

Editorial tasks

As you read this chapter you will find it helpful at times to sit at your computer with a Microsoft Word file open so that you can view the suggested menu options. Microsoft Word provides the ability to personalize the way that certain menus and buttons appear on your screen. As a result, your screens may appear slightly different from the descriptions in this book. To begin, you should check to be sure that you are able to view the Standard, Formatting, Drawing, and Outlining toolbars. To do so, select “Toolbars” under the “View” menu and be sure that Standard, Formatting, Drawing, and Outlining⁷ are selected.

Figure 1 – The selections on the Microsoft Word menu bar



Using Microsoft Word to write a better manuscript

I'm going to assume that you can write well, or at least well enough to meet the task at hand. You will need to look elsewhere if you want advice on issues of grammar or style. However, there are a variety of tools available through Microsoft Word that can make it significantly easier for you to write a typo-free,

⁷ Outlining may not appear as a selection if you are not already in Outline mode.

grammatical manuscript. You would be foolish not to make liberal use of them. Many of these tools are found under the Menu Heading called “Tools”, including a spelling and grammar check tool, a custom spelling dictionary, an automatic correction tool, a thesaurus, and a word count tool.

Spelling and grammar

The “Options” menu under “Tools” includes a section called “Spelling & Grammar” where you can set Microsoft Word to check your spelling and grammar as you type. This convenient tool points out when you make simple mistakes. When the spell checker uncovers a spelling error, it underlines the word in red. When it spots a grammatical error, it underlines the phrase in green. You may choose to ignore these recommendations, but in general they do provide helpful advice. If you find the spelling and grammar check distracting, you can always turn it off, but you should remember to do a final spell check before submitting your job for print.

Working with a custom dictionary

The spell check feature works best when it is used in conjunction with a custom dictionary. Word comes installed with a custom dictionary called CUSTOM.DIC. If you use technical or other unusual words that aren’t in Word’s dictionary, you can add them to the custom dictionary so that they will be recognized in the future. If you right click⁸ with your mouse on a word underlined in red, it gives you the choice “Add”, which if selected, will add the word to the custom dictionary. Be careful with this. If you add a misspelled word to the dictionary, Microsoft Word will no longer alert you to the error. If you want to check the words in the custom dictionary, then go to the Tools menu, click “Options”,

⁸ The mouse for a Microsoft Windows personal computer comes with two buttons. You can use the left button to make menu selections, select blocks of text, and to place your cursor. Clicking the right button often provides a menu of additional choices.

and then click the “Spelling & Grammar” tab. If you click on “Dictionaries” and select the one you want to edit, it will open the file so you can add, delete, or edit words. While you do this, Word will automatically turn off spell checking, so you’ll need to turn it back on when you’re through.⁹

Automatic correction

Word’s AutoCorrect tool corrects or even formats automatically while you type. You can turn this feature on in “AutoCorrect” under the “Tools” heading. Depending on the options you choose, this feature can correct common typing mistakes (for example, it can change “freind” to “friend”) or capitalize the first word in a sentence. All of this happens automatically. You can also customize Microsoft Word to make substitutions. For example, it can automatically change “percent” to “%”.

At first it is unsettling to see the computer correcting your mistakes, but after a while I think you’ll come to like it. It is also handy if you frequently type foreign or accented words. Instead of having to remember the “press the Alt key plus some number” keystroke sequence¹⁰ for an accent, you can simply set AutoCorrect to replace the unaccented word with the accented version. To do this, go to “AutoCorrect” under the “Tools” heading and type in the unaccented version under “Replace” and copy and paste the accented version under “With”. Be sure that the box labeled “Replace as you type” is checked. However, if you are working in a foreign language you can set the appropriate language for the spelling dictionary in “Languages” under the

⁹ You do this by selecting “Tools” and then “Options”. Under “Options”, select the “Spelling & Grammar” tab. Then activate “Check spelling as you type”.

¹⁰ To find the correct keynote sequence for an accented character, you must use a Windows utility called “Character Map”. To open Character Map, go to the “Start” button in Windows and select “Programs”, then “Accessories”, and then “Character Map”. Within Character Map you can select the appropriate font and then copy the desired character.

“Tool” Menu. You can even do this for individual segments of your document that appear in a foreign language.¹¹

AutoCorrect also helps improve the use of certain specialized characters. For example, if you type three consecutive periods when AutoCorrect is on, they will be converted to an ellipsis, which is the correct typographical character to be used in such a situation. AutoCorrect will also change straight quotes to curly quotes. These may seem like minor corrections, but they give a much more professional appearance to a document. However, just because you have AutoCorrect on does not mean that all such specialized characters will be inserted. If you cut and paste a section from another document, AutoCorrect will not be applied to it. You will need to proof your document closely for these errors.

