display 1-2-3 and Symphony .PIC files, and the .DRW files of Lotus Freelance Plus and Lotus Graphwriter II. Magellan 2.0 also can restore deleted files, automatically update your index after a specified period of keyboard inactivity, and search new files without indexing them first.

WordCruncher

We're very pleased with WordCruncher 5.0. We took a look at the Demo and were quite impressed with its new features. WordCruncher 5.0 isn't an off-the-shelf product, as is 4.3; it's licensed to institutions for their in-house use or for resale purposes. Version 5.0 operates under Microsoft Windows, and it boasts enhanced capabilities in the areas of:

- search and retrieval
- multi-lingual text searches
- graphics retrieval
- hyperlink features
- program support

Searching text has become much easier. Under Microsoft Windows, you can select a single word from the text window by double clicking on the chosen word. This immediately positions you at the word in the text segment, displaying page number, paragraph number, and highlighted search terms. Search strategy formulation resembles Sonar Professional's dialog box interface. Rather than having to remember search operators, you can just click on the logical connectors within the dialog box. Version 5.0 also functions as a multi-lingual text-retrieval software package. For example, if the text you're using is in German, you can change the menu language to the language of your text.

Version 5.0 allows for graphic retrieval. WC Edit lets you import TIFF files and Microsoft metafiles. New hyperlink features from graphics to text are available. For example, if you retrieve a graphics image, say a diagram of a PS/2, you can double-click on a specific item part of the diagram and hyperlink from the graphics image to the text location. We could go from a picture of the keyboard to text describing its function.

We mentioned in our report that WordCruncher was weak in the area of documentation. Version 5.0 remedies that problem. In the new release, the addition of context sensitive help and an enhanced user interface has remedied this problem, and has resulted in a more powerful product.

Summary

As you've seen in our report, text-retrieval products are useful for a number of reasons. They allow us to manage our personal text files for fast and efficient retrieval. Rather than manually searching through printed documents, we are able to enter a word or a search string command and execute a search within seconds. More and more, academics, information professionals, and individual PC users with large amounts of textual information will find programs like these indispensable.

List of Study Contributors

Ezra Berman
Veterinary Toxicologist
HERL, MD-72
Environmental Research Center
Research Triangle Park, NC 27711
(919) 541-2784

Donald R. Block
Systems Analyst
NSI Technology Service Corporation
2 Triangle Drive
Research Triangle Park, NC 27711
(919) 541-2301

Roy Huntley
Chemical Engineer
OAQPS, MD-19
EPA Page Road Facility
Research Triangle Park, N.C. 27711
(919) 541-1060

Sophia Jeffries
IC Consultant/UNC-CH Graduate Assistant
Information Centers Branch, MD-35
EPA National Computer Center
Research Triangle Park, N.C. 27711
(919) 541-3661

Len Milich
Technical Writer
Computer Sciences Corporation, MD-43
EPA Environmental Research Center
Research Triangle Park, NC 27711
(919) 541-3500

Theresa Rhyne
Information Center Consultant
Information Centers Branch, MD-35
EPA National Computer Center
Research Triangle Park, NC 27711
(919) 541-0207

Sanjay Satsangi
Senior Computer Operator
EPA National Computer Center, MD-24
Research Triangle Park, NC 27711
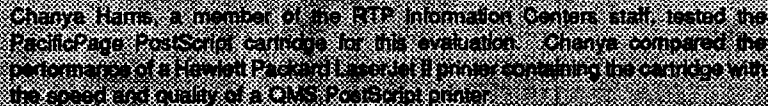(919) 541-1035

List of Study Contributors

Ralph Smialowicz
Research Microbiologist
HERL, MD-82
Environmental Research Center
Research Triangle Park, NC 27711
(919) 541-5776

Libby Smith
NDPD Telecommunications Branch, MD-34
EPA National Computer Center
Research Triangle Park, NC 27711
(919) 541-2795

Open Forum

Open Forum provides an opportunity for users to share with others their own innovations, or the results of their own technology assessments. The PC Technology Assessment Program neither verifies nor endorses the contents of Open Forum items, but we are pleased to offer them as a service to users.

PacificPage: PostScript Language Emulation Cartridge

Chanya Harris, a member of the RTP Information Centers staff, tested the PacificPage PostScript cartridge for this evaluation. Chanya compared the performance of a Hewlett Packard LaserJet II printer containing the cartridge with the speed and quality of a QMS PostScript printer.

The RIC II Information Center recently obtained a copy of Pacific Data Product's new PostScript emulation cartridge. This cartridge allows the HP LaserJet Series II to emulate a PostScript printer, making it possible to print a wide range of typefaces and point sizes. You must have a minimum of two megabytes of memory to use the cartridge. The PacificPage cartridge is compatible with all major application programs that support PostScript.

