DupeFind v1.0 Users Manual

The Worlds Fastest Duplicate File Remover

Version 1.0

Dupefind v1.0
Users Manual

Major Features:

DupeFind is the worlds' fastest duplicate file finder for OS/2. It is a 32 bit native OS/2 application that can ferret out duplicate files on your hard disk or network, optionally show them to you, and then assist you in deleting them.

DupeFind prioritizes your duplicate files in order of file size, so that you can concentrate on the "big ones" first. This will help you free up disk space in the shortest time possible.

Did you ever unzip a ZIP file, and keep the original Zip laying around? DupeFind can look inside ZIP files and tell you if any of the Zip members are unzipped anywhere on your disk or network! DupeFind can find individual duplicate files in multiple Zips. DupeFind will even let you surgically delete individual files inside a Zip!

If you are not sure if you want to delete a file that is duplicated, DupeFind will help show it to you with its Smart Viewer feature. DupeFind can be configured to pop up your favorite graphics viewer on graphics files, your word processor for duped DOC files. It will even pop up your MIDI or WAV player and sing your duped music or sound files to you. Yes, DupeFind can even whistle Dixie [if you happen to have two identical DIXIE.MID files] You can use DupeFind's Smart Viewer with any OS/2 application that accepts file names on the command line.

DupeFind uses sophisticated signature analysis of the files on your computer or network to make comparisons. It can find duplicate files even if you rename them. If you have a large hard disk or network, try our database feature that takes a snapshot of all your file signatures. Next time you run DupeFind, use the database to vastly speed up scan times.

DupeFind can do "fuzzy" comparison of GIF files. This allows you to find and delete GIFs with similar contents. Use the "Graphics Smart Viewer" feature to view the GIFs prior to deletion.

DupeFind handles FAT, HPFS and NTFS via Peer Network. Long file names? No problem!

DupeFind is small - About 200k executable size! Compare our feature set with the competition that weighs in at 1.9 megabytes for slower and poorer performance. We are trying to save disk space, not use it up with bloatware!

DupeFind is a VIO application, compatible with "Presentation Manager". That means you can run it on your PM desktop like any other PM application, or run it as a full screen OS/2 application without PM. You get the best of both worlds!

DupeFind is compatible with VoiceType Navigation. Use your voice to run many DupeFind features!

Any missing features you can't live without? Please write or call the support line. Custom versions are available.
Benchmarks:

All products tested on:

>   Warp 4.0, FP5 applied
>   200 MHz Pentium (std)
>   64 Mb EDO RAM
>   256k Pipeline Cache
>   Stealth64 Video (4mb)
>   (2) 1-Gb Fujitsu Fast-Wide SCSI drives
>   Adaptec 2940UW SCSI adapter
>   C: = 148 Mb FAT partition (Physical Drive #1)
>   D: = 485 Mb HPFS partition (Physical Drive #1)
>   2254 Directories total
>   10227 Files total
>   2063 duplicate files

All products were configured to look in Zips and to compare files by signature.

Product	Configuration	Time to display Dupes (Min:Sec)	Exe Size
DupeFind v1.0	Read existing database (R/O)	1:23	200k
DupeFind v1.0	No database (No DB)	2:41	200k
Clone Cleaner v2.20	CRC Mode	4:48	1.9 Mb
Dups 3/26/96	Content compare High priority	14:42	190k
ISAD/SAD v1.1		N/A	4.3 Mb EXE+ 2.7 Mb DLLs

Dups cannot search inside Zips or compare GIFs by palette.

ISAD/SAD failed to search all sub directories. It apparently has a problem with recursion or paths/filenames with spaces in them - it found only 6910 of 10227 files during its search. It has no capability to look inside Zip files.

Required Equipment and Software:

       OS/2 Warp v3.0 or greater
       16 Mb RAM, (minimum)
       20-50 Megabytes of free space on the Swap partition (depending on # of files to scan)
       MMOS2 and Presentation Manager optional for Smart View feature
       OS/2 compatible network - optional

       Note: A special version of DupeFind is available for versions of OS/2 prior to Warp 3.0. It is untested.

Installation:

1. Unzip the installation ZIP into an empty subdirectory.
2. Run INSTALL.CMD from the OS/2 command line or by clicking on its icon in your new directory folder.

INSTALL.CMD will create a folder on your desktop with the DupeFind icon within. You can move the folder or the icon to someplace more appropriate, if you like.

After installation is complete, you may wish to install VoiceType Navigation support if you are running Warp 4.0. See the VoiceType section below.

Quick Start

After initial installation, DupeFind is initialized with reasonable default values. For a quick tour, do the following 4 easy steps to locate all the duplicate files on your drive or network:

1.   Start DupeFind
2.   Select File|Setup
                Pick some drive letters to scan
                Select OK on the Setup dialog
3.   Select Run from the main menu.
                DupeFind will scan the drives you selected and create a list of duplicate files.
4.   Press Esc to quit from any screen.

We strongly suggest that you DO NOT attempt to erase any duplicate files until you have read this manual.

See the Advanced Startup section for more details. The Advanced Setup can improve DupeFinds speed by a factor of 2!

Support

Support for DupeFind is available via the following methods:

E-Mail:	dupefind@idk-inc.com
	CIS 72470,1233
Voice:	(805) 492-9190 (M-F, 0800-1800 Pacific time)
Fax:	(805) 492-1533
Snail Mail:	IDK, Inc.
	2217 Aspenpark Ct.
	Thousand Oaks, Calif. 91362-1731
	USA.

If you are having problems running DupeFind, please send us the DUPEFIND.ERR and DUPEFIND.LOG via e-mail. Both of these can be found in your DupeFind installation directory.

Caveats and Warnings:

DupeFind, like any sharp power tool can do a lot of beautiful work - or cut a wide swath of death and destruction. It is entirely possible for you to delete files that are critical to your workstation, your server or peer workstations. This program can get you into a lot of trouble, FAST! Carefully review each duplicate filename that DupeFind locates. There are many legitimate reasons for having duplicate files on a hard disk or network, so do <u>not</u> assume that just because a file ends up in the duplicate list, that its fair game for deletion.

We strongly advise you to make a full backup of any drive on which you plan to run DupeFind. DupeFind has no "undelete" feature. IDK, Inc. is not responsible for misuse of this product!

License Information

  DupeFind is shareware. This means:

  1. All copyrights to DupeFind are exclusively owned by the author - IDK, Inc.

  2. Anyone may use this software for a test period of 40 days. Following this test period of 40 days or less, if you wish to continue to use DupeFind, you MUST register.

  3. Once registered, the user is granted a non-exclusive license to use DupeFind on one computer (i.e. a single CPU), for any legal purpose, at a time. The registered DupeFind software may not be rented or leased, but may be permanently transferred, if the person receiving it agrees to terms of this license. If the software is an update, the transfer must include the update and all previous versions.

  4. The DupeFind unregistered shareware version may be freely distributed, provided the distribution package is not modified. No person or company may charge a fee for the distribution of DupeFind without written permission from the copyright holder.

  5. To register you must complete the registration form and send it, with registration fee, to one of the authorized registration sites.

  6. DUPEFIND IS DISTRIBUTED "AS IS". NO WARRANTY OF ANY KIND IS EXPRESSED OR IMPLIED. YOU USE IT AT YOUR OWN RISK. THE AUTHOR WILL NOT BE LIABLE FOR DATA LOSS, DAMAGES, LOSS OF PROFITS OR ANY OTHER KIND OF LOSS WHILE USING OR MISUSING THIS SOFTWARE.

  7. There are NO additional license fees, apart from the cost of registration.

  8. You may not use, copy, emulate, clone, rent, lease, sell, modify, decompile, disassemble, otherwise reverse engineer, or transfer the licensed program, or any subset of the licensed program, except as provided for in this agreement. Any such unauthorized use shall result in

immediate and automatic termination of this license and may result in criminal and/or civil prosecution. All rights not expressly granted here are reserved by IDK, Inc.

   9. Installing and using DupeFind signifies acceptance of these terms and conditions of the license.

   10. If you do not agree with the terms of this license you must remove DupeFind files from your storage devices and cease to use the product.

   11. Unregistered copies of DupeFind are subject to a limitation of a maximum of 10 file deletions in a single session. See the Registration instructions for more information on this limitation.

Legal Stuff

All products and company names mentioned in this manual are trademarks or registered trademarks of their respective companies.
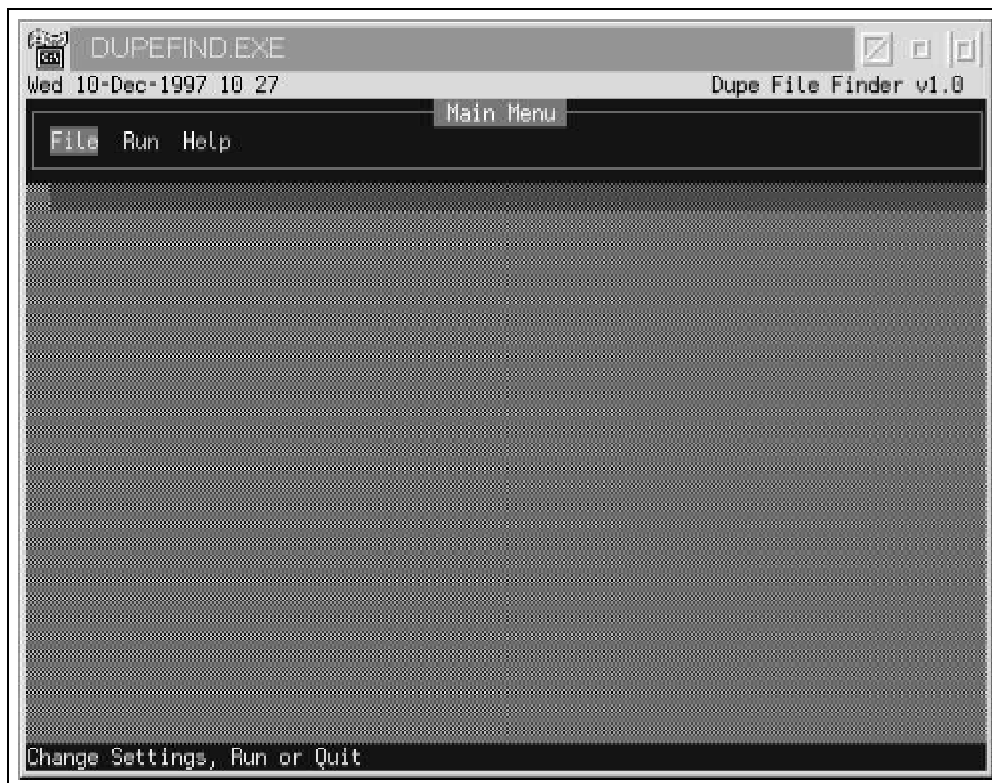
DupeFind Setup:

