

SAVING YOUR *WORD 2000* DOCUMENTS TO HTML

USING *WORD 2000* AS A WEB PAGE EDITOR

The transition from using *Word97* to *Word 2000* is perhaps most noticeable when using *Word 2000* to convert your documents to HTML for posting on the Web. *Word 2000* will allow you to save much of your existing document easily to Web (HTML) format, retaining many elements (including indented paragraphs, tabs, and others) that simply would not display properly in a Web page using earlier versions of *Word*. *Word 2000* uses various page-generation techniques, including XML (extensible markup language) and Cascading Style Sheets (CSS), to allow users to easily switch between editing their *Word* and their Web-based documents with little or no difference in formatting.

POTENTIAL PROBLEMS USING *WORD 2000* AS A WEB PAGE EDITOR

Word 2000 uses many Office-specific markup tags in its documents when saving them to HTML format. The end result may be that users who do not have version 5.x or above of Microsoft's *Internet Explorer* may **NOT** see the document the same way as you had intended! In addition, the Office-specific markup that is inserted when converting *Word 2000* documents to HTML can lead to dramatic increases in the file size--and therefore the amount of time users must spend waiting for a document to load. This trade-off is necessary in order to have the document on the Web appear as close as possible to your original document.

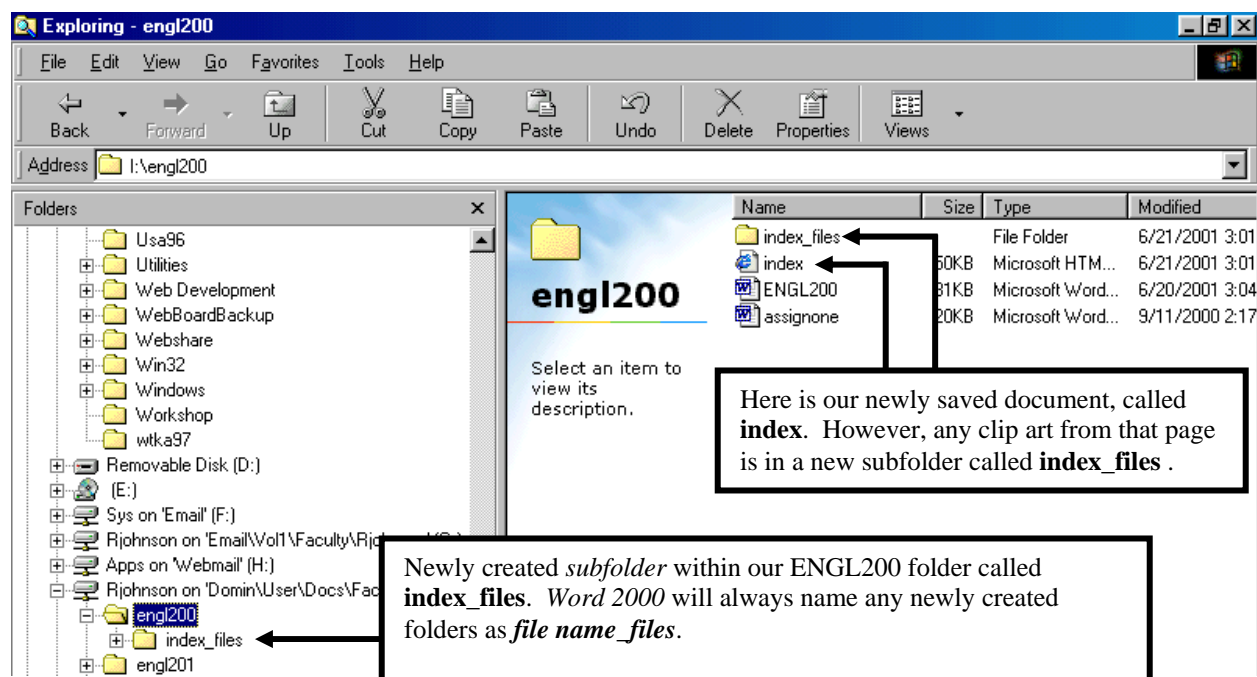
As an example of the increase in file size, we created a basic three-page document that included two tables and one piece of clip art. In *Word97*, using the *Save as HTML* command, this document was **8 KB** in file size, with the clip art adding an additional 10 KB. In *Word 2000*, using the *Save as Web Page* command, this same document became **36 KB** in file size. In addition, a new subfolder was automatically created on the Web server storing the image and accompanying XML files that added an additional 10KB.

