# Contents

What Is *Word Sniffer* ?

Word Sniffer is a search utility for Windows® NT and Windows 95 that you can use to do three things:

1. Search for files that match a certain pattern specified in the 'Search in:' text box. Used this way it's a quick and simple file finding tool.

2. Search for words that you don't know the exact spelling of. This is the most important feature of the program. Just indicate the file or files to search in, type the word you're looking for the way you *think* it's spelled and press Enter. *Word Sniffer* will turn up all the words that sound more or less like the one you entered (of course, it will also find those words that are spelled exactly like it). See Word Searching for details on this search mode.

3. Regular expression searching. This mode is useful when you need to find files contents of which match the expression you specify using words and logical operators (OR, AND, NOT, XOR). See Expression Searching for further details on this mode.

See How to Use   *Word Sniffer*   for help on buttons and options.

How To Use It

Click anywhere on the picture below where the cursor turns into the hand symbol to find out what each area in the application's window means and what options are available.
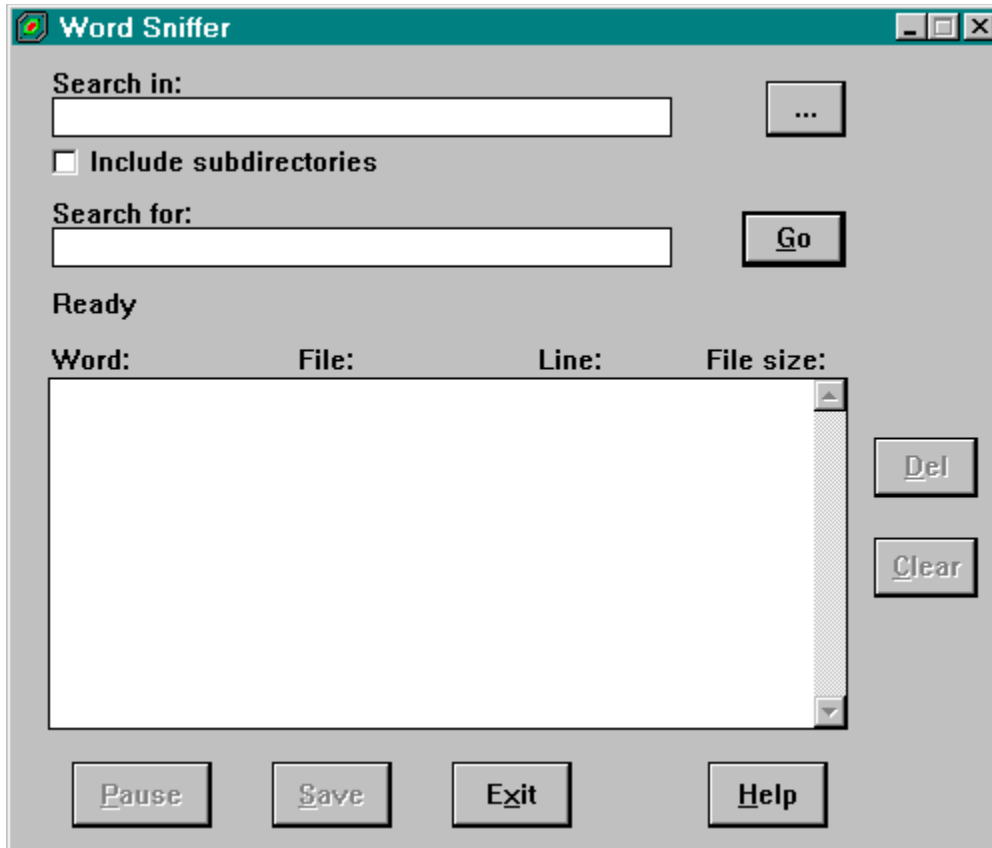
See <u>Word Searching</u> and <u>Expression Searching</u> for more details.

Word Searching

This is probably the most useful feature of the *Word Sniffer* – the ability to find words that you have only a vague idea how to spell. It is handy when searching a large word-processing file, a spreadsheet, or a database (unencrypted, of course) for names of people or places with most of their possible variations (like Caren, Karen or Kerin). *Word Sniffer* is English-language specific and is not very useful for texts in languages that have different pronunciation rules.