Advanced Startup

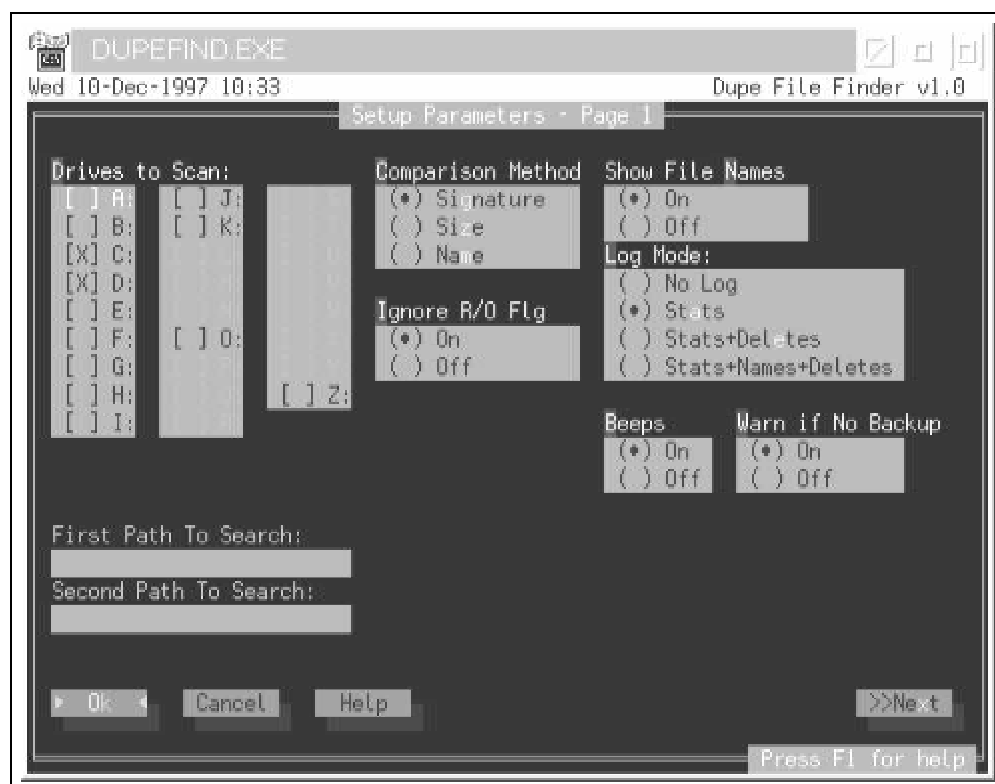See Quick Start for information on how to do a quick tour of DupeFind.

You should not run DupeFind without setting it up first. Here's how:

Start DupeFind by either clicking on the icon, or from an OS/2 prompt, CD to the directory containing DUPEFIND.EXE and type "DUPEFIND" at the OS/2 prompt.

DupeFind will display a menu:

Click on File|Settings to open the Settings dialog. We need to set up your defaults before running the program.

```
┌──────────────────────────────────────────────────────────────────┐
│ ▓▓  DUPEFIND.EXE                                    ☑  □  □│      │
│ Wed 10-Dec-1997 10:33                        Dupe File Finder v1.0 │
│                ╒═══════ Setup Parameters - Page 1 ═══════╕         │
│                                                                    │
│   Drives to Scan:     Comparison Method   Show File Names          │
│   [ ] A:   [ ] J:     (•) Signature       (•) On                   │
│   [ ] B:   [ ] K:     ( ) Size            ( ) Off                   │
│   [X] C:               ( ) Name           Log Mode:                │
│   [X] D:                                  ( ) No Log               │
│   [ ] E:                                  (•) Stats                │
│   [ ] F:   [ ] O:     Ignore R/O Flg      ( ) Stats+Deletes        │
│   [ ] G:               (•) On             ( ) Stats+Names+Deletes  │
│   [ ] H:       [ ] Z:  ( ) Off                                     │
│   [ ] I:                                  Beeps    Warn if No Backup│
│                                           (•) On    (•) On         │
│                                           ( ) Off   ( ) Off        │
│                                                                    │
│   First Path To Search:                                            │
│   [                    ]                                           │
│   Second Path To Search:                                           │
│   [                    ]                                           │
│                                                                    │
│   ► Ok ◄    Cancel    Help                          >>Next         │
│                                                                    │
│                                             Press F1 for help      │
└──────────────────────────────────────────────────────────────────┘
```

Above is a fairly straightforward setup. Notice that the available drive letters in the *"Drives To Scan"* cluster are in Black on Cyan. The drive letters that are not available on this workstation are "grayed out".   On this workstation, A: through F: are local hard disk drives, G: is an external Jaz drive; H: is a CDROM; I:, J:, and O: are Novell Network drives; K: is a Peer network drive on a nearby workstation. Z: is a CryptStream file system on the Jaz.

The above example shows that we have selected C: and D: for duplicate file scanning. The other drives will be ignored. Click on the desired drive letters to toggle them on and off. Alternatively, you can tab to the desired cluster and use the arrow keys to move around. Tap the spacebar to select/deselect drive letters.

The *"Comparison Method* cluster selects the way we will compare files to see if they are duplicates. DupeFind Version 1.0 supports Signature comparison only, so make sure it is selected at all times.

"Show File Names" selects whether DupeFind will display the name of the file being Scanned or Processed. Scanning simply locates files, Processing occurs when we calculate file comparison signatures. Showing file names while scanning slows DupeFind slightly, so you may want to turn it off once you are used to DupeFinds pace.

"Log Mode" selects the verbosity level of DUPEFIND.LOG. We can elect to not log at all (No Log), Log statistics only (Stats), Log Stats and the names of deleted files (Stats+Deletes) or the

whole works: (Stats+Names+Deletes). This logs Stats, Names of all found files (dupes or not) and the names of all the files you end up deleting. The Stats+Names+Deletes option can create quite a large log file if you scan a large network. It also can slow down DupeFind by several percent. Stats only or Stats+Deletes is probably your best bet.

"Warn if R/O File" controls DupeFind's warning message if you attempt to delete a Read-Only file. If you select ON, DupeFind will cheerfully delete a file marked Read Only without warning you. This can get you into trouble if you delete a critical system file or something you actually wanted to protect. If you select OFF, DupeFind will warn you if you attempt to delete a Read Only file. It will still let you delete the file though, so watch out!

"Beeps" controls whether DupeFind will use your PC speaker to make various warning sounds during operation. OFF makes DupeFind go mostly quiet. Beeps also controls the multimedia "fanfare" in the About box. This field does not control the playing of multimedia files using the Smart Viewer.

"Warn if no Backup" controls the warning message that pops up if you attempt to delete a file that is marked as "not backed up". Most backup programs flag each file with an "I'm backed up" indicator. If you modify a file after backing it up, the indicator is reset to "I'm NOT backed up". If you never back up, or infrequently do so (shame on you!) the DupeFind warning message will get pretty annoying after the first 10 or 20 times. Shut it off here, by selecting OFF.

"First Path to Search"
"Second Path to Search"
Both fields are shown as grayed out in the screen shot, since we have selected the C: and D: drives. If we use the Drive Selector, we start scanning for duplicate files in the "root" of each drive - in this case we will start on C:\ and sweep through all the sub directories on C:. Next we will automatically switch to the D:\ drive and scan all the sub directories there.

So what if you want to just scan a subdirectory tree or two?

First, shut off <u>ALL</u> the drive letters in the Drives to Scan box, and then click on First Path to Search. Enter a fully qualified path, like "D:\DOCUMENTS". If you want to compare this subdirectory tree with another, simply click on Second Path to Search and type in another fully qualified path like "C:\MEMOS". This will cause DupeFind to first search/catalog the DOCUMENTS directory and then look in the MEMOS directory. Note that the search order is not important. If you fill in one or both of these fields, DupeFind will recurse into any sub directories below them.

If you want to return to searching entire drives, just click on one or more of the drive letters. You First and Second paths will gray out, and DupeFind will go back into full drive scanning mode.

Buttons
*OK*
Click OK to return to the main menu.

Cancel
Click Cancel to return to the main menu. Note: DupeFind uses "perfect save" similar to OS/2 Properties dialog boxes. If you change something, you can't back out of changes by canceling.

Help
Click Help (or press F1) to view help for this dialog. If you press F1, help for the currently highlighted dialog field will appear. Pressing F1 while the help screen is popped up displays the help table of contents. Alt-F1 jumps to the previously viewed help topic. Another trick: Press BOTH Left and Right mouse buttons on the area of the dialog box you want help on - help for that field will appear!

>>Next
Click Next to move to the second Setup screen.
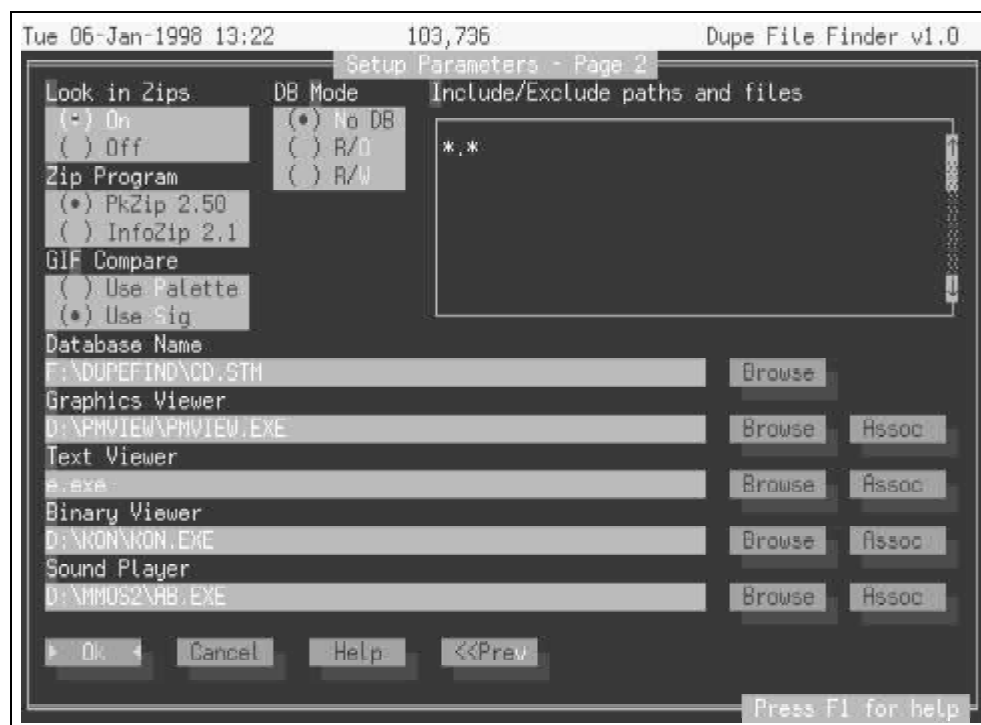
The picture above shows a standard setup page 2. Lets go through each field and its use.