A quick way to see if AutoCorrect is turned on is to type a colon, a close parenthesis, and then a space. If AutoCorrect is on, the two characters will turn into a “smiley face” (☺) character.

Thesaurus

Microsoft Word’s Thesaurus can help you find the right word (or even help you find a substitute for an often-repeated word). To find a synonym, just place your cursor in the word that you want to check and then go to the “Tools” menu. “Thesaurus” will be an option under “Language”.

Word count

Microsoft Word also provides an easy way to get a word count. Just go to “Word Count” under “Tools” and it will give you not only the word count, but also counts for pages, paragraphs, characters, and lines. Word can even give your document a “Readability” rating after you have done a spelling and grammar check. To get

¹¹ Later, in the chapter on design tasks, you will learn how to create a Word style that assigns the appropriate language for spell checking.

this rating, be sure to click “Show readability statistics” in the “Spelling & Grammar” tab in “Options” under “Tools”.

Don’t overestimate these tools!

Despite these wonderful tools, you will still make mistakes that go undetected. The grammar check tool misses many simple problems¹² and the spell check is powerless to help you identify a “wordo” (i.e., when you type “site” instead of “sight” or “dig” instead of “dog”). These errors are more likely to be found by a competent proofreader who reads your manuscript carefully. I highly recommend that you hire the services of a qualified proofreader before you take your book to print. However, you can make the proofreader’s job easier if you use the tools described above and you print out and show your manuscript to others for comment. Keep in mind that someone who reads your manuscript and tells you that it is “good” has not provided you with any useful information. (I’m hoping that you thought it was “good” already.) Give your early readers something to look for, specifically typos and wordos. In addition, ask them to underline or mark sections where they were confused or puzzled. If one reader is puzzled, chances are that other readers will struggle with that section too. This type of information is invaluable, but many readers will hesitate to tell you what they think because they don’t want to hurt your feelings. Choose your early readers with this in mind. They should be people whose opinion you trust and who will tell you what you need to know. Your final proofreader should be willing to give an extremely careful look at your manuscript. However, you may wish to have them proofread one chapter first to see if they are identifying important problems instead of nit-picking you to death.

¹² Sometimes the corrections it suggests are outlandish. For example, when I wrote: “In fact, a lot of times these roles get all mixed up and tasks get done out of order”, it put a green line under “get done” and suggested the following options: “get did”, “get had done”, and “get has done”. Yet even when the grammar check makes a nonsensical suggestion, it still may point out to you that a sentence needs some work.

Playing the roles of author and editor on your own book is a little like representing yourself in a court of law, so don't forget the old saying: "someone who represents himself in court has a fool for a client." Be sure that you understand when to step aside and take good advice from a trusted reader or proofreader.

Keystroke commands

If you are new to computers, you may not know that many of the features within Microsoft Word and other software programs can be reached through certain keystroke commands. Once you know the keystroke commands, it is generally quicker to use them instead of the mouse. Several important examples appear under Word's "Edit" menu. These usually use the "Control" key plus another character. You will find the Control key on the lower left corner of your keyboard. It may be marked "Ctrl". Here are some keystroke commands that you are likely to use frequently:

- Copy – "Control" plus the letter "c"¹³ copies whatever text you have selected. You can then move your cursor to another spot in the document and paste the text. You may wish to use this for long words or phrases that you intend to reuse.
- Cut – "Control" plus the letter "x" cuts the text you have selected so you can remove it and paste it in another spot.
- Paste – "Control" plus the letter "v" pastes the copied or cut text into the spot you have selected.
- Undo – "Control" plus the letter "z" is the undo command. If you don't like the text you just typed or the change you just made, hit "Control+z".

You also can access the cut, copy, and paste commands by right clicking with your mouse.

¹³ Press and hold down the "Control" key and the "c" key at the same time.

Outlining

“Start with an outline!” I’m sure that these words, spoken by innumerable high school English teachers, still ring in the ears of all writers. I didn’t start using outlines until I was much older, and the reason had nothing to do with my high school English teachers. I started using outlines because of Microsoft Word. It made it very easy to organize and edit my documents.