Because of *Word 2000*'s extensive use of XML in addition to HTML, Web browsers that do not support the latest implementation of XML may not display your document properly. This will not affect basic text, but any additional formatting (indents, tabs) will be affected. Later in this document, we discuss ways to ensure backward-compatibility with Web browsers--though at the cost of some document formatting.

Generally speaking, if your potential viewers have the latest version of *Internet Explorer*, they should encounter no difficulties--though they may have to download some add-ons for proper viewing.

IMPORTANT INFORMATION REGARDING FOLDERS

Word 2000 allows you to convert any existing *Word* document to HTML very easily, as will be explained below. However, if you have *any* clip art, photos, line art, backgrounds, or illustrations of any kind on your document, *Word 2000* places all of this material in a **new folder created by *Word* automatically**. In the example outlined below, we create a document for a class and save it using *Word 2000*'s **Save as Web Page** feature directly to our Dominican academic server drive (the **I:** drive). We name this document **INDEX** and save it within the ENGL200 folder. Because this document has some clip art in it, note what happens after saving it when looking at the content to our **I:** drive:



This may lead to difficulties when you choose to copy these files elsewhere, as you *must* remember to copy this newly created subfolder, too. For example, if you originally saved this file in your **My Documents** folder and wanted to copy it to your academic server (**I:**) drive, you would have to remember to also copy the subfolder, or else the clip art and other illustrative material would not appear on your Web page.

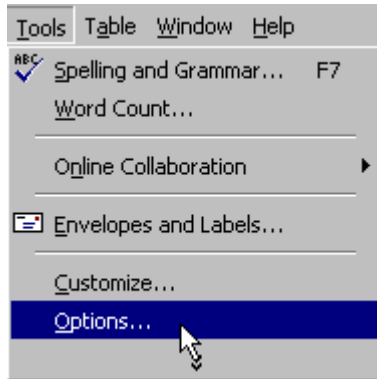
Be sure to see the section on **IMAGES** later in this document if you are using clip art, or any other type of image on your Web page, for information on how *Word 2000* converts these images.

The next page details how to set *Word 2000* so that everything is kept in the same folder as where your document is kept.