"Look In Zips"
On = open each Zip we find and add each Zip member to the catalog of files. This allows us to find other copies of a Zip member either uncompressed or inside another Zip. Off= Only calculate the signature of the entire zip - do not consider the individual zip members.

"Zip Program"
Selects the Zip program you want to use. DupeFind currently supports InfoZip v2.1 and Phil Katz's PKZip v2.50 for OS/2. You must use a native OS/2 Zip/Unziper. PKZip for DOS or Windows is not suitable. InfoZip is freeware and widely available on the Internet. If you opt for the InfoZip solution, you need both ZIP.EXE and UNZIP.EXE.

PKZip 2.50 is shareware (I think) and fairly new, so availability is somewhat limited. The IDK copy was downloaded from and paid for via Phil's PkWare web page.

"*DB Mode*"
(*D*ata*b*ase Mode) Selects how you will use a file catalog database.

"No DB" (No Database) is good for small systems where the time savings of using a database is negligible. If No DB is selected, DupeFind does a smart scan of your drives, finds the dupes and shows them to you.

"R/W" (Read/Write the Database) is used to create or update a database with all the files you find in a single scan. So, if you scan C: and D: and R/W mode is on, DupeFind will place ALL the

names and signatures of every file found on C: and D: into the database. Later, you can use this database to avoid re-scanning files that have already been analyzed. This can vastly speed up the scanning process on large networks. The time to create this file the <u>first time</u> is lengthly, but if you frequently scan your large drive or network, it can really pay off in time savings.

"R/O" (Read Only Database) is good for large drives or networks AFTER you have created a database using R/W mode (see above). R/O mode is good for re-scanning previously scanned drives, since it is the fastest possible mode for DupeFind.

NOTE: if you have a database that contains all the files from C:, D:, E: and F:, and then re-scan only C: with R/W mode turned on, you will effectively erase the database entries for D:, E: and F:. This does not erase the actual files, just the catalog entries. How to prevent this? Create database files with meaningful names like CDEF.STM, C.STM , CD.STM and so forth. Alternatively you can use R/O mode which Reads the database but does not update it. This allows you to create a master database for your entire network and then surgically scan only a few drives here and there without deleting catalog entries.

So, to sum up, R/W mode forces DupeFind into scanning EVERY file on each selected drive to create a signature/catalog entry in the database. This can be very time consuming on a busy network or if there are many thousands of files. R/W mode can also create a fairly large database file. Make sure you specify a path for the database file where there is plenty of disk space. Specifying a fast drive for the database file location can help speed things up too.

Suggestion: Depending on how often your files are modified, use R/W mode periodically and scan all your drives. In between times, use R/O mode on the big database to look for dupes. Most systems will only need a full R/W scan every month or two for best performance. DupeFind "knows" if you modify a file, and will reanalyze files that have an out-of-date signature in the database.

R/W mode skips over recalculating signatures on files that have not changed.

If you always scan ALL the visible drives on your system, then go ahead and use R/W mode for all your DupeFind runs.

"Include/Exclude paths and files"

This field contains a list of wild cards used to accept or reject files for comparison. You can enter up to 32 Include/Exclude wild cards. If a given file matches one of your wild cards, it will be added to the catalog for comparison with other files. If a file does not match a wild card in this list or is "excluded" using a "~" wild card, it will be ignored and <u>not</u> compared against the other files. Most users will use a single wild card set to "*.*". This will cause DupeFind to accept (include) all files for comparison / duplication matching, excluding none.

You may want to use the Exclude feature to lock out files that you really don't care about. A prime example is your Netscape CACHE directory. This directory contains hundreds of GIF and

JPG files used to draw the buttons and artwork for your favorite web pages. Netscape does a pretty good job of cleaning this directory up, so it may save you some time if you exclude it from the scan.

To exclude the CACHE directory, enter the following:

~?:\netscape\cache*.*

In the example above, the Tilde (~) character tells DupeFind to *Exclude* any files that match the wild card. The \netscape\cache sub directory can be on any drive, since we used a question mark (?) in the Drive Letter position of the wild card path. In this example, the "?" will prevent DupeFind from looking in the cache on any peer workstation.

You may also wish to exclude the OS/2 system directories on peer networked workstations. This will save you substantial scan time and bogus duplicate file entries. Use syntax similar to the example above.

Other clever tricks you can do with wild card file matching:

[a-m]*.*	Any files beginning with letters A-M
[a-z,~ux]*.*	Any files beginning with a any letter except X or
*.[~q]?	Any files except those that match *.?Q?
foo[abc]*.*	Files of FOO?*.* where '?' is A,B or C
foo["abc"]*.*	Same as above
foo[a-c]*.*	Same as above
test{0-9}.*	Files of TEST0.* through TEST9999.*
x{}z.*	Filenames beginning with X and ending with
x{0123456789}z.*	Same as above, only with numbers between X and Z
	Example: X1234Z.ABC
("read")*.*	Filenames that contain the text READ"

undocumented DOS wild cards (eg. FILENAME = FILENAME.* and .EXT = *.EXT).

Filenames or wild cards can be in quotes (eg. "*.*" is equivalent to *.*).

"GIF compare"

DupeFind can check for duplicate GIF files in two ways.   You can *Use Palette* or *Use Sig*.

GIF files have an internal table of all the colors used in the image. This table is called the "palette", because it is like the board an artist uses to hold the various colored paints necessary to paint a picture. Generally speaking, every GIF image will have a different palette, unless they are duplicates.   Do you see where we are headed with this concept?

"Use Palette" comparison to find GIF images that are similar, including images that do not match pixel for pixel. For example, it can find a sister GIF image that is rotated or slightly cropped. Heavy cropping, gamma correction or color shifting of an image will cause this comparison to fail, so its not foolproof. Another shortcoming of Palette Comparison is that gray scaled images frequently have a "standard palette" of all 256 possible gray scale values - even though the image does not use all 256 gray levels or "colors". This will cause all images with a standard gray scale palette to show up as duplicates. If you use Palette Comparison, don't forget to use the Graphics file Smart Viewer feature before deleting GIF files!

DupeFind only compares the first palette it finds in a GIF file. If a GIF image has multiple pages, only the first image's palette is used for comparison. DupeFind only "looks" at the palette and does not utilize the LZW compression algorithm used to store the actual image. It is therefore exempt from LZW patent and royalty problems. DupeFind can handle both GIF87a and GIF89a formats.

"Use Sig[nature]" to compare GIF images in a manner similar to other non-GIF files. It will find exact duplicates, but will not find cropped, rotated or resized images with similar content.

Why didn't we add a JPG palette comparison feature? JPGs don't have a palette, and use "lossy" compression. A given JPG image can have a nearly infinite number of file signatures if you make even the smallest change to one. This adds up to making them virtually impossible to match two or more visually identical JPG images. Even re-saving a JPG with PMView will make the new file signature completely different than the original file, even though both images will appear to be absolutely identical.

If there is enough interest, we might add a JPG comparison feature. Note that building the special JPG signatures for comparison with other images will be a *very slow* process if you want fairly accurate comparisons. Also possible, (but not included in v1.0) is a Windows/OS2 BMP and PNG comparitor. Please contact IDK for other file format requests.

"Database Name"

Enter the location and name of file catalog database file. DupeFind stores catalog information in a proprietary binary file format.

"Graphics Viewer"

Enter the location and name of your favorite Graphics File viewer. OS/2 ships with one (IB.EXE), but is not very good. Use it only if you are desperate. The built in OS/2 graphics viewer crashes on many common variants of JPG files. It is installed only if you installed MMOS2.
We highly recommend PMView for OS/2. It is the best file viewer on ANY platform. The program is truly amazing and well worth the very reasonable shareware price. PMView is widely available as full-featured Shareware on the Internet.

The Graphics Viewer must be a native OS/2 application that accepts the filename of the file to view as a command line parameter.

"Text Viewer"
Enter the location and name of your favorite Editor or Word Processor. E.EXE that ships with OS/2 is fine, as is IBM Works. Lotus Word Pro96 for OS/2 (or the impending new release) would be another possible choice, but it will cause major slowdowns, waiting for it to load each time. The Text Viewer must be a native OS/2 application.

"Binary Viewer"
Enter the name and location of your binary file viewer here. We suggest using KON, the shareware binary file viewer. It allows you to look inside EXEs, database tables, etc. and view the contents in Hex, ASCII and other formats. If you are not a programmer, the stuff you see with the binary viewer will be mostly meaningless. DupeFind also invokes the Binary Viewer when you attempt to view a file type it does not recognize.

"Sound Player"
Enter the name and location of your multimedia sound player application. For most OS/2 users with MMPM installed, this will be AB.EXE (short for Audio Browser, I guess). AB will play WAV files (recorded sounds and voices) and MID files (instrumental music). When you "view" a WAV or MIDI file, DupeFind will pop up the multimedia player applet and begin playing the sound or music file out your multimedia PC speakers. You need a sound card and speakers to use this feature, obviously. We have not tried other multimedia browsers, but any native OS/2 application that accepts filenames on the command line should work here. If someone finds a decent WAV, MIDI and MOD player that works well, let us know.

