A LOOK AT TEXT RETRIEVAL

By John Whelan

Are computers ready yet? How do we store documents and find them again? Traditionally, as computers get more powerful, we tend to use them to handle larger and larger tasks. Text retrieval is probably one of the largest tasks that has been tackled to date.

A paper document costs an average of 25 cents per page to store and about $2 to retrieve. These figures have been substantiated by many consulting companies. Can we reduce these costs, achieve better control (i.e., reducing misfiling, keeping the number of copies down to one), and enable users to remotely access these stored documents through the use of computers?

At first glance, the problem seems simple to solve. We can find any WordPerfect document in a subdirectory simply by using WordPerfect's word search function. However, when multiple subdirectories are involved, we must employ specialized utilities such as AskSam, Magellan, Fultext or Topic. All of these programs will handle not only WordPerfect documents but also other types of files such as ASCII text.

At the moment, the software market seems to be split between programs that handle either small or large volumes of stored computer data.

One of WordPerfect's strong points is its mass market take-it-out-of-the-box-and-use-it philosophy. Products such as Fultext that come from the mini and mainframe worlds can handle the volumes, but they do not yet provide user-friendly menus nor the quality of the documentation which we have come to expect from most computer applications.

From an organizational point of view, a large number of companies have a "central records" procedure that sends files into the depths. This is where numbers really start to get interesting. It's not unknown for the number of these paper files to be as high as 1,000,000, and I've heard estimates of an average of 100 pages per file.

We could try to scan these files and store the images through the use of data compression techniques. For example, in Group IV fax format, one scanned page adds up to about 50K, but we still cannot access its contents with the use of a search utility.

Before any search can be done, we must make an ASCII copy of the text with an optical character reader and define key words. I personally don't believe that optical character recognition is totally reliable yet.

The volume of data generated by the scanner in the above-mentioned example would take up about 5,000 gigabytes. However, if we simply store these same documents in WordPerfect format, they would occupy 200 gigabytes or about 2K per page.

Large conventional magnetic two-gigabyte disk drives, which are available from hardware companies such as Fujitsu for approximately $10,000, typically offer the Small Computer Systems Interface (SCSI) bus and 12 milliseconds average access time. The price works out to about one cent per page for media storage. Scanning is probably still above the "magic" 25 cents per page cost.

Is it practical? Well, we can attach several SCSI drives to PCs, LAN servers or mini computers. However, we will still need to scan those paper files that are not available in WordPerfect format.

We need better documented high-end programs or more powerful low-end programs, but it is beginning to look as if we really can cost-efficiently handle large computer databases consisting not only of numbers but also of text.

* * *

NEXT MEETING

The next meeting of the Ottawa PC Users' Group will be held on Wednesday, November 28, 1990. Unfortunately, the name of the guest speaker was not available when the Newsletter went to press.

* * *

FROM THE EDITOR

Hi folks! First, it appears that Competition Computers is no longer offering OPCUG members a 10% discount on store items.

Many thanks to thank Colette Lacroix of Ventura for an informative presentation on Ventura's Gold series. Peter Heney and Y.S. Yoong provided an excellent demonstration of Presentation Task Force. Congratulations to all those who won draw prizes.

I would like to take this opportunity to thank all those members who so kindly and willingly contributed to the production of the Newsletter.

The holiday season is almost upon us, and I wish all of you and your families a safe and happy holiday. Best wishes to you all for the coming year. Happy reading!

CONTENTS

Meetings are held at 100 Sussex Drive, Ottawa
Last Wednesday of the month - 8:00 p.m. to 10:00 p.m.
No meetings in July and December

REPLYING TO MESSAGES ON THE PUB

By Chris Taylor

If you wish to compose your PUB messages off-line, try the following procedure:

1) Make sure you use an editor that will produce straight ASCII text. Avoid using a word processing package unless you are certain that it can save the file as plain ASCII. Even then, you are generally better off using an editor, because it is all too easy to save text as a word processing document by accident.

2) Do not be concerned with fancy formatting. Actually, avoid any formatting other than blank lines between paragraphs. The PUB automatically reformats every line to whatever users have set their screen widths to.

The only way to prevent a line from being reformatted into the line above is to either leave a blank line between them or to leave some "white space" at the start of the line. A space or tab can do this.