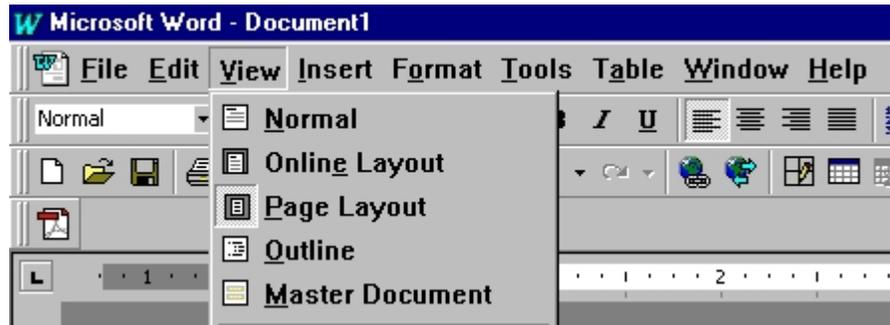
Are you old enough to have used a typewriter? Did you ever type up some assignment and then realize that it needed reorganizing? Did you get out your scissors, cut it up, tape it together again in the improved order, and then type it up again? This is what Word allows you to do, but it’s much easier, and you don’t need tape or scissors. I find that my writing benefits tremendously from the structure imposed by Word. You know those writers who won’t give up their Underwood typewriters? I feel the same way about writing with Microsoft Word.

Using the outline structure has another important benefit: it provides an easy framework for adding styles that greatly simplify the typesetting and formatting of a document. It also sets the stage for making a table of contents that is easily updated. (The table of contents entries are pulled directly from the headings in your document and assigned the appropriate page number.) I love the outlining feature and encourage people to use it. It surprises me how many people don’t know that it exists or understand the power that it wields.

Viewing a document in Outline mode

The outlining feature of Word is related to how the document is viewed. There are five different “View” settings in Microsoft Word: Normal, Online Layout, Page Layout, Outline, and Master Document. You can switch between these views by choosing the appropriate setting under the “View” Menu.

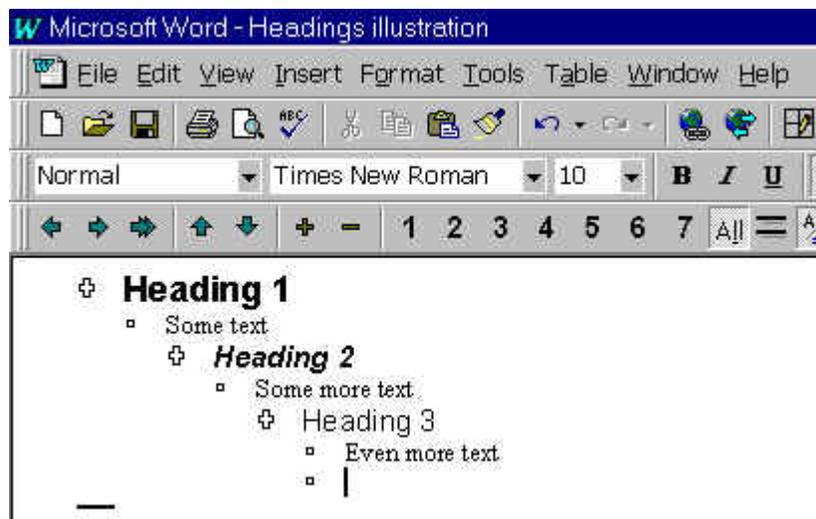
Figure 2 – Menu bar with the View options



I find that I use three of these views regularly:

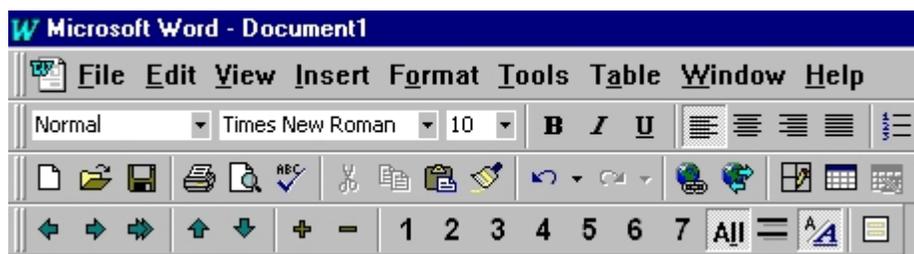
- Normal – “Normal” provides an easy way to view the document if you are just writing and don’t want to be bothered with seeing the edge of the page or the headers and footers.
- Page Layout – If you do want to see the page, then the best setting is “Page Layout”, which shows the page and all of the associated elements in position.
- Outline – “Outline” is the best setting if you want to view the hierarchical structure of the document, particularly if you intend to move or insert sections.

Figure 3 – A page viewed in Outline mode



There are two options when you are in Outline mode: “headings” and “body text”. The highest heading level is Level 1. In general, four or five heading levels should be more than enough for a document. Many documents require only two: Heading 1 for chapter titles and Heading 2 for subheads within chapters. The heading levels that you choose will be reflected in your table of contents later, so choose carefully. You can select whether all of the heading levels are reflected in the table of contents, but you cannot include items that are not headings in the table of contents.