Dupefind v1.0
Users Manual

Smart Viewer Associations
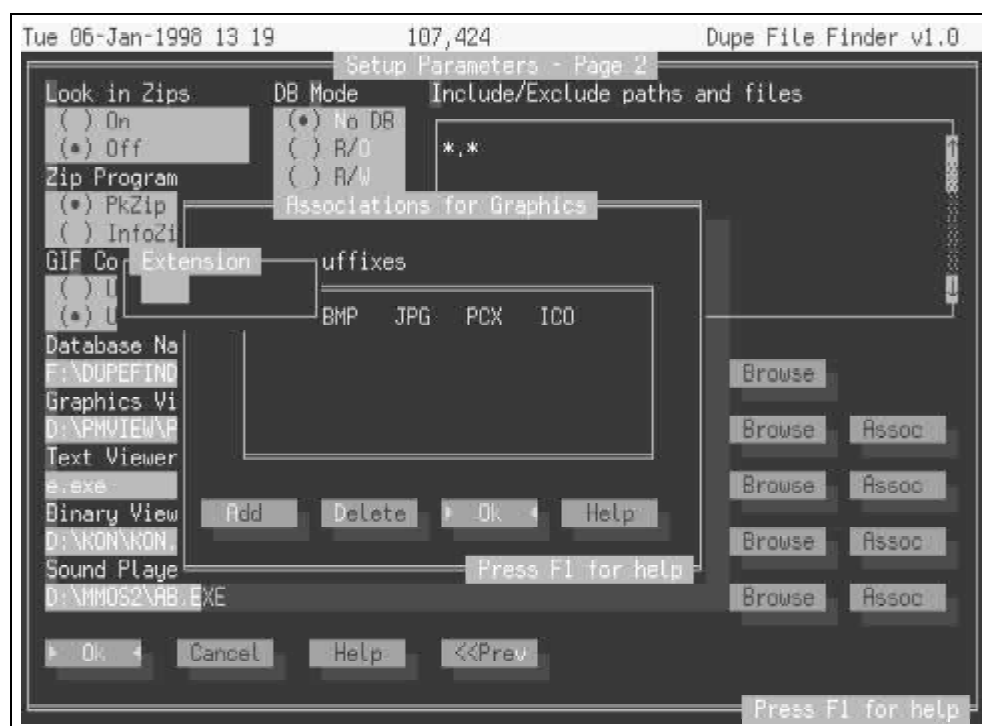
The screen shot above shows the effect of pressing the *Assoc* button on the Graphics Viewer line. The Graphics Viewer (PMView) is currently configured to pop up when the user attempts to view a BMP, JPG, PCX or ICO file. We can add another file suffix by pressing the *Add* button and typing in the new extension into the Extension field. You can delete a file extension association by highlighting it and clicking on the *Delete* button. The order of entry is not important.

Buttons

Browse
Click Browse to pop up the file picker. The file picker dialog allows you to change drives and directories, and to pick an existing file. You may also hand-type a filename into the dialog filename field. There is a separate File Browser to the right of each path/filename field.

Assoc
Click Assoc to edit the file associations (suffixes) that will activate the various Smart Viewers. There is an Association button associated with each Smart Viewer.

OK
Click OK to return to Setup Page 1.

Cancel
Click Cancel to return to Setup Page 1. Note: DupeFind uses "perfect save" similar to OS/2 "Properties" dialog boxes. If you change something, you can't back out of changes by canceling.

17

Help
Click Help (or press F1) to view help for this dialog. If you press F1, help for the currently highlighted dialog field will appear. Pressing F1 while the help screen is popped up displays the help table of contents. Alt-F1 jumps to the previously viewed help topic. Another trick: Press BOTH Left and Right mouse buttons on the area of the dialog box you want help on - help for that field will appear!

<<Prev
Click Prev to move to the Setup Page 1 screen.

Setup Complete, Now what?
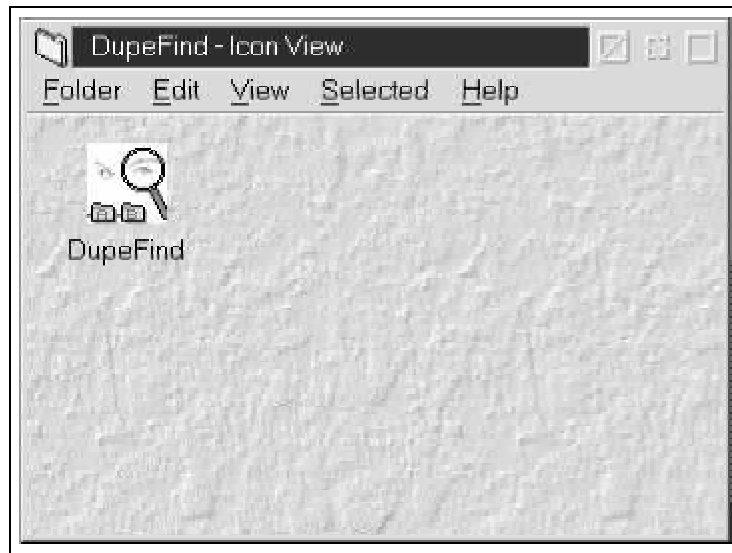
Above is a shot of a typical DupeFind folder on your desktop. Double click on the icon to start DupeFind.
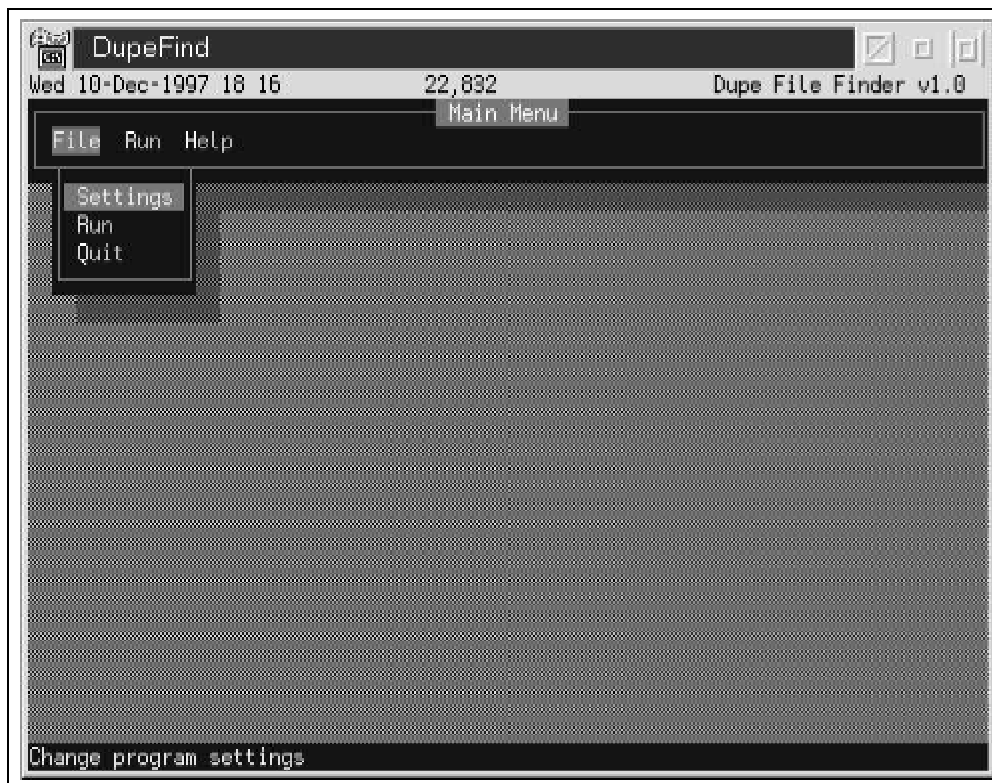
Here is a shot of the File menu pull down. Note that quick menu help is available at the bottom of the screen. You can also highlight a menu item and press F1 for more extensive help on that item. Most menu items have a hot key. Press the key that corresponds to the highlighted (White

On Blue) letter to activate the desired menu item. See the section on VoiceType Navigation below for speech control of DupeFind.