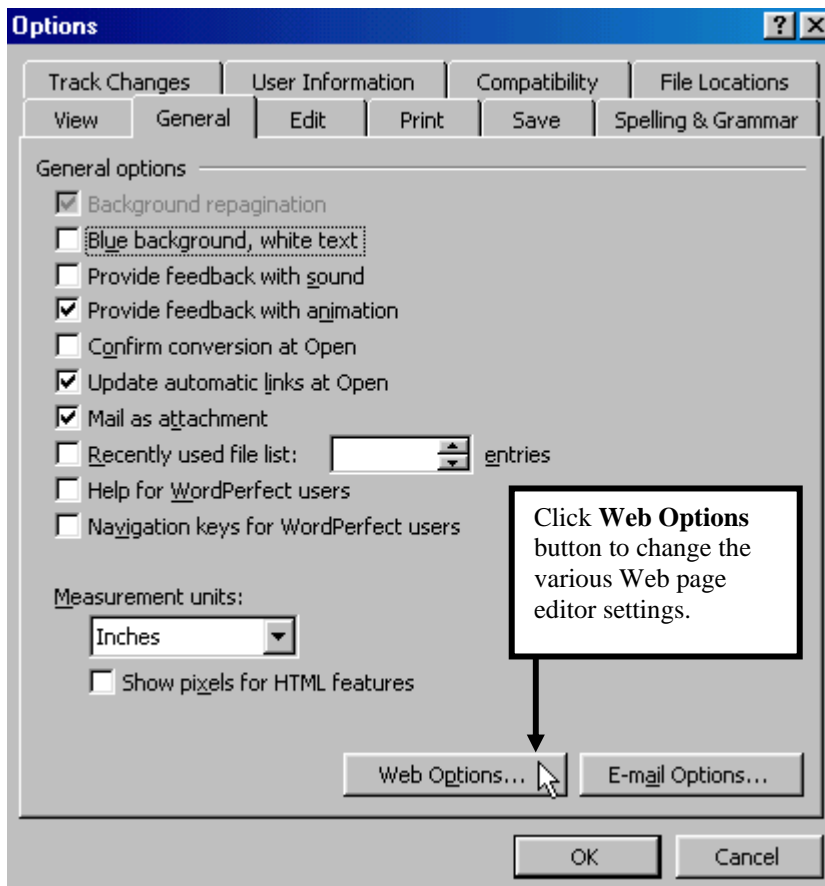
Keeping Everything in One Folder

To force *Word's* HTML editor to keep *everything* in a single folder, you must change settings:

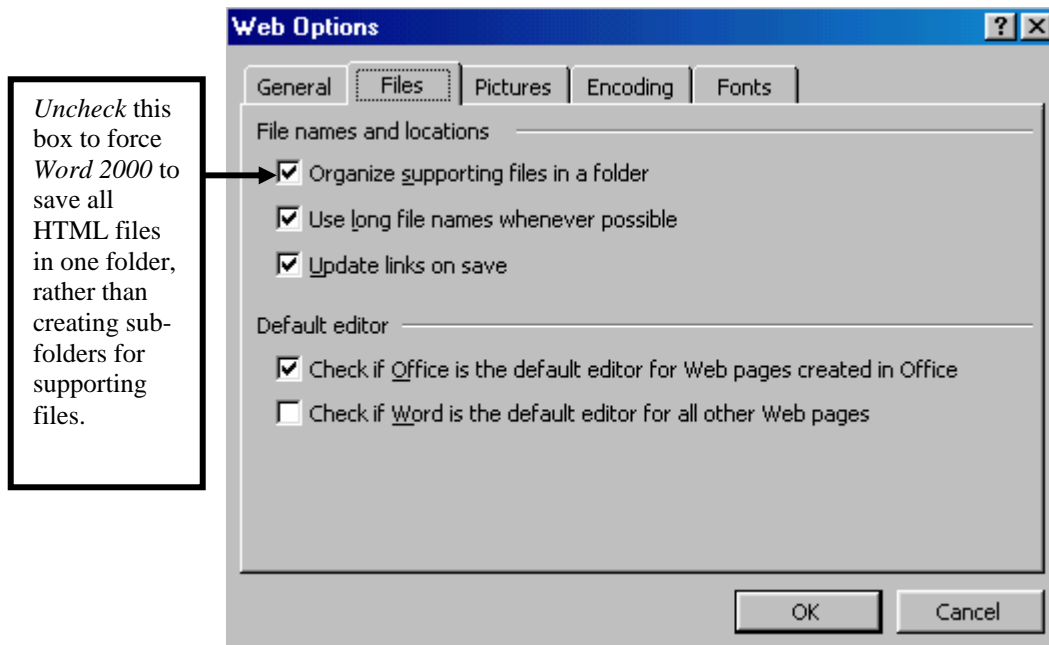
1. Click the **T**ools pull-down menu, then click **O**ptions:



2. In the next menu, make sure the **G**eneral tab is selected, and click the **W**eb Options button:



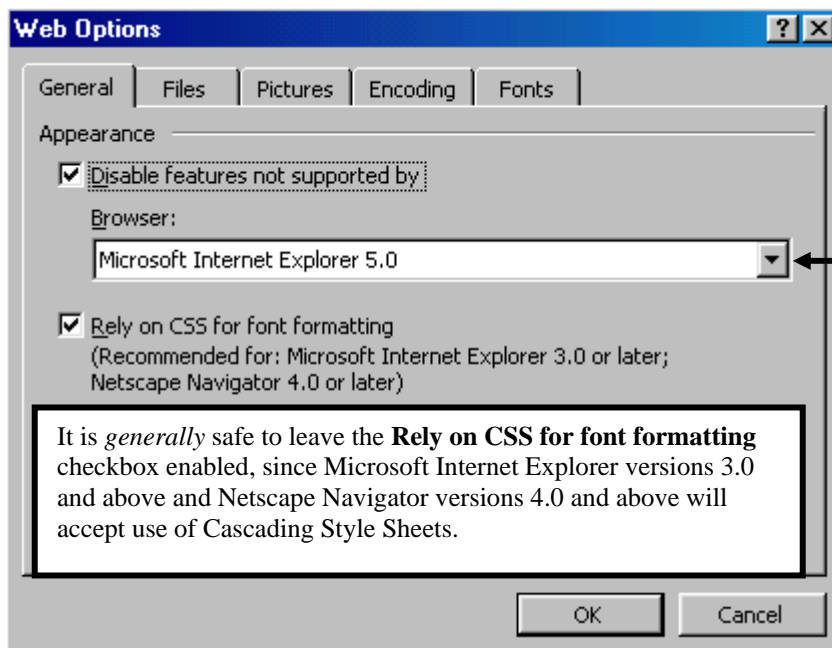
3. In the following dialog box, click on the tab labeled **Files**, and then *uncheck* the checkbox labeled **Organize supporting files in a folder**:



IMPORTANT INFORMATION REGARDING BROWSER COMPATIBILITY

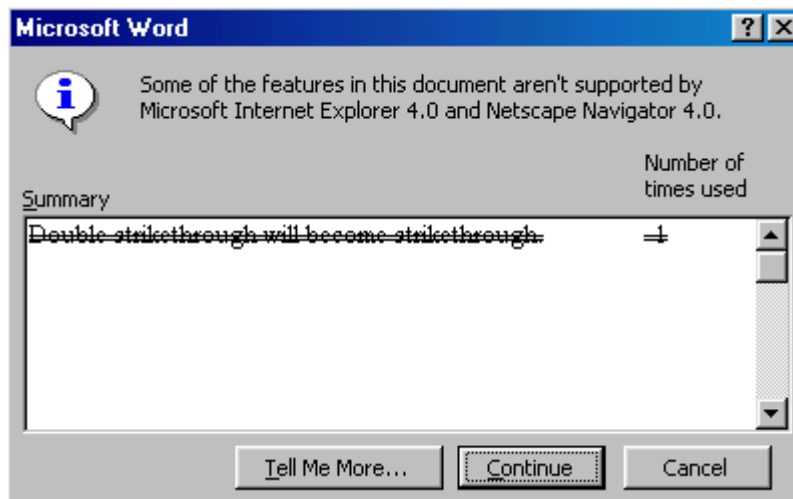
By default, *Word 2000* is configured to save Web pages with formatting that will be acceptable to *Internet Explorer* versions 5.0 and above. If you believe that your users will be accessing your page with earlier versions or with certain versions of Netscape Communicator, you should be sure to set options for *Word* to alert you to items *not* viewable in these earlier versions of browsers.

To check these, go to the **T**ools pull-down menu, and select **O**ptions, as outlined above. From the ensuing dialog box, make sure the **G**eneral tab is selected, and click the **W**eb **O**ptions button. (See steps one and two on preceding page for screen shots of these steps.) In the **W**eb **O**ptions dialog box, click the **G**eneral tab, as shown on the next page:



Click this down arrow to set the **Disable features not supported by** box to **Microsoft Internet Explorer 4.0 and Netscape Navigator 4.0** instead of the default setting of Internet Explorer 5.0 if you would like to be warned of possible browser compatibility problems when saving your Web pages.

In this dialog box, you want to leave enabled the checkbox labeled **Disable features not supported by . . .** in order to receive error messages when saving a file if a certain feature is not supported by the browser selected. For example, if we set **Disable features not supported by** to Microsoft Internet Explorer 4.0 and Netscape Navigator 4.0, we will receive the following error message when saving the file if we used the double-strikethrough font feature somewhere in the document:

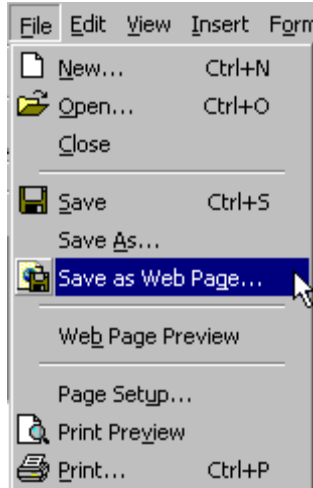


This warning message would not have appeared if we cleared the **Disable features not supported by** check box, so it is best to leave it checked.