Figure 4 – Outline menu bar (with arrows and numbers)



When you are in Outline mode, a menu with arrows and numbers will appear below the menu bar.¹⁴ The arrows on the left-hand side of the menu bar allow you to “promote” and “demote” headings, and also to change a heading to body text. Clicking the arrow that is furthest to the left raises a heading a level (i.e., promotes it). Pressing the next arrow brings it down a level (i.e., demotes it). Pressing the double arrow changes a text block from a heading to body text.¹⁵

If you have never used Outline mode, you should practice using it. Try moving heading levels up and down and changing them to body text. One nice feature of Word’s Outline mode is that you

¹⁴ If for some reason it isn’t there, go to “View” then “Toolbars” and be sure the box next to “Outlining” is checked.

¹⁵ A block of body text can be sentence or a paragraph. Every time that you hit “Enter” on the keyboard you create a new block of body text.

can easily collapse or expand what you see within the document. If you select the number 1 on the Outlining menu bar, you only see Level 1 heads. Select number 2 and you'll see Level 1 and 2 heads. Select "All" and you see everything, including both headings and body text. Notice that when you select the number 1, the collapsed headings have either a plus or a minus sign to their left. A plus sign indicates that there are more heads or body text beneath. A minus sign means that nothing is below. You can reveal the heads or text below a plus sign by double clicking on it. Body text blocks have a small square to their left. You can move a heading or body text block by left clicking on the plus/minus sign or square, and dragging it to the desired spot.

Outlining is a deceptively simple yet powerful feature. I use it constantly. It also reflects the power of keying your text into a computer. You only have to type it once and from then on you can copy it, paste it, and reformat it as much as you want.

Creating a table of contents

Say that you have created a document using Word's Outline mode and you want to create a table of contents. You have built the structure, now automatically creating the table of contents is simple. Insert your cursor where you want the table of contents to go and then go to "Index and Tables" under "Insert". Choose the "Table of contents" tab. You will see a selection of styles. Choose one and select the number of heading levels that you want reflected in the table of contents.¹⁶ Once you choose "Okay", Word will build a table of contents. If you don't like what it has done you can either delete it or "Undo" it. Once you have created a table of contents that you like, you can update it by placing the cursor within the table of contents and clicking the right button of

¹⁶ Selecting one level provides a very basic table of contents. If you include multiple heading levels, your table of contents may extend for several pages. You might choose to have a lengthy table of contents if you are not planning on having an index.

your mouse. You will have the option of updating the entire table of contents or just the page numbers.

I placed my table of contents under a bold heading that matched my chapter heading, but was not a true Level 1 heading. I did this so that the words “Table of Contents” would not appear in the table of contents. I’ll explain how to create such special styles in the chapter on “Design tasks”.

Creating an index

Creating an index is similar to creating a table of contents, but it requires a little more work. First, you must decide whether your book really requires an index. I think that most non-fiction books benefit from an index, even when there is an extensive table of contents. You can make it easy for a reader to find the names of people or products mentioned in your book. It also helps to unify concepts that may appear in multiple chapters. I think that most non-fiction writers would like to have an index, but are discouraged from doing so because creating an index was a very tedious manual process in the past. Microsoft Word can help you automate this process so that page references in your index can be updated very easily.

Once you have decided you want an index, you have to identify which words you want to appear in the index. There are automated and manual ways of doing this. To automate the process, you can create a “concordance”. A concordance is a file containing all of the words that you want indexed. To make a concordance file, open a new Word document and make a two-column table.¹⁷ The first column should include the word or words that you want to index. The second column should include the way you want the entries to appear in the index. For example:

¹⁷ The table feature in Word appears on the main Menu under the “Table” heading. Select “Table” and then “Insert Table”. The menu that appears next allows you to set the number of rows and columns. For practice, make a table with two columns and six rows.

Baker	Baker, Joseph F.
Birdwatching	Birdwatching
Joe Baker	Baker, Joseph F.
Joseph Baker	Baker, Joseph F.
Oriole	Baltimore Oriole
XYZ	XYZ, Incorporated

Select these words carefully, because once you apply the concordance to your document, it will assign index entries to these words whether it makes sense or not. For example, if there are any other members of the Baker family mentioned in the text, their names will also be indexed as “Baker, Joseph F.” because of the single word “Baker” in the concordance. Any discussion of a baker (who bakes) will also appear under “Baker, Joseph F.” in the index. Only use an individual name, such as “Baker”, if you are confident that it will not create confusing index references.