If the setup information is correct, then select either Files|Run or Run. You can also use the speech macro "Run".
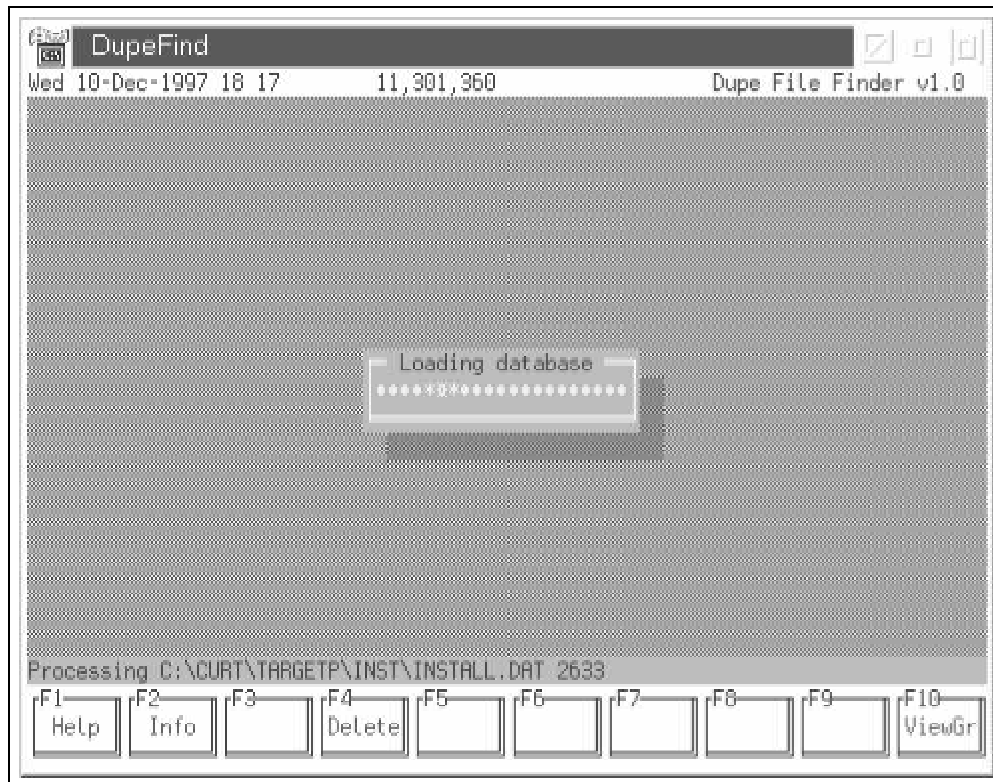
The screen shot above shows DupeFind scanning for files. Each time it finds a file, the black on cyan "processing" line updates with the current file name, if you elected to show file names in the Setup section of DupeFind. DupeFind is simply adding to its database all the files it can find that match your include/exclude criteria, drive letter list and explicit paths. The little squares at the bottom of the screen represent the function keys (F1 - F10) on the keyboard. They are not active until later - read on...
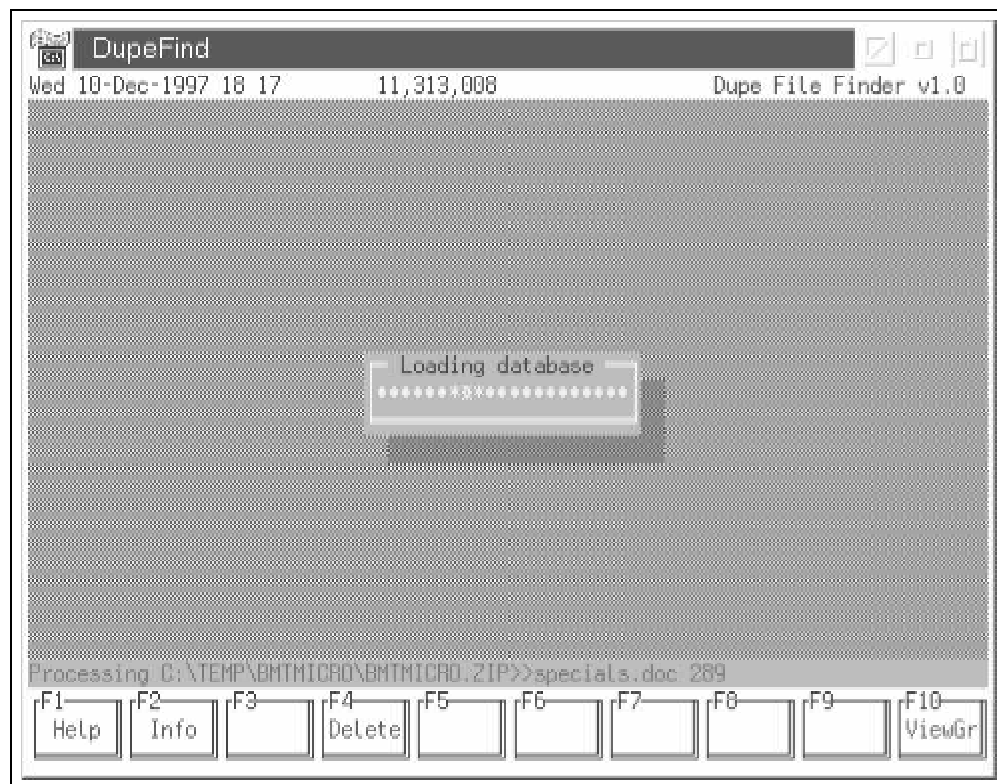
The screen shot above shows that the "Processing" status line color has changed. Why? DupeFind just found BMTMICRO.ZIP and is adding the Zip member file, named SPECIALS.DOC to the database. The number 289 means that SPECIALS.DOC is 289 bytes long. The moving dot indicator (Cylon) shows that files are being added to the database. DupeFind uses the ">>" character sequence to indicate members of a parent ZIP file.
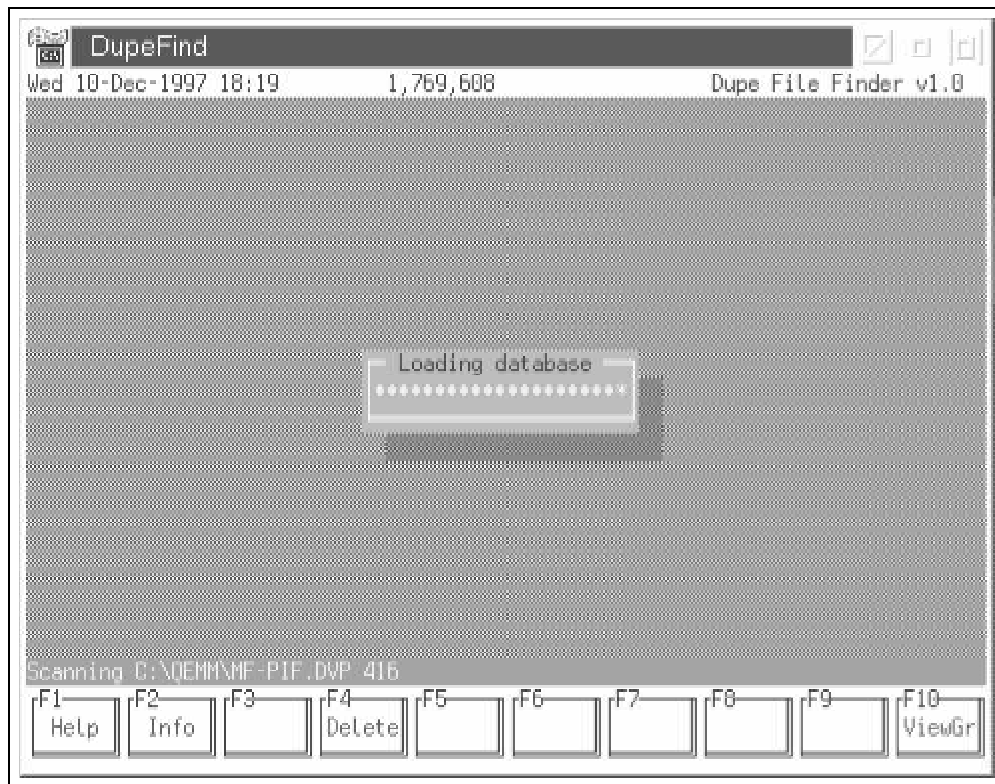
This screen shot shows yet another color status line. Yellow on Green means that DupeFind is calculating a signature for this file. 416 means that MF-PIF.DVP is 416 bytes long.
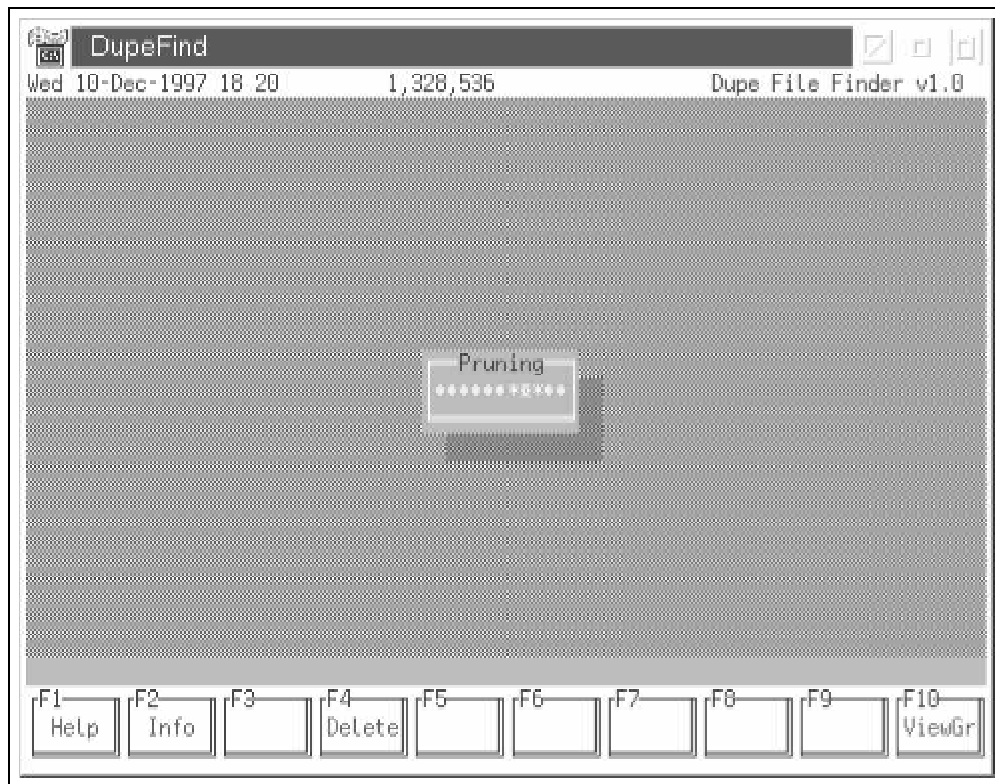
DupeFind is cleaning up the database of unneeded (non-duplicate) files. It is NOT deleting files from your hard disk yet.
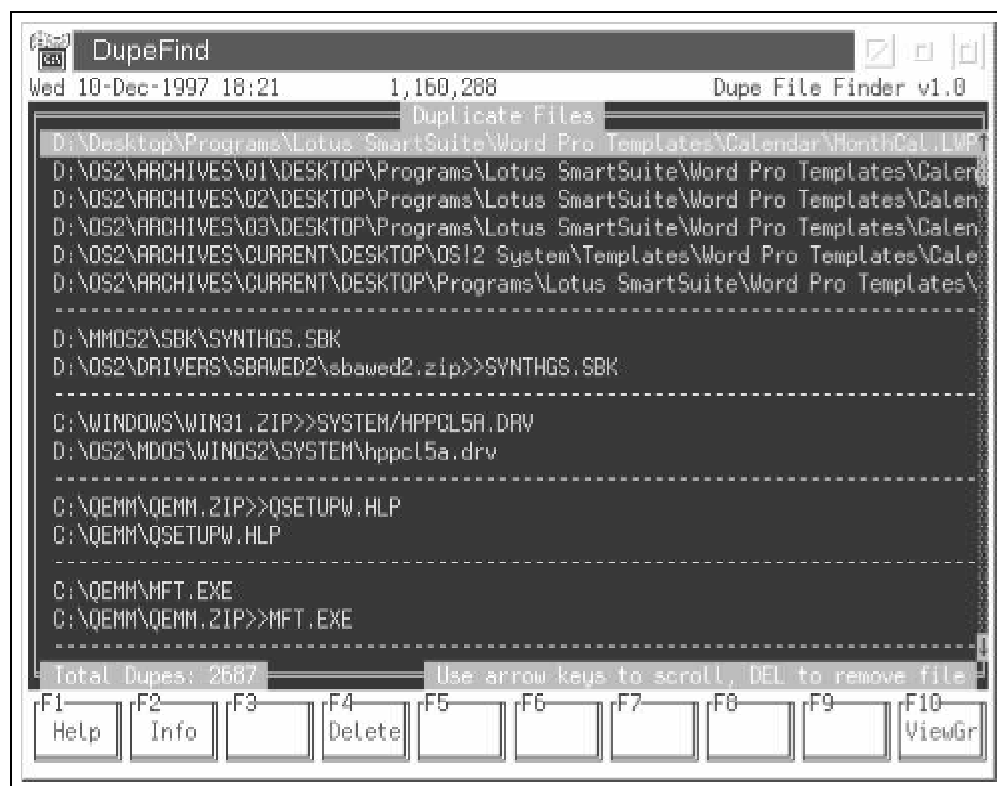
This screen shot shows the list of duplicate files found while searching C: and D:. At the top are 6 duplicate files located in the D:\OS2\ARCHIVES sub directories. While the files are huge (Lotus SmartSuite Templates), it is not a good idea to delete any of these. The ARCHIVES sub directories are used by OS/2 to reconstruct trashed desktops. This is a classic example of some files that you should leave alone.