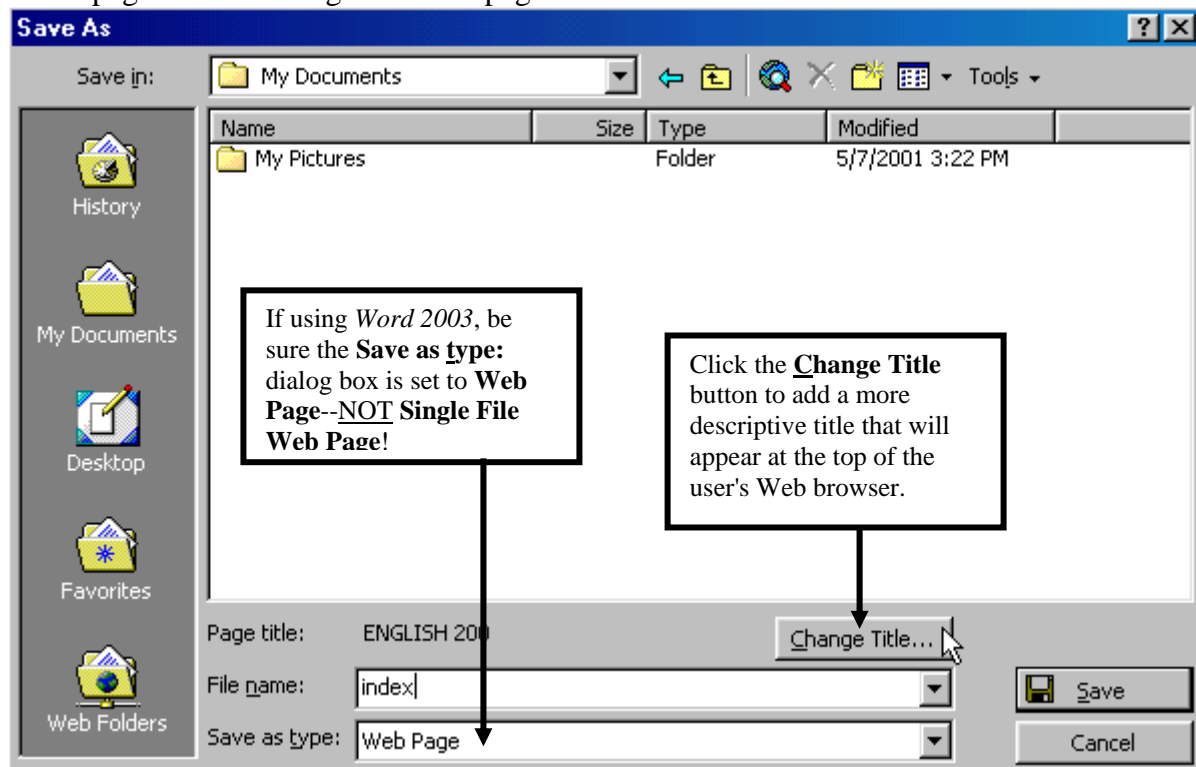
If all this talk about browser support, XML, Cascading Style Sheets, etc., is a bit confusing, the good news is that *Word 2000's* default settings will match the Web browsers installed on all of Dominican's computer labs, which have the latest version of *Internet Explorer*. Compatibility problems may arise, however, for home users without the latest version of the browser. If you plan on using clip art, be sure to check the section on **IMAGES** later in this document!