3) When you choose to <W>rite a message, after you respond to the prompts for who you want it sent to and the subject, you will be prompted with "Submit Prepared Msg Text(Y/N)?". Answer Y(es). If you do not get this prompt, you can change this option in the User/Profile menu on The PUB. You will then be presented with a list of available protocols. If you are uncertain about which protocol is best to use, choose XModem. Virtually every communications program supports XModem.

4) The PUB then sits and waits for the upload to start. Follow the commands your communications software requires to upload the file containing your message.

5) After the file is transferred, your cursor will be placed on the PUB's message prompt line. Press V(iew) to see how the message will be seen by others. If everything looks satisfactory, press S(ave) to save the message.

By composing messages off-line, you will have more PUB time at your disposal.

* * *

THE DIR COMMAND

By Jean Fortier

There are many DOS shells on the market today and most of them deliver what they promise. You can bypass the basic DOS commands with such utilities, but, for whatever reasons, there will always be times when you have to resort to DOS.

DIR is one of the most commonly used internal commands of DOS, and yet a number of its capabilities elude most of us. Its main purpose is to display a list of files in a directory along with the file size, date and time of creation. It shows the volume label that you may have assigned to a disk and reports available disk space.

By the way, available disk space represents the number of bytes in unallocated disk clusters and not the difference between the formatted capacity of the disk and the number of bytes in files. Note that DIR does not list hidden files.

If you only type DIR, you are not using the command to its full power. Let's look at the syntax for this command:

DIR d:path\filename /p /w

where d: is the drive, path\ is the path name, and filename is the name of the file to be displayed.

The DIR command allows you to navigate across subdirectories. Any of these three command lines will give you a listing of the subdirectory LETTERS no matter which directory you are in on your hard disk:

DIR C:\WP\LETTERS*.*
DIR C:\WP\LETTERS
DIR \WP\LETTERS

The /P switch

The /P switch (**DIR /P**) will pause the listing after a full screen, but you can also use the MORE filter (**DIR | MORE**) and obtain the same result. However, the /P switch is more convenient than MORE because it is embedded in the DIR command and doesn't require MORE to be in the search path.

The /W switch

The optional /W switch (**DIR /W**) is quite handy in that it displays filenames across the screen in columns, thus greatly increasing the number of files appearing on the screen.

It also lists subdirectories as well as filenames, but doesn't distinguish between a subdirectory and a file, and you cannot see the size of files.

Wildcards

The DIR command contains two very useful wildcards, ? and *. A ? represents a single character (including a blank or character), whereas an * can represent up to eleven characters. **DIR *.*** is really the same as **DIR ????????.???**, **DIR**, **DIR *** or **DIR** .

DIR .. lists the files in the parent directory, **DIR ** lists the files in the root directory, and **DIR ..\..** lists the contents of the grand-parent directory.

Redirection

Redirection sends a directory listing to either a file or a printer. For example:

DIR > ALLFILES.TXT

will send a list of all the files in a directory to a file named ALLFILES.TXT. Because the > symbol will overwrite the contents of an existing ALLFILES.TXT file within the same directory, using >> will either append the list to this file (if it already exists) or create a new one. For example:

DIR >> ALLFILES.TXT

would append the listing to the previous ALLFILES.TXT. You can send a listing of a root directory to a printer by typing:

DIR D:\ > PRN

Linking DIR with other DOS commands

DIR can be used in conjunction with other DOS commands to perform special types of searches. For example:

DIR | SORT

gives an alphabetical list of files in ascending order. Add the /R switch to obtain the opposite effect. To weed out certain files or elements of information which are not needed, use FIND.

DIR < FIND "<" > LIST.TXT will screen out all names that contain this symbol, for example, the information about the directories themselves, and send the output to a file named LIST.TXT.

Useful tip

Use the DIR command to delete specific files in a directory. For example, **DIR *.BAK** will list all the files ending with the extension .bak that you plan to delete. By doing this, you can verify beforehand that you are not deleting the wrong files.

When you are satisfied that the listing is accurate, press <F1> once, type the letters EL and press <F3> <ENTER> to display the rest of the command line.

I will continue, in my next article, with other helpful DOS commands.

* * *

WORD PROCESSING DOCUMENT CONVERSION

By Gord Philips

Since there are a number of people here who are interested in word processing, I thought the following Envoy 100 message might be of general interest.