The next group of duplicates shows that we have 2 copies of SYNTHGS.SBK. The first copy is being used by OS/2 to run a sound card on this workstation. The second copy is located inside a zip called SBAWED2.ZIP. We don't need this second copy, so its fair game for deletion. Deletion is easy: Just click on the "sbawed2.zip>>SYNTHGS.SBK" line, and press F4. You can also delete by pressing the DEL key, activating the "delete" speech macro or by clicking the mouse on the F4 function key label.

All function key labels may be activated by clicking them with the mouse. The labels update if you hold down the Shift, Alt or Ctrl keys. Example: Hold down the Alt key to see that the function key label for Alt-F4 is "Close".

Use the cursor keys, speech macros, or the mouse to page this display up and down. You can mark/unmark files for deletion by clicking with the mouse, tapping the spacebar or using a speech macro.

If you want to take a quick peek at the details of a file in the pick list then move the highlight bar to the desired file and press F2. A dialog box will appear showing the file size, location, signature and other information.

When deleting duplicate files note that DupeFind shows ALL instances of duplicate files. If you delete all duplicates in a cluster, you have deleted ALL the copies of your file. In most cases you will want to leave at least one copy showing on the screen!

Example:

I would probably want to delete C:\QEMM\QSETUPW.HLP and leave C:\QEMM\QEMM.ZIP>>QSETUPW.HLP. This would cause DupeFind to delete the larger copy of QSETUPW.HLP (the uncompressed one), and leave the master copy (compressed in the ZIP) alone. If I deleted both copies, I would be out of luck trying to run QEMM again.

Deleting Multiple Files

DupeFind allows you to "tag" multiple files in the duplicate file pick list. Press the spacebar, click with the mouse or use a speech macro to select a file. The "Toggle" speech macro is used to select/deselect a file. A file is selected when you see a check mark to the left of the filename in the pick list. Once you have selected a group of files, press or click on F4, use the DEL key or a speech macro to delete the group of files. If, prior to deleting a group of files you decide to spare a file or two from the group, simply reselect the file using any of the above methods to remove the check mark.
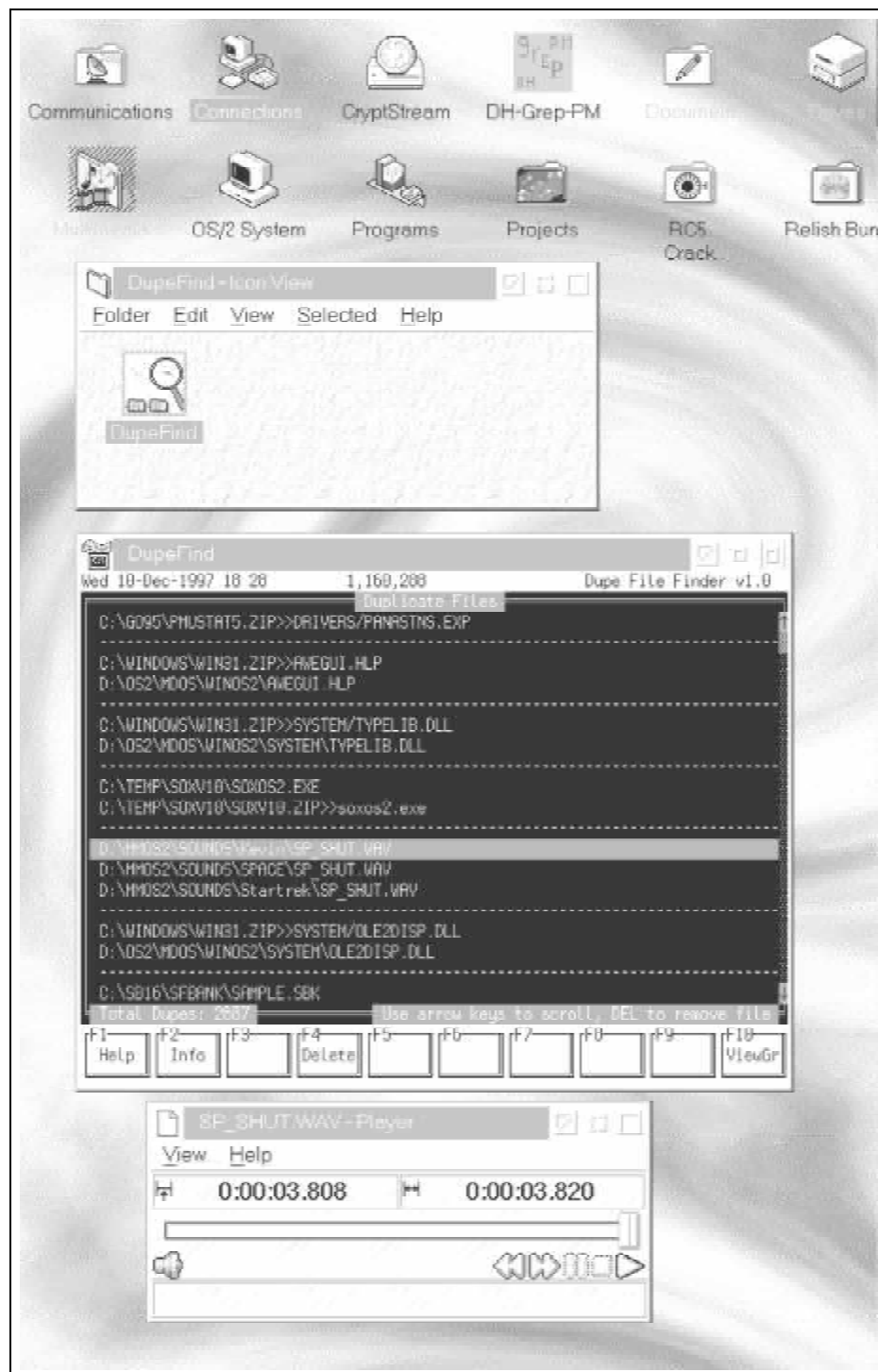
This screen shot shows the Smart Viewer feature in action. We have highlighted D:\MMOS2\SOUNDS\Kevin\SP_SHUT.WAV and pressed the F10 key. This activates the appropriate browser for that file type. In this case the audio browser was activated, since we are trying to "view" an audio file. If you highlight a graphics file, the graphics file browser will activate, and so on. See the second setup page documentation for instruction on how to set up this feature.

```
┌────────────────────────────────────────────────────────────────┐
│ [▤] VP DebAppl: F:\DUPEFIND\DUPEFIND.EXE          ☑ □ ▯         │
│ Wed 31-Dec-1997 16:09          448,216        Dupe File Finder v1.0 │
│ ┌──────────────────── Duplicate Files ──────────────────────┐  │
│ │ D:\OS2\ARCHIVES\03\DESKTOP\OS!2 System\Templates\GIFImage.GIF  ↑│ │
│ │ D:\OS2\ARCHIVES\03\DESKTOP\OS!2 System\Templates\Multimedia Templates\GIFImag│ │
│ │ D:\OS2\ARCHIVES\0X\DESKTOP\OS!2 System\Templates\Multimedia Templates\GIFImag│ │
│ │ D:\OS2\ARCHIVES\CURRENT\DESKTOP\OS!2 System\Templates\GIFImage.GIF│ │
│ │ D:\OS2\ARCHIVES\CURRENT\DESKTOP\OS!2 System\Templates\Multimedia Templates\GI│ │
│ │ ------------------------------------------------------------- │ │
│ │ D:\BRIEF31\MACROS\LANGUAGE.CB                                  │ │
│ │ D:\Desktop\Communications\Internet!(Modem)\Ultimedia^Mail!2 'Lite'\In-basket │ │
│ │ D:\Desktop\Communications\Internet!(Modem)\Ultimedia^Mail!2 'Lite'\Mail Cabin│ │
│ │ D:\Desktop\Communications\Internet!(Modem)\Ultimedia^Mail!2 'Lite'\Names and^│ │
│ │ D:\Desktop\Communications\Internet!(Modem)\Ultimedia^Mail!2 'Lite'\New Letter│ │
│ │ D:\Desktop\OS!2 System\Templates\AVCImage._IM                 │ │
│ │ D:\Desktop\OS!2 System\Templates\AVIVideo.AVI                 │ │
│ │ D:\Desktop\OS!2 System\Templates\Backup Set                   │ │
│ │ D:\Desktop\OS!2 System\Templates\Contact List                │ │
│ │ D:\Desktop\OS!2 System\Templates\Data File                   │ │
│ │ D:\Desktop\OS!2 System\Templates\DIBImage.DIB                 │ │
│ │ D:\Desktop\OS!2 System\Templates\Digital Video.FLI            │ │
│ │ D:\Desktop\OS!2 System\Templates\FORMATTED TRACE.FTF          ↓│ │
│ └──────────────────────────────────────────────────────────────┘ │
│ Total Dupes: 1852        Use arrow keys to scroll, DEL to remove file │
│ ┌F1──┐ ┌F2──┐ ┌F3──┐ ┌F4────┐ ┌F5──┐ ┌F6──┐ ┌F7──┐ ┌F8──┐ ┌F9──┐ ┌F10────┐ │
│ │Help│ │Info│ │    │ │Delete│ │    │ │    │ │    │ │    │ │    │ │ViewGr │ │
│ └────┘ └────┘ └────┘ └──────┘ └────┘ └────┘ └────┘ └────┘ └────┘ └───────┘ │
└────────────────────────────────────────────────────────────────┘
```

Ok, so why are all these filenames down at the "bottom" of the list displayed in white on blue? These are all zero length files. Since zero length files all have the same signature (none), they all get grouped together. Note: Many of these files are OS/2 Templates that should NOT be deleted. Even though they have a zero file length, they have Extended Attributes (EAs) attached to them - OS/2 needs the Template EAs to help it create various new desktop objects.