The cartridge is easy to install. Follow these five basic steps:

1. Turn off the printer
2. Insert the cartridge in the left slot of the printer
3. Turn the printer on
4. From WordPerfect 5.1 select a PostScript printer as your output device
5. Execute the print command

Preliminary tests were run using WordPerfect 5.1 documents containing large point-size selections (255 point Helvetica), and landscape orientation. The cartridge performed well, and printing speed was relatively fast. We also ran two separate tests, comparing output quality and speed between the LaserJet Series II with the PacificPage cartridge and the QMS PS810 PostScript printer. The test images appear on the next page.

In the first test, a 12k document was printed; it took 2.5 minutes with PacificPage, and one minute on the QMS PS810 (Figure on left). In our second test, we used a document containing rotated text and graphics, with a file size of 8268k. The document took 30 seconds to print on the QMS PS810, and 85 seconds with the PacificPage cartridge (Figure on right).

Although the QMS PS810 is faster, for those who don't own a PostScript printer, there are a number of reasons to consider purchasing the PacificPage cartridge. Not only are 35 different fonts available, but you can randomly select point sizes. For example, it's possible to specify a point size of 12.7. In addition, PostScript language capabilities like scaling, rotating, reverse text, screens, and patterns, are available. Furthermore, we found the print quality comparable to that of our QMS PostScript printer. It's nice to be able to purchase a cartridge that emulates a PostScript printer, especially if buying a PostScript printer is not feasible at the time. The PacificPage cartridge lists for $695 retail.

This & That

This & That is a new feature. From time to time, we will use it to bring you up to date about the latest in areas we have reported on before, or to tell you about announcements of new products or other timely items of interest.

HP Introduces LaserJet III

Hewlett Packard has recently introduced the LaserJet III, a powerful printer that combines the best features of its LaserJet II, with added PostScript printing capabilities. The LaserJet III has marked improvements in the following areas: productivity, output quality, and ease of use. And at $2395, it even costs $300 less than the LaserJet II. The LaserJet III allows you to print rotated text, shaded images, curved lines, and an assortment of scalable fonts. Hewlett Packard eventually plans to offer 300 scalable fonts.

According to Hewlett Packard, the LaserJet III's speeds are 46% faster than those of the LaserJet II for complex pages with graphics, multiple typefaces, and special effects. Here are some of its key features:

(1) Scalable fonts
(2) Prints landscape and portrait on same page
(3) Resolution enhancements
(4) More useable standard RAM

Two of the LaserJet III's most significant features are scalable fonts and resolution enhancements. Scalable fonts are vector fonts rather than bitmappped fonts. So the quality of the fonts, whether at 10,12, or 20 point, remains the same. Resolution enhancement, also known as "fuzzy dot technology," allows the printer to analyze the jagged edges of text, and to fill in the corners with small gray dots. The output quality of characters and curved lines is significantly enhanced.

Before the LaserJet III, users had to choose between a pricey PostScript printer, and an HP LaserJet (or compatible) with three or four fixed-size typefaces built in. There are a number of reasons you may consider purchasing a LaserJet III:

(1) The LaserJet II is no longer manufactured
(2) The LaserJet III costs only $300 more than the LaserJet II
(3) Pending is a release by HP of an Appletalk Interface to allow Macintosh computers to use the HP LaserJet III printer
(4) Resolution has been enhanced substantially

The LaserJet III is on order at the RTP Information Centers. We will be testing its capabilities and features, and we'll keep you informed about our findings.

How to Submit Items for Open Forum

In keeping with the PC Technology Assessment Program's objective to have the user community actively involved in TAP projects, users are encouraged to submit items for inclusion in future *PC TAP Consumer Reports*. If you have independently investigated the capabilities of a software product or a hardware component, we would like to hear from you. We'd also like you to share with others your solutions to any problems you may have encountered with a particular application or device, and about tricks, shortcuts, or unique applications you have devised. Although we can't promise to publish every contribution, we will evaluate them all in terms of their potential interest to our readers and their conformance to the spirit and intent of PC TAP.

There are no additional rules for *Open Forum* contributions, but here are some guidelines:

1. Contributions must be typed. Our first preference is that they be submitted on a floppy disk in WordPerfect format. If that isn't possible, the next best method is to EMAIL the text to PCTAP, EPA30647. The least preferable method, but still acceptable, is to mail a typewritten article to TAP at the address on the cover of this publication.

2. The length of your contribution will be determined somewhat by its complexity. However, keep in mind that we're primarily interested in the purpose of your study project and how pleased you were with the results, not in the nitty-gritty details of how you did it. We will publish your name, address, and phone number for those who want more details. Two to three pages is probably a reasonable maximum length. On the other hand, a paragraph containing a nugget that may be useful to others would be equally welcome.

3. All material submitted by users is subject to our editing, and you will not be given an opportunity to review the final manuscript before publication. Sorry, you'll just have to trust us. If we have questions or don't understand any part of your text, we'll contact you for clarification.

We hope you enjoy *PC TAP Consumer Reports*, and we look forward to hearing from individuals who have insights or discoveries to share with others. Thanks for your interest and your participation in the PC Technology Assessment Program.