"Telecom Canada, in conjunction with the Department of Communications (DOC), Government of Canada, announces the commercial availability of a new messaging feature, Document Conversion. The Document Conversion feature was jointly developed by Telecom Canada and DOC's Government Telecommunications Agency, utilizing Canadian technology from Keyword Office Technologies Ltd. of Calgary.

The Document Conversion facility provides iNet 2000, Envoy 100 and GEMDES (Government Electronic Message and Document Exchange Service) users with the ability to exchange fully revisable documents between a variety of incompatible word processing systems preserving most of the format and special features of the document.

Approximately fifty different word processing formats are supported covering PC, MacIntosh mini and mainframe environments. The feature allows, for example, documents created in Microsoft Word on a MacIntosh computer to be converted into Wordperfect format and then be sent to an IBM computer user who can upload the document directly into Wordperfect on his/her PC. The document would retain its original format. Regular Ascii text files may also be converted to word processing formats through the use of the document conversion feature.

There are three methods of accessing the Document Conversion utility:
1. Users who wish to upload an existing document/file for conversion may enter "Compose Convert" at the "Command?" prompt. A series of prompts will then guide the user through the procedure.
2. Users who wish to send an existing workspace or a message received from another user, (with or without a binary attachment) may do so interactively by entering "Convert" at either the "Command?" or "Action?" prompt. Again, a series of prompts will guide the user through the procedure.
3. The third method of access is for more experienced messaging users. It is similar to method number 2 but allows for a one line string of commands to accomplish conversion rather than prompting the user for a series of responses.

Further information on this or any of the iNet 2000 or Envoy 100 features can be obtained from the Customer Assistance Centre, Monday to Friday 7:30 - 22:00 EDT or Saturday and Sunday from 10:00 to 18:00 EDT, at 1-800-267-8480. GEMDES users should contact their departmental Telecommunications Co-ordinator or their local Government Telecommunications Agency consultant.

* * *

THERE'S A NEW BARTENDER AT THE PUB!

By Mike Schupan

The Ottawa PC Users' Group Bulletin Board, The PUB, is getting a new SYSOP. Jean Fortier, who has been helping out with the Group's Newsletter over the past few months, will be taking over responsibilities for the bulletin board system at the end of November.

Jean has been learning the ins and outs of TBBS, The PUB's BBS software and may be implementing some new features in the next few months. One feature that's in the works is to improve the Vendor Area for vendors posting their latest sales and specials. It is hoped that it can be turned into the "Yellow Pages" for local dealers.

The PUB has evolved quite a bit over the last few years: four 2400 baud phone lines, MNP support, 600MB+ of on-line file capacity, etc. (but still no ZMODEM support though!). That trend will surely continue with Jean's leadership and support. Best wishes, Jean, as the new SYSOP!

* * *

THE TWILIGHT ZONE

The following is the new Company policy regarding computers. Please adhere to the following regulations:
- No two employees will have the same brand/model computer.
- No two computers will have the same size/capacity disk drive.
- No two offices will have the same brand/model printer.
- Hardware and software manuals must be filed away in some obscure place and forgotten.
- Copies of the above-mentioned hardware and software manuals must be illegible in those sections that the employee will need most.
- Copies of the above-mentioned hardware and software manuals must contain no indexes.
- Filenames must have no relation whatsoever to what is contained in the file.
- Always place the newest employee at the computer that contains the most important data so the data can be inadvertently and hopelessly destroyed.
- Never send employees for computer training on the software they are required to use during their workday. It is too expensive for the company, and besides, the above-mentioned copies of the appropriate software manuals are available.
- Employees are encouraged to use other employees' computers and reorganize the files as they prefer.

From The Vancouver PC Users' Group. By Doug Russell, Pinellas IBM PCUG.

* * *

NEW EXECUTIVE

The Nominating Committee is searching for people to form the 1991 Executive. Extensive computer knowledge is **not** a requirement and most Board activities require only a few hours per month. Several of the existing Board members have indicated that they will not be standing for re-election. Thus there is an urgent requirement for your services.If you are able to help guide the Group into the PC's second decade, please contact a member of the Nominating Committee either through The OPCUG's PUB or by telephone.