CONVERTING AN EXISTING WORD DOCUMENT TO HTML

If your document is already in *Word* format, open the document in *Word*. To convert to HTML, go to the **File** pull-down menu, and click the **Save as Web Page** command to convert the document:



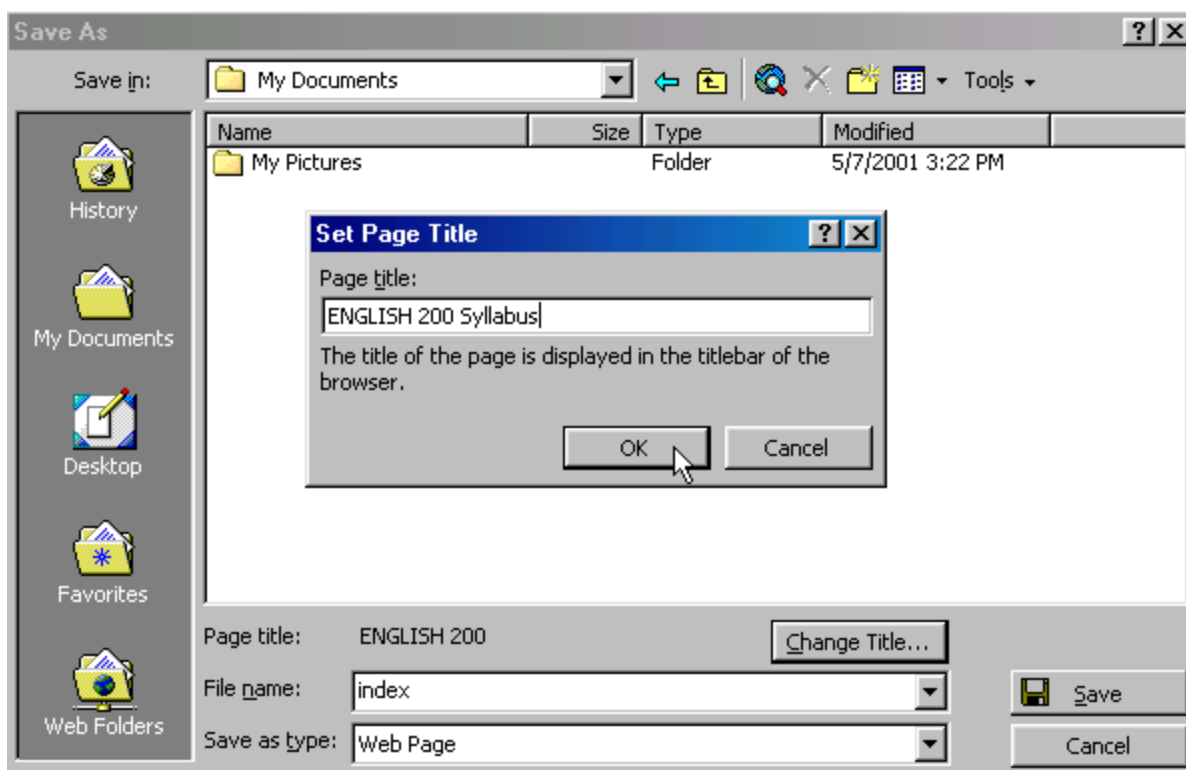
You will then see the following dialog box. For users of *Word 2003*, be **sure** to set the file type to web page and **NOT** single file web page:



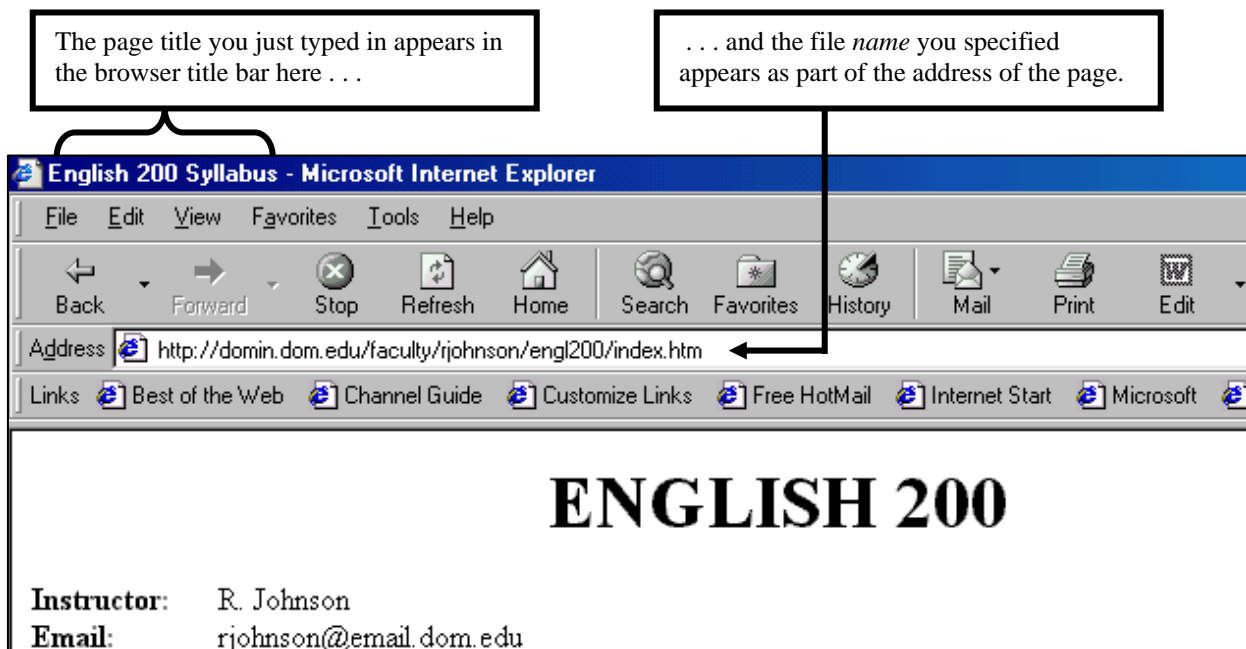
If your document has illustrations or clip art of *any* kind, be sure to read the above section **IMPORTANT INFORMATION REGARDING FOLDERS** for where this clip art may wind up, and the section on **IMAGES** later in the document for how they are formatted!

If you are an **instructor** at Dominican and saving a syllabus, you should name your syllabus INDEX and save it to the proper directory on the DOMIN server (your **I:** drive), as explained in the handout *Steps for Loading Your Syllabus onto the DOMIN Server*. If you are a **student** and creating a home page, you should name your file INDEX and save it to your directory on the DOMIN server (your **I:** drive). The reason for this is to ensure that users can not type your Web address into their browsers and gain access to other files on your **I:** drive. Using a file name of INDEX will mean, unless they know the file name, their browsers will go to the INDEX file by default.



On the previous screen shot, notice there is a box labelled **Change Title...** in the dialog box. Clicking on this will bring up the following screen:



This screen allows you to type in the title of the page that appears in a browser window. When a user opens your page in a browser, the title bar on the browser will show this title, rather than the file name you saved it under, as shown in the screen shot on the following page:



You will likely have to make a few changes in formatting, as outlined in the **Tips on Text Formatting** section below. You may also find it easier just to create a Web document from scratch in *Word 2000*--particularly if, on your old *Word* document, you used the **Tab** key several times. There is no such thing as **Tab** in HTML, though thanks to *Word 2000*'s implementation of using XML this is no longer as big a problem as it was in previous versions of *Word*.

However, it is still strongly recommended you place tabbed text into tables since there is no guarantee the text will line up properly the next "tab" stop over. Alternatively, use the increase indent or decrease indent buttons in *Word*   if you plan on indenting the text just one "tab stop" over.