For this book, I selected words for the concordance that covered some of the important concepts and identified key products that were mentioned in the text. For example:

5D PDF	5D PDF Creator
Acrobat	Adobe Acrobat
Microsoft Word	Microsoft Word
Outline mode	Outlining
PostScript fonts	Fonts, PostScript
Print shop	Print shop
Proofread	Proofreading
Proofreader	Proofreading
Proofreading	Proofreading
TrueType	Fonts, TrueType

In general, you should create your index relatively late in the editorial process so that you can apply the concordance when the document is near completion. Save the concordance file in an easy-to-access location.

To apply the concordance file to your book, open the document you want to index. If the document is already open, then save it before proceeding (in case something goes wrong). Go to the “Insert” menu, click “Index and Tables”, and then click the “Index” tab. On the “Index” tab you will see a button named “AutoMark”. Press this button and locate the concordance file that you created. Select “Open”.

Once you have done this, Word will search the document for the selections in the first column of your concordance file. It will then use the text in the second column to create the index entry.

After you have applied the concordance you will be able to view the index entries that Word inserted in your document. There are two ways to make these entries visible. Press the Show/Hide button (see below), or go to the “Tools” menu and select “Options”. Under Options, select the “View” tab and activate the check box named “All” under “Non-printing characters”.

Figure 5 – The Show/Hide button is the symbol that looks like a backward “P” at the lower right corner of this illustration



Once you have made the non-printing characters visible you will be able to see the index entries (and spaces and carriage returns and line breaks¹⁸). Index entries appear within special curly

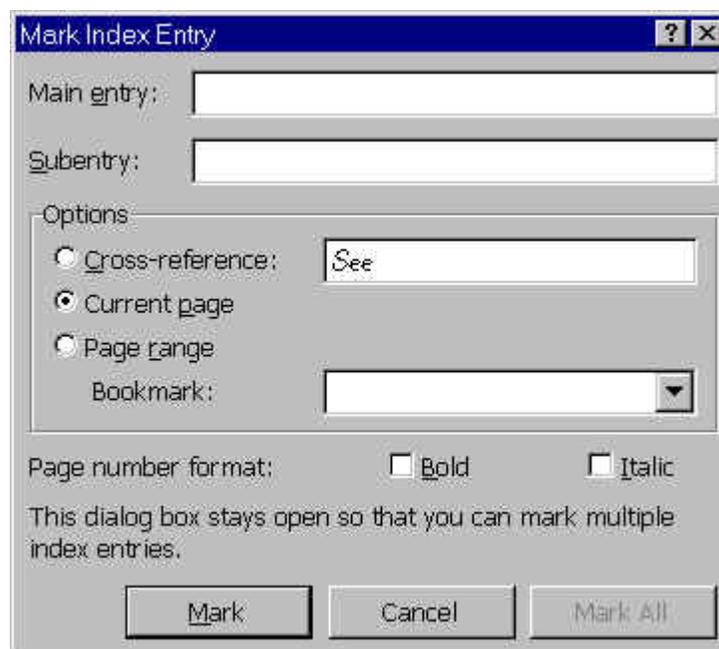
¹⁸ A space appears as a small dot. A carriage return, which you get by pressing the “Enter” key, looks like a backwards “P”. (Microsoft Word refers to this as a “paragraph mark”.) A line break looks like a small, curved leftward pointing arrow. You get a line break by holding the shift key while pressing the enter key.

brackets. Each entry contains the letters “XE” (for index entry) plus the exact wording the way it will appear in the index. You can use Word tools to place index entries manually within your document, but it is just as easy to cut and edit existing index entries. To do so, select the index entry, press “Control+c” (i.e., the keystroke command for copy), then move your cursor next to the desired spot (i.e., next to the words for the next index entry), and press “Control+v” (i.e., the keystroke command for paste). Edit the words in the index entry as required.

If you find that you want to automatically mark all the occurrences of a word that you forgot to include in your concordance, you should select “Mark All” in the “Mark Index Entry” dialog box.

I have found that using a concordance is a great time saver, and it also forces you to think about the exact wording of the entries in your index. However, there will always be additional words that you will want to reference. Therefore, you will almost certainly end up manually marking and editing index entries.

Figure 6 – The Mark Index Entry dialog box



You can also create cross-references in your index (for example, “See Birdwatching”). To do so, select the text you want to use as an index entry and then press “Alt+shift+x”. This will bring up the “Mark Index Entry” dialog box. Under “Options”, pick “Cross-reference”, enter the desired cross-reference text, and then press the “Mark” button.

Once you have marked the index entries, you can create the index. I placed my index under a Level 1 (i.e., chapter) heading. To do this, I typed “Index” and then set the heading to Level 1. Then I hit “Enter” and set the style of the next line to “Normal”. To insert the index, go to “Insert” and select “Index and Tables”. Word gives you multiple options for design choices. Later in this book, we will cover how you might adjust those style selections. For now, pick the “From Template” format, set the columns to two (this is a good choice for an 8½” x 11” page), and check the “Right align page numbers” box. Once you press “Okay”, Word will create an index.¹⁹ It will collect the index entries, sort them alphabetically, reference the correct page numbers, find and remove duplicate entries from the same page, and display the index in the document.