Stuart Moxley 832-1278

CHECKING FOR DIRECTORIES

By Robert Parkinson

. Many .COM and .EXE utilities check for the existence of a directory by relying on the fact that any subdirectory, whether with files or empty, contains the various DOS default device drivers (NUL, CON, AUX and PRN). This same action can be done in a batch file as follows. The drive is optional unless you wish to search a drive other than the current one.

```
IF EXIST <drive>\<directory>\*.* GOTO FILES
IF EXIST <drive>\<directory>\AUX GOTO DIREXIST
ECHO    The directory does not exist.
REM     Action to be taken if the directory does NOT exist.
GOTO XYZ
:FILES
ECHO    The directory exists and has files in it.
REM     Action to be taken if the directory exists and has files.
GOTO ABC
:DIREXIST
ECHO    The directory exists, but is empty.
REM     Action to be taken if the empty directory exists.
GOTO UVY
```

Line 2 above could use NUL, CON, or PRN instead of AUX, and it would still work. If both of the tests in lines 1 and 2 fail, then the directory definitely does not exist, at least not on the designated drive.

You can also use this technique from the DOS command line, where the drive and directory name are given as a named parameter (%1) in a batch file command (e.g. EXIST C:\TEMP), with the drive again being optional, in that the command EXIST \TEMP will also work. An example would be as follows:

```
@ECHO OFF
REM     This is EXIST.BAT.
CLS
IF "!%1"=="!" GOTO PARAM
IF EXIST %1\*.* GOTO FILES
IF EXIST %1\NUL GOTO DIREXIST
ECHO    Directory %1 does not exist.
REM     Action to be taken if directory does not exist go here.
GOTO END
:PARAM
ECHO    Correct syntax is "%0 <directory>",
ECHO    or "%0 <drive>:\<directory>".
GOTO END
:FILES
ECHO    Directory %1 exists and has files in it.
REM     Actions to be taken if directory
REM     exists, and has files, go here.
GOTO END
:DIREXIST
ECHO    Directory %1 exists, but is empty.
REM     Actions to be taken if empty directory exists go here.
:END
```

You can also use the DOS environment to help with both the check and any subsequent actions, as in the following EXIST1.BAT:

```
@ECHO OFF
REM     This is EXIST1.BAT.
CLS
```

```
SET DRV=C:
SET DIR=\TEST
IF EXIST %DRV%%DIR%\*.* GOTO FILES
IF EXIST %DRV%%DIR%\NUL GOTO DIREXIST
ECHO    Subdirectory %DRV%%DIR% does not exist.
REM     Actions to be taken if directory does not exist go here.
REM     For example:  MD %DRV%%DIR%
REM     XCOPY C:\FILES\*.* %DRV%%DIR%
REM     DIR %DRV%%DIR% /P
GOTO END
:FILES
ECHO    Subdirectory %DRV%%DIR% exists and has files in it.
REM     Actions to be taken go here if directory exists
REM     and has files.
REM     For example:  DIR %DRV%%DIR% /P
GOTO END
:DIREXIST
ECHO    Subdirectory %DRV%%DIR% exists
ECHO    but contains no files.
REM     Actions to be taken if empty directory exists go here.
REM     e.g., XCOPY C:\FILES\*.* %DRV%%DIR%
REM          DIR %DRV%%DIR% /P
GOTO END
:END
REM     Always best to delete unneeded environmental variables.
SET DRV=
SET DIR=
```

In the above example, the use of the drive in DRV is optional. If you delete the fourth line (SET DRV=C:), the batch file will still work, but will search only the current drive.

The actual setting of the environmental variables DRV and DIR need not be embedded in this particular batch file. It could be done in a previous batch file or from the command line.

In this next example, EXIST2.BAT, you can also include the option to check multiple drives. If you include floppy drives, make sure that they have a diskette in them, or the system is guaranteed to hang with an error message.