Note *The letters in the files must be in the ASCII/ANSI format. Binary, Unicode, encrypted or otherwise different letter formats are not recognized. The words also cannot contain any non-alphabetic characters inside them (e.g.,* **gla4ss** *is an invalid word for the purposes of searching). This is* **not** *an exact text search tool.*

You conduct a word search (in contrast to expression search) when you type a *single word* in the 'Search for:' text box. The program will then look for any words that sound more or less like the one you specified. Vowels are omitted (except the beginning one), that's why you may get lots of extraneous words (and, then, you may not; depends on what is in the files you search). The rest of the letters are matched as follows. Case does not matter.
Where 'c' might sound like 's' both letters are looked for. For example, if you type in "ciense" (I know you mean well) you will get your "science" (if it's there). And if you enter "sycle" you will get "cycle" (and "sycle", too, if you have friends). On the other hand, where 'c' might sound like 'k', 'k' will also be checked for (that's the Kerin from above, which you thought was spelled Caren). 'K' will also match 'c' (where 'c' can sound like 'k', like the second 'c' in that 'cycle' from before). 'Sh' combinations will also match 'sch' (here comes that Schneider you didn't know how to spell) and 'ch; and 'ch' will turn up 'tch' and 'sh', except when the word begins with it (were you thinking about catch, cache, or maybe cash? Try 'cach' and you'll net all three). Where 'g' may sound like 'j', 'j' is also looked for, where appropriate, and vice versa ('j' to 'g'). Where appropriate means, if you type 'jumbo', 'gumbo' will not be there to laugh at. Where 'x' can be thought to be spelled with 'z', 'x' will find its way into the list (you know Xena, that girl with the big ring, now you can be sure it's spelled Zeena and still get to her. What is she doing in your files, anyway?). And if you think 'thorax' is spelled 'thoraks' (or even 'thoragz'), it's ok, too. But note, that 'x' in the beginning of the word is not matched to 'z' (that is, 'xenith' is no good if you're looking for 'zenith'). It **is** matched to 'ks' in the beginning of a word . Anything that sounds like it has 'shn' (like 'promotion') can be searched for using 'tion' (or 'tian', or 'tiun'). So, 'permission' and 'pension' are yours to spell with a 't'.

But note, that 'sion' (or 'ssion') won't look for words correctly spelled with 't'.

Beginning 'ps' that sounds like 's' can be looked up using 's' ('soriasis', for example).

'Ph' can be matched with an 'f' (i.e., 'farmacy' is ok), and vice versa (find 'f' using 'ph').

The words in the files can be separated not only by spaces to recognize them as words, but also by any the following punctuation marks:

. , ! ? " # $ % & ' ( ) * + - / \ [ ] { } | ~ : ; < > = @

These characters are delimiters, which means there don't really have to be any spaces to separate the words in the file. The funny characters with ASCII codes above 127 are also delimiters, which means you can search binary files like .EXEs or .DLLs.

See also Expression Searching.

Expression Searching

This type of searching is different from the word mode in that in the regular expression mode the program's objective is not individual words but *files* that conform to the pattern you specified. All the rules of word matching described in the <u>Word Searching</u>   topic apply to expression searching. This means that matches can be looser than you'd like.

All of the four logical operators can be used, according to the following syntax:

OR – use **o** or **O** to separate two words to find all the files (subject to what's in the 'Search in:' edit box) that contain any one of the words that *sound like* the two words separated by this operator. For example, if you type "classic o pentium" (without the quotes) in the 'Search for:' box, the program will list all files that have any word that sounds like 'classic' or 'pentium'. So, if file A contains a word that sounds like 'classic' but not one that sounds like 'pentium', and file B contains a word that sounds like 'pentium' but none like 'classic', both files will be in the result list, because each one of them contains at least one word from those specified. This is the default operator; that means if you type two words without an operator between them, they will be looked for using the OR operator.

AND – use **a** or **A** to find all files that contain both of the words that sound like the ones you specified. Using the previous example (substituting "classic a pentium" for the search string and assuming that both A and B do not contain any more words that sound like 'classic' and 'pentium') the search will list zero files as matching the expression. But if there is a third file C that contains a word that sounds like 'classic' **and** one that sounds like 'pentium', you'll get a match.

NOT – use **n** or **N** to check for files that have the first word in the expression (the first one is the one to the left of the operator), but not the second. That is, "picture n magazine" will turn up all files that contain at least one word that sounds like 'picture' and no words that sound like 'magazine'.

XOR – this is exclusive OR. Use **x** or **X** to search for files that contain at least one word that sounds like either one of the words in the expression, but not both of them. So, "picture x magazine" will show all files that contain words that sound like 'picture' or like 'magazine', but will exclude all files that have both of them.

You can specify a complex expression that contains more than one operator and more than two words, like this "Johnson a Parker n Stillwell o Perkins". Don't forget, the search order is always left to right, which applied to the previous

expression means this: first all files (out of those matching the 'Search in:' pattern) that contain anything that sounds like Johnson will be marked. Then, out of these found will be chosen those files that in addition contain at least one word that sounds like Parker. Then, this narrowed down list will be checked against the 'n' operator to exclude all files that have at least one word that sounds like Stillwell. And at last, the whole file list (not only those that came up after the previous comparisons, because the next operator is 'o Perkins' which matches files having something that sounds like Perkins regardless of the results of the previous expression) is tested against the 'o' operator. So, at the end of this search any files that do not contain anything that sounds like Perkins will be excluded from the results list, but those files that didn't make through the "Johnson a Parker n Stillwell" test still have a shot if they can come up with something that sounds like Perkins (of course, those files that did make it through the "Johnson a Parker n Stillwell" will also be listed).