Once you have the index in the document, you can update it by right clicking with your mouse anywhere within the index. This brings up a dialog box that includes the message “Update field”. If you select this, the index will automatically update the entries and page numbers. Before you update the page numbers, be sure to make the non-printing characters invisible by pressing the “Show/Hide” button. If you don’t do this, the page numbering will be incorrect. When the index entries are visible on the page, their presence changes the length of the book. Most of the page references in index entries will be wrong if you update the index

¹⁹ In many cases, you may wish to view a menu without actually setting the process in motion. If you don’t want to create an index, simply press “Cancel” to leave the menu without doing anything.

when the index entries are visible. (The same holds true of the table of contents.)

Proofing the index

Once you have created an index, you will also need to proofread it. Important things to look out for include:

- Accurate page references – The page numbers will be wrong if you update the index while the index entries are visible because their presence changes the number of pages in the book.
- Duplicate entries – Duplicate entries crop up when there are differences between index entries that you intended to be identical. For example, “Baker, Joseph F.” and “Baker, Joseph F” will result in two index entries because they aren’t identical (the second is missing the period after the “F”). An additional space can cause similar problems. These mistakes frequently crop up when you are manually adding index entries. They can appear even if you use a concordance, if there are slight variances in the spelling of the words in the second column. Luckily, these kinds of errors are easy to spot and correct because they generally appear side by side in the index.
- Incorrect entries – If you are doing a significant amount of editing, or if you are adding a lot of index entries manually, it is quite easy to insert an incorrect index entry. For example, you copy an existing index entry and paste it elsewhere. Before you have a chance to edit the text to reflect the new entry, the phone rings. You forget to edit the text. That means that an incorrect reference to the first topic will be identified elsewhere. The best way to correct this is to do an index entry proofing cycle by reading the document with the non-printing characters visible. Then you can check that all of the index entries are appropriate.
- Weird entries – The format for an index entry does not handle words inserted in quotes very well. Word’s index entries are

identified within the curly brackets by quotes. Therefore, if your index entry includes a word in quotes, the index reference will most likely appear truncated. The only solution I have found is to omit any quotes from the index entries.

Word's Find/Replace feature

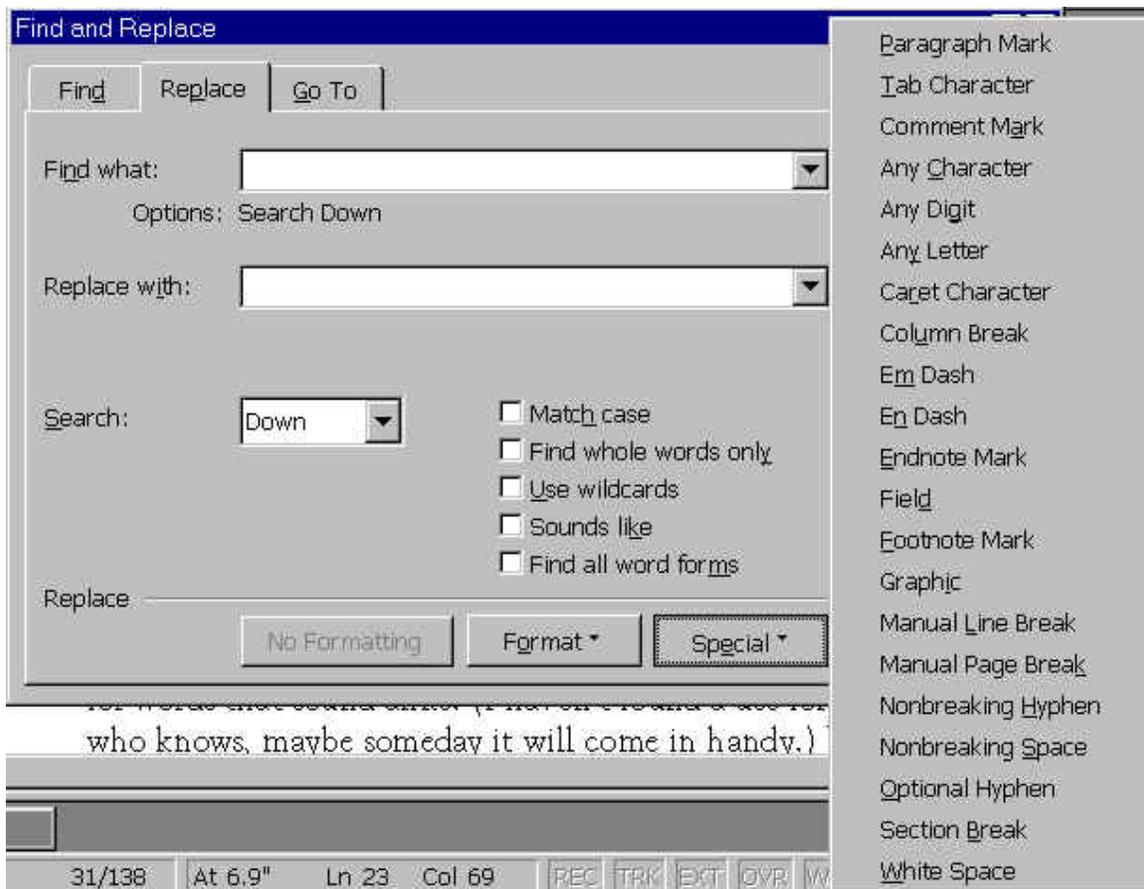
One of the most powerful tools in Word is its “Find/Replace” feature under “Edit”. I use this tool in a variety of ways. Sometimes I use it to navigate to an exact spot in a long document when I know a sequence of words that only appears there. I’ll use it to look for and delete double spaces or multiple carriage returns, which sometime crop up in a document.²⁰ In the same way, you can use it to search a document and replace all instances of something, for example, a product or person’s name that you want to assure is spelled consistently throughout the document. You can do this by searching for a sequence of letters that will appear uniquely in a certain word. For example, I might search for “icros” if I wanted to find all instances of “Microsoft” and I feared that I had capitalized the “s” in some instances.