To do this sort of multiple checking, you must change the command line slightly, as in EXIST2 TEST C D E. In this example, EXIST2.BAT would then look like the following:

```
@ECHO OFF
REM     This is "EXIST2.BAT".
CLS
ECHO    If checking a floppy drive, is there a diskette in it?
PAUSE
SET DIR=\%1
:TOP
REM     The next line operates with SHIFT to conclude the
REM     operation when all parameters have been processed.
IF "!%2"=="!" GOTO END
SET DRV=%2:
IF EXIST %DRV%%DIR%\*.* GOTO FILES
IF EXIST %DRV%%DIR%\NUL GOTO DIREXIST
ECHO    Subdirectory %DRV%%DIR% does not exist.
REM     Actions to be taken if directory does NOT exist go here.
GOTO NEXT
:FILES
ECHO    Subdirectory %DRV%%DIR% exists
ECHO    and has files in it.
```

```
REM   Actions to be taken if directory exists,
REM   and has files, go here.
GOTO NEXT
:DIREXIST
ECHO  Subdirectory %DRV%%DIR% exists but contains
ECHO  no files.
REM   Actions to be taken if empty directory exists go here.
GOTO NEXT
:NEXT
REM   This section pauses the operation so that you can read
REM   the screen, then invokes the SHIFT command to move
REM   command line parameters one to the left, i.e., %1
REM   becomes %0, then goes back to the SET DRV=%2:.
PAUSE
CLS
SHIFT
GOTO TOP
:END
ECHO  Search is finished.
PAUSE
SET DRV=
SET DIR=
```

The technique of testing for a DOS default device driver was initially suggested in the User-to-User section of PC Magazine, Volume 8, No. 6 (March 28, 1989). Have fun!

* * *

BEGINNERS' CORNER

FILENAMES AND EXTENSIONS

By Bonnie Carter

Computer files, just like people, have first and last names. The first part of the name can be up to eight characters long and is called the **filename**. The last part is called the **extension** and can be from one to three characters long. A filename is always required, but not an extension.

A letter to your mother, for example, is a **text file** and usually ends with the extension .txt or .doc. The instructions that carry out the functions of your word processing package are called **program or command files** and end with the letters .exe or .com.

Another type of file is a **data file** (e.g., records of sales). Some common extensions for data files are .dta, .wks, and .dbf.

Two other common types of program files are **system files** and **batch files**. The former uses the extension .sys and the latter, .bat. One of the most important system files is config.sys. It sets up the way your computer interacts with its various peripherals (i.e., the keyboard, the monitor, etc.).

The best-known and most widely used file is autoexec.bat. It is both a text and program file. Its role is to automatically execute a series of instructions when you boot your computer.

Text filenames and extensions have the following conventions:
- You must put a period between a filename and its extension.
- No spaces are allowed within filenames or extensions.
- You cannot use .com or .exe since these names are reserved for program files.
- Although not recommended, you may use the following symbols in filenames: $ # @ ! ^ ( ) - { } ' ~ _ % &.
- You cannot use the following symbols in filenames: + = / [ ] " : ; , ? * \ | < > because they have a special meaning for DOS.
- You cannot use control characters, including ESC or DEL.
- Every file must have a different name within the same directory, otherwise the second will overwrite the first. However, you can have the same root name if the extension is different, and vise versa.
- You can create a filename without an extension, but you cannot have an extension without a filename.

COMMON FILE EXTENSIONS & WHAT THEY MEAN

.$xx	- A temporary or incorrectly stored file.
.ASM	- An Assembler programming language source file.
.BAK	- A backup file.
.BAS	- A BASIC programming language file.
.BAT	- A batch file.
.BIN	- A binary program file.
.C	- A C programming language source file.
.CFG	- A configuration file. Saves the details of a program's settings.
.CGM	- A special type of graphics file used by Microsoft Windows.
.COM	- A command or program file. Used when the file is less than 64 kilobytes in length.
.CPI	- Code page information file (DOS)
.DAT	- A data file (i.e., spreadsheets).
.DBF	- A dBASE database file.
.DCT	- Usually a dictionary file.
.DOC	- A document file.
.DTA	- A data file.
.EXE	- An executable program file (used when the file is over 64K).
.FNT	- A font file (produces a specific typeface for printing).
.HLP	- A help file.
.IDX	- An index file (associated with a database).
.KEY	- A keyboard macro file.
.LIB	- A program's library file.
.LST	- Listing of a program (in a file).
.MAC	- A keyboard macro file.
.MSG	- A program message file.
.NDX	- An index file (associated with a database).
.OLD	- An original version of a program which has been renamed.
.OVL	- Program overlay file
.PAS	- A Pascal programming language source file.
.PCX	- Industry standard paint program file for DOS.
.PIF	- Program Information File (TopView/Windows).
.PRN	- A list of printer files.
.SYS	- A system or device driver file.
.TIF	- Tagged-image file (scanners and desktop publishing).
.TMP	- A temporary file.
.TXT	- A text/ASCII file.
.WK1	- New format of a Lotus spreadsheet file.
.WKS	- Old extension of a Lotus spreadsheet file.
.WPG	- A WordPerfect graphic file.