One more thing. You can use a single operator followed by a word to make a quick listing of files that have at least one word sounding like the one in the expression (operators **o**, **a**, and **x** are the same used this way), or those that do not have any words sounding like the one specified (operator **n**). Like this: 'o Shnoller', or 'n Shnoller' (use the latter if you have bad associations with either Shnoller or Schnoller - the search will skip all files containing anything that sounds like him).

Tips for Searching

Based on what's said in the <u>Word Searching</u> topic, here are some tips
to get you where you're heading.

Use 'sh' (instead of 'sch' or 'ch') letter combination to match both 'sh' and 'sch'
in the beginning of a word. Use 'ch' to catch 'ch', 'tch' or 'sh' anywhere else.
Use 'tion' (instead of 'sion' or 'ssion') in anything that sounds like 'shn'.
Use 'ks' rather than 'x' to match both 'x' and 'ks' (and also 'gz').
Use 'z' in the beginning of a word if you're not sure whether it's spelled with a
'z' or an 'x'. Use either of them anywhere else.

Although *Word Sniffer* will catch words that start with 'ps' but sound like just
's' (psychology, etc.) even if you start the search word an 's', there are letter
combinations that aren't looked for that you should be aware of. Here they are:

If a word sounds like it has an 'f', try using 'gh' (like in 'tough').
If a word seems to have an 'o', it might really be spelled with 'ough' or
'augh'. Similarly, 'i' or 'y' can have 'igh', 'ygh', or 'aigh' behind them.
If it looks like there is a 'd', try searching on 'ld'.
If you think a word starts with a 'z', try 's' - some beginning s's sound like z's.
If you can't find a word starting with 'ku' or 'cu', try starting it with 'q' instead.
And try 'wr' after trying 'r' to start a word that might belong to the 'wrong' family.
And the most common pitfall: an 'n' in the beginning or middle of a word.
Do try to find it with 'pn', 'kn', or 'gn'.

'Search in' text box

Type the name of the file(s) you want to search in this edit box.
You can also use the usual Windows® wildcards (? - to match any
one letter in the filename, * - to match zero or more letters). Check
the 'Include subdirectories' checkbox to search all the subdirectories
of the directory you specified for the filename (or the wildcard specification).
If you don't know the name of the file, press the Browse button ('…')
to find your way to it.

'Include subdirectories' checkbox

Place a check in this checkbox to search all the subdirectories of the directory specified in the 'Search in' edit box for the file (or files) you indicated.

Browse button ('...')

Press this button to browse for file you want searched.

'Search for' text box

If you just want to find the *words that sound* like the one you need,
type it in this box the way you think it is spelled.
If, on the other hand, you are looking for *files* that conform to certain search
conditions, you would type in this box the regular expression you want matched.
See Word Searching and Expression Searching for the details
on   differences between the two types of searching.
You can also leave this field empty to generate a simple listing of all the files
that match the pattern specified in the 'Search in' field.

'Go' button

Press Enter or click this button to start a new search or stop a search in progress (while a search is in progress the caption of this button changes to 'Stop').

You can temporarily pause and then continue the search using the 'Pause' button.

Status line

Displays the names of the files as they are being searched and the results of the search when it's finished.

Search results list

Displays the results of the search.

If you're conducting a word search, the following information will be displayed under approriate labels:

'Word:' – The word matched in the file. Long words are 'cut to size' to keep them from invading the next field.

'File:' – The name of the file in which the word to the left was found. It is also cut short if needed.

'Line:' – The line number and the position in the file (in percents of its total size) at which the word occured. The line number is only accurate for text and word-processing files. It does not reflect anything useful if the file is a database, spredsheet or the kind.

'File size:' – What it says it is. The size is in kilobytes, so files smaller than 1 K will be reported as having a size of zero.

If you specified a regular expression as the search string or left it empty, the list will display the filenames of the files matched.

Double-clicking an entry in this list will open the file in its associated program. Note that double-clicking a file whose type isn't associated with a program has no effect.

'Delete' button

Select an entry in the list and press this button to remove it.

It **does not** delete a file or a word from the file, only helps you to narrow down your search.

'Clear' button

Press this button to clear the results list and all information about them.

'Pause' button

Press this button when the search is in progress if you decide to be gentle to the machine and give it a rest. Press it again (the caption will be 'Resume') to remind your computer of its responsibilities. Or press 'Stop' to stop the search altogether.

'Save' button

Enabled when there is something in the list. Press it to save the search results in a text file. The information will include the search word, number of words and number of files that matched (there can be more than one match in each file). The listing will include the complete filenames (not the short names in the results list) and also the files that could not be opened (because they were used exclusively by another program, or some other reason).

'Exit' button

Press it to exit the program. (Do not press it while a search is in progress - press 'Stop' first.) Any existing search results will be lost unless you save them. You will **not** be prompted to save results before exiting.

'Help' button

The button to use if you're lost and start panicking.

It will fetch you the help file you're reading right now.