I also use the Find/Replace feature to help find reminders that I have written to myself. I enclose these comments in square brackets, for example “[re-work]” or “[come back here]”. Square brackets are rarely used in a document, so it is easy for me to use the Find/Replace feature to locate them. I find that I use square brackets not only for reminders, but also to include ideas or subject areas that I want to write about, but don’t have time to flesh out right then. I have done this for so long that it is second nature to me now, and I never complete a document without searching first for open and close square brackets.

²⁰ For the record, I think it is appropriate to use a single space after a period when working on a computer. In addition, two or more carriage returns (i.e., paragraph marks) should never appear consecutively. As we will see in the “Design tasks” chapter, there are better ways to add space.

You can also use Find/Replace to look for format attributes like font or style. Therefore, Find/Replace can come in handy during formatting if you want to change a consistently mislabeled style to the appropriate style. To search for style and other format attributes, press the “More” button on the “Replace” menu. This brings up a “Format” button with multiple selections. Also on the “More” menu is a selection of frequently used characters, including line breaks and paragraph marks, that you may wish to include in a search.

Figure 7 – The Find and Replace menu with special characters



Find/Replace has advanced features that will even let you search for words that sound alike. (I haven’t found a use for this yet, but who knows, maybe someday it will come in handy.) You can even use “wildcards” to do searches for words that have certain characteristics in common. (For example, searching for “w?rd”

would return any instances of “ward” and “word”.) Chances are you are more likely to use the “Match case” feature, which assures that the words that are found exactly match the capitalization of your entry.

Whenever you use Find/Replace, save your file beforehand. The Replace feature is a powerful tool and you may unwittingly change more than you had planned. Also be extremely careful before selecting “Change All”. You want to be entirely sure that you are changing only the attribute you want, and not something else. For example, you might replace a character string in one word, but not realize that the same string appears in other words. I recently read a book in which the editor must have done a search for the letters “US” intending to replace them with U.S. (as an abbreviation for the United States). Unfortunately, the search also brought up every instance of the word “us” and as a result, the book contains sentences like: “Tell U.S. about it.”

You should be aware that the Find/Replace feature works best in “Normal” or “Page Layout” mode. In “Outline” mode, it will only search within the text that is currently displayed. Therefore, in “Outline” mode if you search for a word that you know is in the text of the document, you may not find it if you are only viewing the headings when you do the search.

Scanning an existing manuscript to convert it to text

Up to this point I’ve assumed that you are working from a manuscript that you already have or are creating as a Word file. However, you may have an older typewritten manuscript that you would like to convert into a Word file. A software technology called optical character recognition (OCR) can help you do this if you own or have access to a scanner. You simply provide the OCR software with a scanned image of the page, and it converts that page into computer readable text. Many scanners sold today come with very capable OCR software.

There are two major issues to consider regarding OCR:

- Scanning the pages
- Cleaning up the file after the OCR software is done

We'll look at these issues separately.