* * *

BASIC STRINGS

By Harry Gross

· In this article, we will focus on string commands and their uses in BASIC. There are a variety of commands and functions, that, when used together, give a wide range of programming utilities. An individual string in interpreted BASIC may be up to 255 characters long and in Turbo BASIC, 32,767 characters, but the total length of all strings must be less than 64K, no matter how they are set up.

First, consider this triplet of functions:

LEFT$(A$,n), RIGHT$(A$,n), and MID$(A$,n,{m})

The first one, LEFT$(A$,n), returns a substring of A$, consisting of the first n letters. The second returns the end of the string. For MID$, the substring consists of m characters starting at and including character n. If m is omitted, then the substring consists of all characters from n to the end.

```
$="Ottawa PC Users' Group"
B$=LEFT$(A$,6)
C$=RIGHT$(A$,6)
D$=MID$(A$,6,11)
```

Then printing B$ gives us

B$="Ottawa": C$="Group": D$="PC Users'"

LEFT$ and MID$ can also be used to change part of a string.

```
A$="Harry Gross"
B$="Ester"
LEFT$(A$,1)=B$
```

and A$ is now "Ester Gross", my sister. For my cousin Harvey, with a longer string, we could use:

```
C$="Harvey"
A$=C$+MID$(A$,6)
```

to give us "Harvey Gross".

Another function is INSTR({n},A$,B$). This tells the program to start searching at character n in A$, for a substring B$. If a match is returned, the function returns the position of the start of the match. If no match is found, 0 is returned.

Consider the string "Jack,Jill,Jim,Henry,James". We have a list which we would like to parse out into separate items. So:

```
DIM List$[100]              ' allow up 100 items
A$="Jack,Jill,Jim,Henry,James"  ' this is the input list
cm=1                        ' start at character 1
k=0                         ' k counts the list
StartLoop:                  '
  c=INSTR(cm,A$,",")        ' where is the comma?
  IF c >0 THEN              ' found one
    k=k+1                   ' bump k by one.
    List$[k]=LEFT$(A$,c-1)  ' extract it
    A$=MID$(A$,c+1)         ' shorten the list
    cm=c+1                  ' next place to look
    GOTO StartLoop          ' look again
  END IF                    '
k=k+1: List$[k]=A$          ' last item in the list, no commas.
```

Another example, where we only want a yes or no:

```
A$="Jack,Jill,Jim,Henry,James"  ' this is a class list.
INPUT "Pupil's name",P$
IF INSTR(A$,P$) THEN
  PRINT P$ "was in class today."
  ELSE
    PRINT P$ "played hooky today."
```

END IF

If we want a position in a list, the following is useful.

```
Mon$="Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec"
INPUT "Month please",m$
m$=LEFT$(m$,3)          ' only three letters
j=INSTR(Mon$,m$)        ' where is it?
m=(j+3)/4               ' change to a number
```

The previous example shows how to find where in a list an item is, and the prior one, how to break down a list into its parts.

LCASE$ and UCASE$ are next. The former changes a string into lower-case letters, and the latter to uppercase. We can use these to allow for some leeway in the input. It is good practice, if no ambiguities arise, to let the program look after some details.

```
Mon$="Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec"
Start:
    INPUT "Month please",m$
    m$=UCASE$(LEFT$(m$,1))+LCASE$(MID$(m$,2,2)
    j=INSTR(Mon$,m$)    ' where is it?
    m=(j+3)/4           ' change to a number
    IF j=0 THEN PRINT "Error in input.": GOTO Start
```

If you noticed, we have combined two functions into one, the LCASE$ and MID$. If it does not affect the readability of the program, do it, as it will speed up the program's execution.