Feel free to delete any non-template files with a zero file length. Only delete files that you are sure will not affect workstation or network operation.
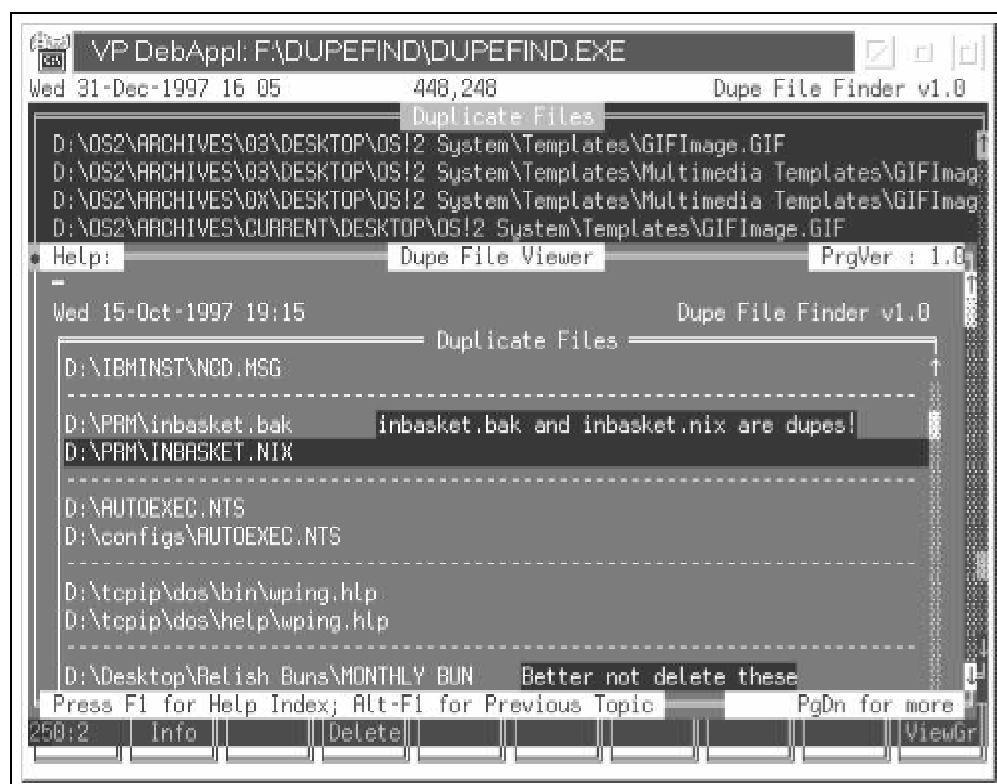
The screen above shows what happens if you press F1. The Magenta help screen appears, showing help for the current screen or field. Use the Page-Up/Page-Down keys to scroll the help page up and down. Once the Help page is on-screen, press F1 a second time to view the help index. The Help screen remembers all the places you have "visited" in the help system - press Alt-F1 to view the previous help topics. Press Escape key or click the right mouse button to close the Help screen.

Many help topics are "cross-referenced" like a web page. Cross-referenced topics are characterized by Yellow-On-Blue blocks of text. Double click on the yellow-on-blue block of text to jump there. Press Alt-F1 to return back to the previous topic.

Arcane Spells and Incantations

Command Line Options

Command line options override any corresponding INI file options. INI file options will be used as defaults if not overridden.

Usage:
DUPEFIND [[-Ad:\path] [-Bd:\path] | [-La..z]] [-Mx] [-S] [-O] [-Q] [-Dmm/dd/yyyy] [-Thh:mm:ss] [-R]
Where:
-A   First drive & path to search/scan - skips drive selector dialog
-B   Second drive & path to search/scan - skips drive selector dialog
-H Show help (this screen)
-Lxxx List of drive letters to search/scan, starting in the root of each.
-Mx Scan mode; x=0 Name comparison, x=1 Signature Comparison
-Q Quiet mode - slightly faster Search
-Rx Generate DUPEFIND.RPT log file; x = verbosity level 0..3
-S No beeps
-O Ignore read-only flag, delete R/O files unconditionally

Example:
DUPEFIND -LCDEFG *scan C,D,E,F,G drives*

INI File options:

DupeFind uses a special INI file to hold the information from the Setup dialogs. This file is editable by you, if you want to do specialized setups. The syntax of the INI file is fairly simple - it uses a format similar to a text outline.   Use E.EXE to edit this file, NOT a word processor!

In the example below, text in italics are descriptive text, and would not be found in the actual INI file.

```
*********************************************************************
* DUPEFIND INI File
*
* This is a hierarchical INI file format, similar to a text outline. Major
* headings are toward the left, with lesser subheadings indented. Values
* belonging to a subheading are preceded by an equal sign (=). You can
* enter data in any order, as long as the hierarchy is preserved. DUPEFIND
* will reorder/regenerate this file next time you run the program however.
*
* You can edit this file with any standard text editor. E.EXE works fine.
*
* Unless you know what you are doing, only edit the SUFFIX lines for the
* various VIEWERs. You can add as many suffix lines as you like.
```

```
*
* If you add additional undefined tokens, DUPEFIND will remove them when it
* resaves this file.
******************************************************************************


DEFAULT CONFIG
{
   DRIVE LIST
   {
        = DEFJK list of drive letters to scan
   }
   SEARCH PATHS
   {
        = f:\netscape first path to search
        = second path to search (none)
   }
   SEARCH MODE
   {
        = 2 mode 1 is dupe-name mode, mode 2 is dupe-signature mode
   }
   SEARCH CRITERIA
   {
        = *.* list of include/exclude search criteria. Use the setup screen to edit!
   }
   AESTHETICS
   {
      SHOW FILENAMES
      {
        = YES this is fairly self-explanatory
      }
      ALLOW BEEPS
      {
        = NO
      }
   }
   REPORT VERBOSITY
   {
     = 1        mode 1 = show stats only in log
   }
   SAFETY
   {
      IGNORE READONLY
      {
        = YES more self explanatory stuff
```

```
    }
  WARN IFNOTBACKEDUP
  {
    = YES
  }
}
ZIP setup parameters for Zip. Not a good idea to mess with this.
{
  PROGRAM BRAND
  {
    = 1
  }
  LOOK INZIPS
  {
    = YES
  }
  ZIP PATH
  {
    = pkzip.exe
  }
  UNZIP PATH
  {
    = F:\pkware\os2\pkzip.exe
  }
  CMDPARAMS
  {
    CRC
    {
    = /vi /nofix
    }
    DELETE
    {
    = /del
    }
  }
}
FILE DATABASE
{
  USE
  {
    = 0
  }
  NAME
  {
    = F:\DUPEFIND\allnet.STM
```

```
        }
     }
     GRAPHICS
     {
        PALETTECOMPARE
        {
          GIF
          {
          = NO
          }
          PCX
          {
          = NO
          }
          BMP
          {
          = NO
          }
        }
        VIEWER
        {
          = D:\PMVIEW\PMVIEW.EXE
        }
        SUFFIX
        {
          = GIF add more graphics file suffixes here. Use similar syntax. Case insensitive. NO
WILD CARDS!
          = bmp
          = jpg
          = pcx
          = ico
        }
     }
     TEXT
     {
        VIEWER
        {
          = e.exe
        }
        SUFFIX
        {
          = ME add more text file suffixes here. Use similar syntax. NO WILD CARDS!
          = TXT
          = C
          = CPP
```

```
          = H
          = PAS
          = BAT
          = CMD
        }
     }
     BINARY
     {
        VIEWER
        {
          = D:\KON\KON.EXE
        }
        SUFFIX
        {
          = EXE add more binary file suffixes here. Use similar syntax. NO WILD CARDS!
          = COM
        }
     }
     DOCUMENTS
     {
        VIEWER
        {
          = ibmworks.exe
        }
        SUFFIX
        {
          = DOC add more document file suffixes here. Use similar syntax. NO WILD CARDS!
          = RTF
        }
     }
     SOUND
     {
        VIEWER
        {
          = D:\MMOS2\AB.EXE
        }
        SUFFIX
        {
          = WAV add more sound file suffixes here. Use similar syntax. NO WILD CARDS!
          = MID
        }
     }
}
```

The INI file gets rebuilt each time you access the setup menus in DupeFind. If you manually put anything into the INI file that DupeFind does not "understand", it will be ignored and erased. If you get into real trouble, just erase a trashed DUPEFIND.INI. DupeFind will recreate it with "reasonable" default values.

VoiceType Navigation

We have included a Warp 4.0 VTN macro file that can be used to activate most of DupeFind's features. Installation of the macro file is optional and fairly simple:

After installing DupeFind, select the Speech page of the Properties dialog for DupeFind. Click the Restore button and enter the name of the VTN macro file: DUPEFIND.SPM.
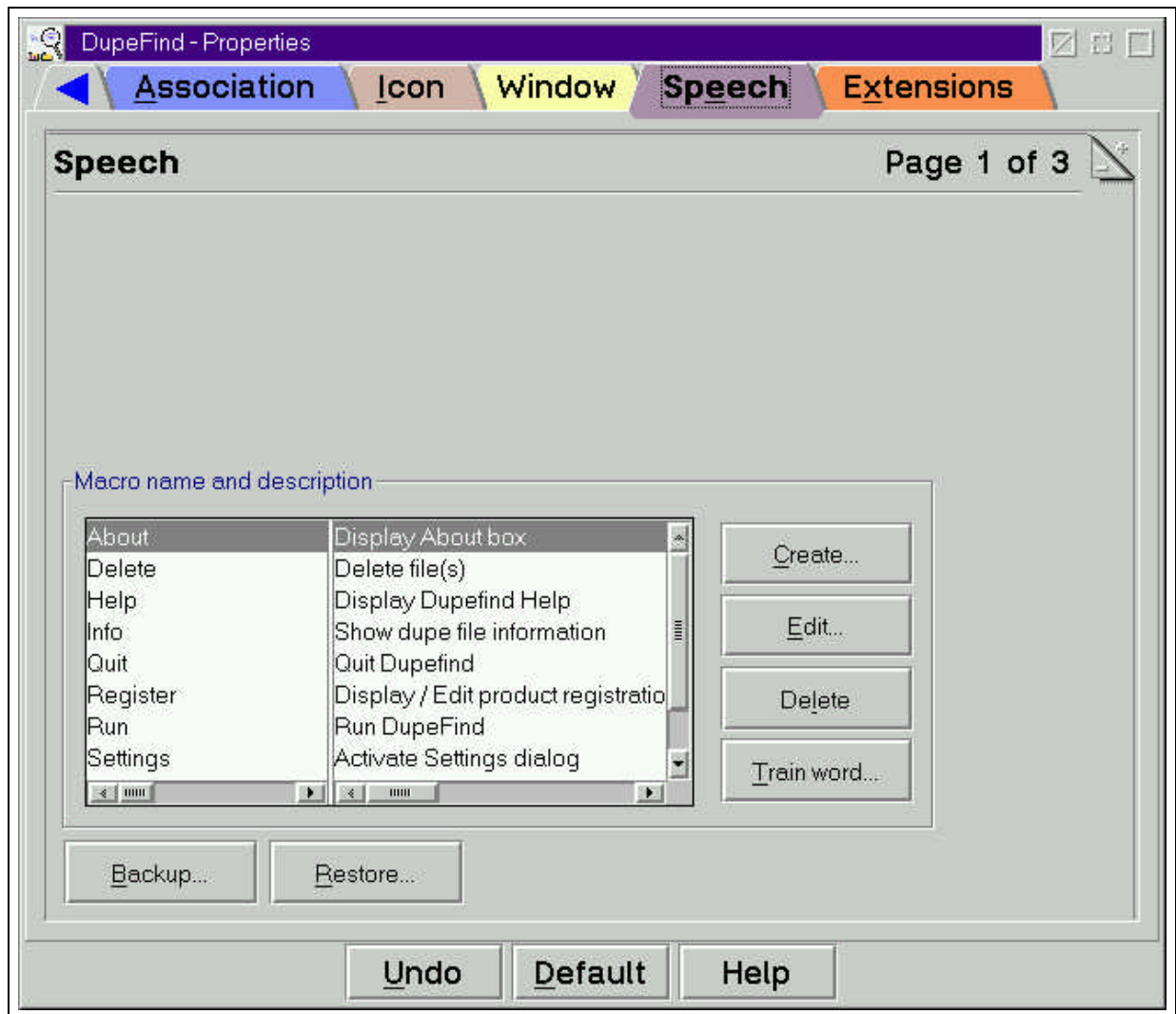
The macros you see above should appear in the Macro name and description list. Any names that look like: ~~toggle~~ needs "training". Select the macro item and click the Train Word button to add it to your custom VTN vocabulary. Use the microphone and say the macro name. Each person that uses your computer will probably have to train the same word. Feel free to change the names of the macros, or use non-English equivalents where appropriate. If you come up with any clever voice macros, send them to us for incorporation into future releases.