Scanning the pages

The first step is to scan the pages of your manuscript. Unless you have an automatic document feeder²¹ for your desktop scanner, this will be a time-consuming process because you will have to place each page manually on the scanner's glass. Chances are that your manuscript is 100 or more pages, so this could take many hours. Before you go through this tedious process, you should check to see if you might have access to an advanced digital device like certain fax machines or digital copiers. (Hewlett-Packard offers a product called "Digital Sender" that is effective for this purpose.) These devices are becoming popular among corporations because in addition to performing typical faxing or copying functions they also allow you to scan a document and turn it into a computer file that you can send to someone as an e-mail attachment. It goes without saying that these high-tech digital fax and copy machines have automatic document feeders, but more importantly the scan-to-file function can save you hours of manual scanning on your desktop scanner. These devices can scan much faster than your desktop scanner, and they also have the advantage that they run unattended.

Whether you use a desktop scanner or a digital copy/fax machine, you need to be sure that you provide the OCR software with the right amount of data. If you give it too little data, it will have difficulty distinguishing certain characters. If you give it too much data, you will be creating large scanned images that will be

²¹ Many copy machines have an automatic document feeder that takes individual sheets from a stack of paper and feeds them into the copier. This feature is much less common on home scanners.

difficult to transport from one computer to another, and whose size will bring no added benefit. In general, a scan resolution²² of 300 dots per inch is sufficient for OCR. Be sure that the scanner is not set to color or grayscale. The proper setting should be bitmap, which may also be called black & white, line art, or one bit. These settings will create a file of around one megabyte²³ in size for an 8½” x 11” page.

Once you have the scanned file, the OCR software can be used to convert it to text. The OCR software that I have used processes pages quite quickly, however the time it takes will depend on your computer and software as well as the number of pages being processed.

Cleaning up the file

OCR software can be very good, but it's not perfect. You will get a much better result if the text on the pages you are scanning is sharp and clear, and is not set in peculiar fonts (script, for example). The software will also have more difficulty interpreting manuscripts that are set with multiple columns or that contain a lot of tab settings. It is also very important to follow the software's guidelines for scanning resolution. However, even in the best of circumstances you will have to scrutinize the resulting file very carefully to find some subtle “wordos” and other errors. In many cases, a single letter or character is consistently mis-read, for example, “our” becomes “out” or “can't” becomes “can1t”. Luckily, you can use OCR's consistency to your advantage. As you identify mistakes, you can use Word's “Find/Replace” tool to locate other instances of these same mistakes.

Although OCR software is powerful, don't count on it to retain any of the formatting of the original manuscript, such as bold or

²² A discussion of scanning and scanner resolution is included in the “Design tasks” chapter.

²³ A discussion of file size is included in the “Design tasks” chapter.

italic fonts or varying type sizes. Once you have the text file, I recommend that you select it all (by pressing “Control+a”) and set it to one consistent style (perhaps Normal). You can then go about setting the proper heading and text levels so that you can apply the appropriate styles.²⁴ This would also be the process you would use if you are starting with a manuscript that was created on another computer system, that is, start with a text file, assign a consistent style in Word, and then begin formatting.

Here are some other typical OCR errors to look out for:

- Extras – OCR often inserts extra tabs, spaces, paragraph marks, or even dashes.
- Character substitution – OCR often cannot discern characters that look alike. For example, an apostrophe often becomes a lower case “l”, the numeral “1” may become a capital “I” or a lower case “l”, a lower case “q” may become a “g”, and a lower case “a” may become a “d”. These substitutions will depend on the design of the type (or typefaces) used in the original. These substitutions will be more likely if the original manuscript is in poor condition or contains broken or distorted text.
- Page numbers/headers/footers – If your manuscript contains page numbers, headers, or footers, chances are that the software will not recognize them. As a result, you will find page numbers inserted at regular intervals in the resulting file. Since you will be repaginating the text, these page numbers are now worthless and you will have to remove them from your document.
- Dropped spaces – You will occasionally find merged words because the OCR software did not recognize the space between them.

²⁴ The “Design tasks” chapter includes additional information on Microsoft Word styles.

You will catch many of these errors with a global spell and grammar check, but you will also need to read the resulting file very closely.

Is it worth it?

Although OCR can be a great tool, you still need to consider whether using it is to your advantage. If you have a long manuscript and you do not have access to a scanner with an automatic document feeder, you will end up spending a lot of time placing documents on your scanner's glass and pressing the scan button. In addition, you will have to allocate some time to cleaning up the resulting file after the OCR process. Realistically, you may be better off retyping the manuscript (or paying someone else to do it if you are not a fast typist). However, whether you use OCR or retype the manuscript, you will still need to pay close attention to any mistakes that might have crept into the document during the conversion process.