The next function, while not a string operation, is closely related -- the LEN(A$). This returns the length of the string. Again using it in the above example.

```
Mon$="Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec"
Start:
    INPUT "Month please",m$
    IF LEN(m$)<3 THEN GOTO Start  ' need 3 letters to be sure.
    m$=UCASE$(LEFT$(m$,1))+LCASE$(MID$(m$,2,2)
    j=INSTR(Mon$,m$)    ' where is it?
    m=(j+3)/4           ' change to a number
    IF j=0 THEN PRINT "Error in input.": GOTO Start
```

We now come to the function STRING$(n,A$) or STRING$(n,a). For the first one, we get as string, n characters long of A$. The second one is the same, except a is an ASCII value.

```
a$="-": a=45
b$=STRING$(60,A$)
c$=STRING$(60,45)
```

will both give us the same thing, a row of dashes.Not very exciting, but look in your ASCII table and notice the characters with values 128 and above. These are not available on a keyboard. Try the following on your machine.

```
A$=STRING$(10,32)+STRING$(1,209)+STRING$(58,205)
    +STRING$(1,187)
B$=STRING$(10,32)+STRING$(1,186)+STRING$(58,32)
    +STRING$(1,186)
C$=STRING$(10,32)+STRING$(1,200)+STRING$(58,205)
    +STRING$(1,188)
LOCATE 10,1: PRINT A$;
FOR N=11 TO 20: PRINT B$;: NEXT N
LOCATE 21,1: PRINT C$
```

If my research and your typing were correct, then there should be a box on your screen. The problem lies in printing this, as not all printers use the same output character for a given input ASCII value. The next article will carry on with more string examples.

* * *

PUB TALK

Ham Club Meetings
From: Brice Wightman

Correction to October Newsletter, PUB Talk: The Ottawa Amateur Radio Club has changed its meeting time to 7:30 pm (That's at the NRC, 100 Sussex Drive, first Wednesday of the month).

The Ottawa Valley Mobile Radio Club meeting time (not given) is 7:30 pm (That's at the Multi-Service Centre, 1480 Heron Road, third Thursday of the month.)

The OVMRC also sponsors a course to prepare for the Amateur Radio Operators licence annually. This is cheaper and better than the Algonquin course. Unfortunately, it is now into its seventh week. Try next year right after Labour Day.

Scanners
From: Harald Freise

Does anyone have experience with flatbed scanners for OCR and images? I would appreciate your views and comments on specific hardware and software for the above. What were the strong points and whatwere the pitfalls? Is there anything you would have done differently with 20/20 hindsight?

From: Andre Lefrancois

I recently had two OCR systems demonstrated for me. The first was from XEROX, the second was Omnipage. Both systems were very accurate. The XEROX system had a so-so end user interface, but created great WP50 files. Omnipage had a great user interface, but the WP50 files had hard returns at the end of every line. I believe Northern Micro may be able to provide a full demo of the Omnipage system. The XEROX system lets you scan several pages at once into TIF files, which can be batch processed into WP50 files at a later time. This is an important time saver if you are going to be scanning several documents each day.

From: John Whelan

Personal Publishing did a comparative survey of software for Hewlett Packard ScanJet Plus and recommended Omnipage. Xerox brought up Kerzweil which is very much the Cadillac of the scanners, but Omnipage on the ScanJet is reputed to be close. The thing to watch is the software to convert the image to characters. Omnipage nlly needs a 386 with 4Mb of RAM and it tends to eat the dollars. Traditionally OCR has been very fussy about the quality of the input. Now I understand that at least magazines can be scanned, there was a time when only IBM Selectric typewriter input could be scanned and that only in some typefaces with a new ribbon. Photocopies do not scan well.

One thing to watch out for is the error rate. Even an apparently small error rate, say 5%, means a lot of proofreading and correction work. Typically it can take as long to retype. You really want 99%+ accuracy. Beware also when scanning in French accented characters. If you're very lucky, you may get the lower case to scan. Dot matrix output is more difficult than the laser or daisy wheel. Photocopies of dot matrix are considerably worse than anything else. Photocopies of photocopies are getting really dodgy. A good spelling checker, especially one that has been "tuned" to eliminate improbable words, can help.

You may need additional hard disk drive space to store images. A single page can take up 3 to 4 megabytes of hard disk space, although data compression will bring that figure under control. Also, images can take a while to print on Postscript printers. Ones that normally print eight pages per minute can typically take two hours to print a single page. The theoretical maximum is 40 hours for a single page. The thing to watch out for here is Postscript typesetters. It takes much longer to print the image at 2,400 dpi than 300 dpi, and if you are paying by the minute beware that one page of a company's newsletter with lots of graphics in it took many many hours to print.