Note: If you don't see the speech tab in the properties dialog box for DupeFind, then you have not installed VoiceType Dictation. If your settings notebook looks radically different than the screen shot above, then you are running an older version of OS/2. VoiceType Dictation is included in Warp 4.0.

Review the list of macro names to see what commands are available. Not all commands are accessible from every location in DupeFind. The "What Can I Say" list of commands should show all the macro names when DupeFind has the focus.

Internationalization

DupeFind is easy to translate into most languages. Lets start with the ENGLISH.MSG file. This file is located in your installation sub directory.

ENGLISH.MSG is a text file list of most of the English phrases used by DupeFind on screens, error messages and menus. Each phrase has an index number like so:

        000021 'Problem adding file to master list: '
        000022 'Scanning '
        000023 ' Registration Information '
        000024 'Sorting'
        000025 'Pruning'

Translation of each phrase is as easy as overwriting the English phrase with the same message in your native language:

        000021 'foo bar blabba stuff: '
        000022 'blabba foo '
        000023 ' BlabbaBlabba FoooooBlaba '
        000024 'BlabbaStuff'
        000025 'FooStuff'

Not that my translation into "pseudo-caveman" left in formatting characters - spaces at ends of lines and colon (:) characters. Many of the phrases found in ENGLISH.MSG are concatenated with other phrases, and spaces may be important. Experiment!

Use E.EXE or equivalent as the editor for this file. Word processors will probably NOT work unless you really know what you are doing. You can use "high ASCII" characters if need be. Don't delete the single quotes at each end of the line. You don't need to "double up" single quotes inside the two end quotes:

        Correct: 000099 'This is Kevin's good example'
        Incorrect: 000099 'This is Kevin''s bad example'

Each line should end with a standard DOS CR/LF sequence. Trailing spaces after the end quote and before the CR/LF are ignored. Don't use TAB characters. Don't use some UNIX or Mac editor that uses only CR or LF or some other weird permutation of end-of-line characters.

Some of the Phrases that have an ampersand (&) in them are menu items or entry screen fields with a "hotkey". The letter to the right of the '&' is the hotkey for that item. You can move the & to a letter more appropriate for your language, or remove it if you like. If you remove it, there will be no hotkey for that field or menu item.

Example:

>000125 '&First Path To Search:'   'F' is the hotkey for this field
>000126 '&Second Path To Search:'   'S' is the hotkey for this field

Don't change phrases that start with a colon and are in all lower case:

>000308 ':gif'
>000309 ':pcx'
>000310 ':bmp'

The phrases above (and others like them) are used in the local INI file. You can change them if you really want, but it may make DupeFind forget the last INI file settings you used with English phrases. It will also make it difficult for me to troubleshoot any tech support problems you might have with the INI file (DUPEFIND.INI, not OS2.INI).

Don't try to change the supported File Suffixes (Smart Viewer) here. ENGLISH.MSG only has the default file suffixes used when the INI file is missing. Change Smart Viewer settings by editing the INI file itself. See INI file format documentation for more info.

The best way to translate the screen messages is to make "screen dumps" of each of the screens you want to translate. Find the corresponding message in the ENGLISH.MSG file, and try to translate it into a phrase in your language that will fit on the screen.

Note that some phrases are used in more than one place! Make sure the translated phrase makes sense in the other contexts where it may be used. If you run into a brick wall with language-context problems, contact IDK for help.

So, how do you compile the new phrases into the ENGLISH.MSX binary format? Easy - just put the modified ENGLISH.MSG file into your DUPEFIND.EXE subdirectory and run DupeFind. It will automatically compile the phrases if the MSG is newer than theMSX. You can delete the MSG after compilation, but we don't recommend this. There is no way for you to decompile the MSX back into its original form.

Don't forget to send me a copy of translated MSG files so I can make them available to other speakers of your language. No nasty words please - I will verify your translation accuracy with other independent sources prior to bundling them in future versions.

I will be happy to give you credit in the About Box, if you provide me with a good translation. Use your real name please.

Translating Help text

The text for online help is found in DUPEFIND.TXT. Since this file contains cross referenced help topics, it is a little complex to deal with. Additionally, you need a special compiler to change the TXT file into a HLP file. If you are interested in translating this file, please write us at dupefind@idk-inc.com. Feel free to begin translation, and I will be happy to compile it for you and send you the results. Again, use a standard text editor.

The following are examples of help topic numbers. You can change the text part, but DO NOT change the topic numbers! Topic names must match in all cross references.

        !CONST
        File = 1
        Settings = 2
        Run = 3
        Quit = 4
        Run1 = 5
        Help = 6
        General Help = 7
        About Box = 8
        Register = 9

Cross referenced topics are indicated by:
@# $ %^ &* ( ) - =
<spade>General Help<club>General Help<club>

The spade character ($04) marks the text as a hypertext link (cross reference). The club character ($05) marks the actual index that will be activated if you click on the link with the mouse.

"!LINE" inserts a line break;

"!Topic Help" marks the following block of text as belonging to the topic "Help"

"!Index 7" marks the following block of text as belonging to index #7 - the seventh item in the index list.

"!NOINDEX" marks the block of text as being only available as a help topic or cross reference. The block will not appear in the Help index list.

"xxxxx" makes "xxxxx" appear in an alternate color.

";" at the beginning of a line marks the line as a comment. It does not appear in the help.

Registration

DupeFind REGISTRATION FORM

(Please answer questions required for registration and
send it to your local DupeFind Support Site with the
registration fee) Please PRINT this section clearly.

Note: The First, Middle Initial and Last Names found below
are used to register your product.

First Name: _____

Middle Initial: _____

Last Name: _____

Company: _____

Address: _____

City: _____

State/Province: _____

Zip/Country: _____

Phone: _____ Phone: _____
         (voice)   (data)

E-mail: _____
         (Internet/CompuServe)

1 copy   ( ) $35.00 each

2 - 9 copies ( ) $27.00 each   10 - 24 copies ( ) $21.00 each

25 - 49 copies ( ) $16.00 each   50 - 99 copies ( ) $12.00 each

100 - 149 copies ( ) $9.00 each 150 - 199 copies ( ) $8.00 each

Site license   ( ) $1500.00 for unlimited usage

Number of copies to register: _____

 Send registration key via:

   ( ) E-mail to your address (usually within 24 hours of funds receipt)
   ( ) Postal (snail) mail to your address

Sending check or money order (U.S. Funds ONLY PLEASE!), please use:

   Postal
   address   :

               IDK, Inc.
               2217 Aspenpark Ct.
               Thousand Oaks, California, 91362-1731
               USA

IDK can provide only minimal support to non-registered copies of DupeFind. If you want us to do further development of DupeFind or other OS/2 products, vote with your pocketbook! Registration shuts off Demo mode, allowing you to delete more than 10 files in a session.

A special trial registration is available that allows unlimited file deletions, but limits the product to a specific lifetime. The product will stop working after the drop-dead date expires. Contact IDK for details.

To register DupeFind, select the Help item from the main menu, and then pick the Register menu item. The following screen should appear:

Enter your name, EXACTLY as you did when you filled out the registration form.

Enter the Serial Number and Registration number exactly as found on the E-mail or letter we send you after receiving payment. If the registration process goes as planned, you will return to the main menu after entering the Registration number without seeing a warning message. If a warning message appears, you did not type something in correctly. Try again.

DupeFind will create a small OS/2 INI file entry on successful registration. If your OS/2 INI file gets trashed, or you do a fresh install of OS/2, the DupeFind registration information will be lost. Simply reenter the name and number data to reregister. We advise saving the registration information in a safe place.

Memory Use and the Swap File

DupeFind uses considerable amounts of virtual memory when scanning large disk drives or networks. On our network containing approximately 150,000 files, DupeFind uses approximately 12 megabytes of virtual memory. If your workstation does not have this much free memory available, DupeFind will begin using Swap file space. Your swap file may grow much larger than you are used to. OS/2 is infamous for not "shrinking" the swap file back down to its original size after an application releases all the virtual memory it used. Your only recourse in this case is to reboot the system to force the swap file to shrink.

DupeFind's database has been tested with 300,000 files without problems.

DupeFind goes to considerable trouble to release all the virtual memory it uses. If you scan a very large number of files, DupeFind will spend a minute or two releasing the memory when you quit the program. DupeFind will display a "gas gauge" showing the progress of virtual memory release. The red/green bar graph above the function key labels shows the progress of memory release. If you find this process to be annoying, simply minimize the DupeFind screen and continue working with other applications on your desktop.

The amount of virtual memory consumed by DupeFind is related to the number of files found, and the length of each corresponding path/filename. Memory consumption is not related to scanned file sizes.

If you have problems with swap file overflow, we recommend moving your swap file to a drive with more free space. This will help all your other applications - not just DupeFind. Try to use a fast drive if you have a choice. The line in CONFIG.SYS that controls the swap file looks something like this:

    SWAPPATH=D:\OS2\SYSTEM\ 8192 16384

This example shows that the swap file is placed on the D: drive, with an initial size of 16 Mb; An OS/2 warning message appears when there is less than 8 mb free. If you want to move the swap file, change the drive letter and path to point to a valid location on a different drive, then reboot. After successfully moving the swap file to a different drive, don't forget to go back to the old location and delete the original swap file: SWAPPER.DAT.