Check the throughput of your job. If you need to scan in fifty pages, check if there are hoppers available. Also, if you can, convert the whole fifty pages to a single file in one shot unattended once you have loaded the paper bin. Do you need information such as underlining, bolding, tabs, indents? You'll need to do quite a lot of formatting work to convert the scanned image into a word processing WordPerfect file with the correct typefaces, (small medium sizes etc.) justified right, left, margin set, etc.

Scanners can be a very good way to go here, and I have the odd blind programmer contact who thinks scanners are the only way to handle written documentation. However, he really would prefer that we all used E-Mail more. I have more questions than answers I'm afraid, but yes it can be done, and its a lot better today than three years ago.

Local Writer Strikes Out
From: Michael Goddard

Warning.....I will use the V*** word below!!! A local author with an interest in computer plagues like viruses has taken it upon himself to educate us all.

He has created a program originally kept in a file called VIRDEM.LZH. This has been uploaded to some BBS in the Ottawa area. Local sysops were just warned about it and asked to sanitize their file lists. The program seeks out .COM and .EXE files and corrupts them. The files are not destroyed, you just get a message on your screen. (It does not seem to be a particularly virulent beast but....). It is NOT detected by McAfee's tools at present. Cautious users can look for the thing by using Norton utility type text search and scan all .EXE and .COM files for the characters 595-0787. This is part of the message that appears on the screen. Sorry about the warning after the latest discussion about viruses, but I wantedthe PCUG users to know fast.

ProComm
From: Dana Weber

Is it possible to redefine the keys in ProComm? And if so, then how? This is annoying several people I know.
From: Gord Phillips

I took a quick look through the index of the ProComm Reference Manual (1504 pages long!), but I did not find anything on redefining keys. There IS a facility (which I have not used) to create your own keyboard macros. You can assign a character string of up to 50 characters to a key (such as ALT-0 through ALT-9). The string can contain embedded control codes and carriage returns and use the translate conventions. To send the string you've assigned, simply press the appropriate key. These macros can also be used in command files. To set up the keyboard macros, press ALT-M from Terminal Mode.

This is all documented in the ProComm 2.4 Reference Manual. Look in chapter 4 (Major Functions - Keyboard Macros, on page 401) for the correct way to setup and use the macros. Chapter 7 (Command Files - Macro Command, on page 714) explains how to use these various macros in command files.

* * *

OTTAWA PC USERS' GROUP

EXECUTIVE

Chairman	David Terroux	238-4895
Past Chairman	Stu Moxley	832-1278 or 832-4378
Treasurer	Tony Frith	671-0401
Secretary	Claude Jarry	521-3366
Newsletter Editor	Bonnie Carter	236-1015
Software Librarian	Chris Taylor	723-1329
BBS System Operator	Mike Schupan	820-0293
Convener	Douglas Poulter	745-8768
Membership Chairman	Carl-Henri Gomez	731-1462
Hardware/Software Broker	Terry Mahoney	225-2630
		226-2615 FAX

ASSISTANTS

Assistant Newsletter Editor	Jean Fortier	236-1015
Newsletter Assistants	Marc Riou	733-2092
	Chris Taylor	723-1329
	Ann Falkner	722-3226
	David Smith	837-9291
Software Assistant	John Ings	235-8132
Membership Assistant	Marc Riou	733-2092

BEGINNERS' SESSIONS	Eric Clyde	749-2387

DISK OF THE MONTH

One-year subscription (10 Disks) 5.25"	$25.00
One-year subscription (10 Disks) 3.5"	$35.00
Individual 3.5" disk surcharge	$2.00

THE OTTAWA PC USERS' GROUP MEMBERSHIP APPLICATION - Please Print

Last Name: _____ First Name: _____

Mailing Address: _____

Postal Code: _____ Telephone - Home: _____ Office: _____

Profession: _____ Date of Birth: _____ Sex: M _____ F _____

Disk of the Month: YES _____ NO _____ Size: 5 1/4" _____ 3 1/2" _____ Amount Enclosed $ _____

Are you: A new member? _____ Renewing your membership? _____ Sponsor's Name: _____

What in particular interests you in the Group? _____

Can you help in Group activities? Check off the activities that apply: Programming Language Instruction _____

Newsletter Input _____ Memberships _____ Software Library _____ Promotion/Publicity _____

Hardware Techniques _____ Meeting Locations _____ Agendas & Speakers _____ Advertising _____

Bulletin Board _____ Other _____

What hardware/software do you own and/or use? _____