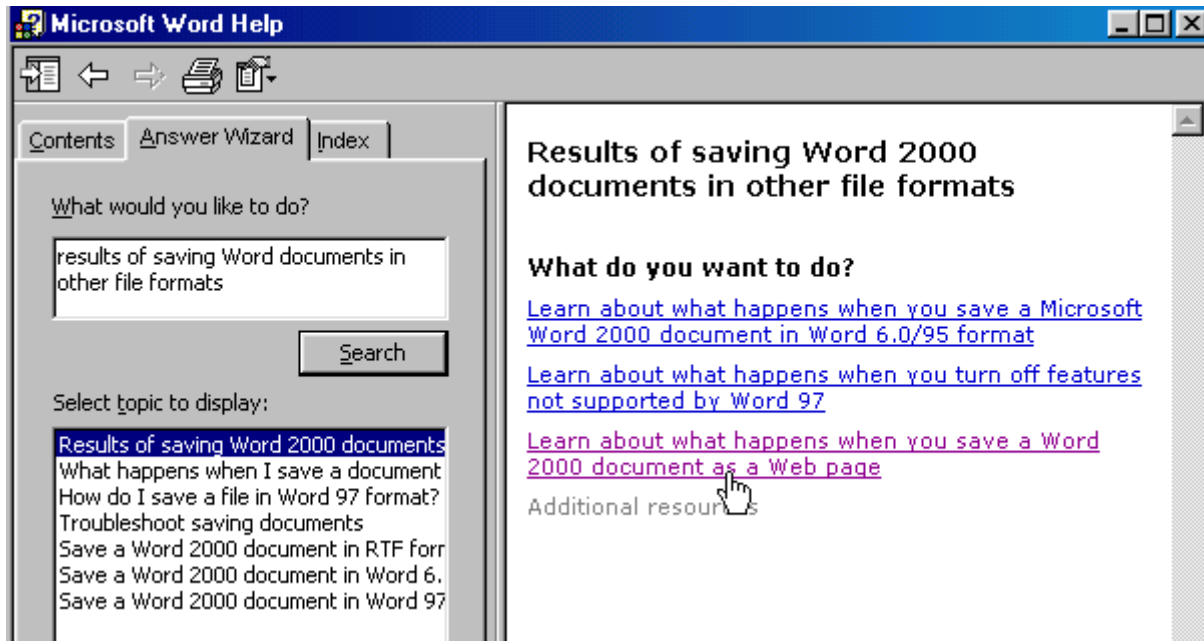
Again, **put tabbed or columned text into TABLES** if you plan on converting from *Word* to HTML. Tables are recognized by all versions of browsers and will guarantee maximum backward compatibility.

EFFECTS LOST TO HTML CONVERSION

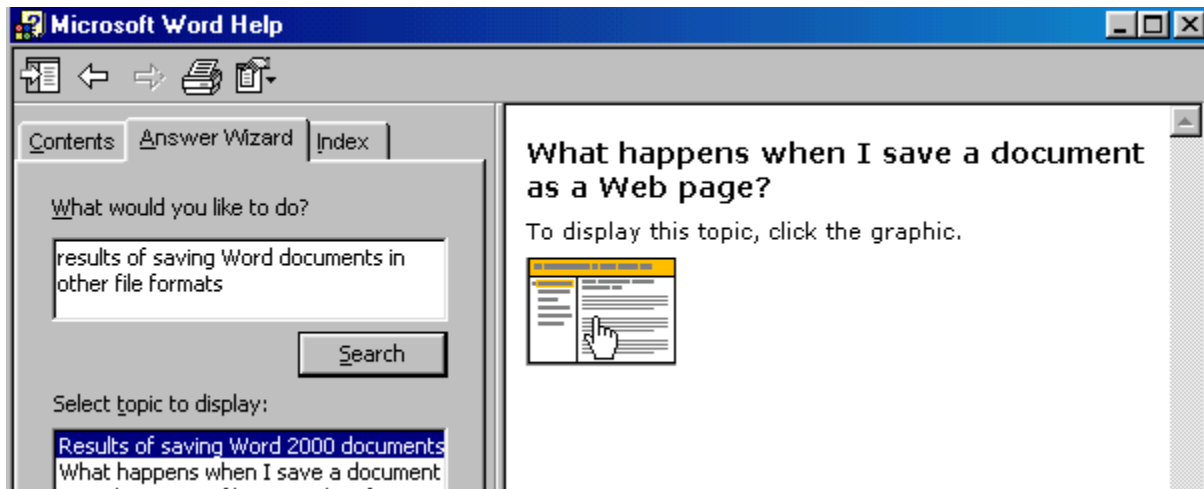
Even with *Word*'s extensive use of XML and CSS to convert documents to Web pages that will be viewable by the latest browsers, there are still some elements that will be lost in the HTML conversion. For a complete listing, go to *Word*'s help facility and type in the following as either a keyword search in the Office Assistant's help box or in the Answer Wizard portion of the help screen:

Results of saving Word documents in other file formats

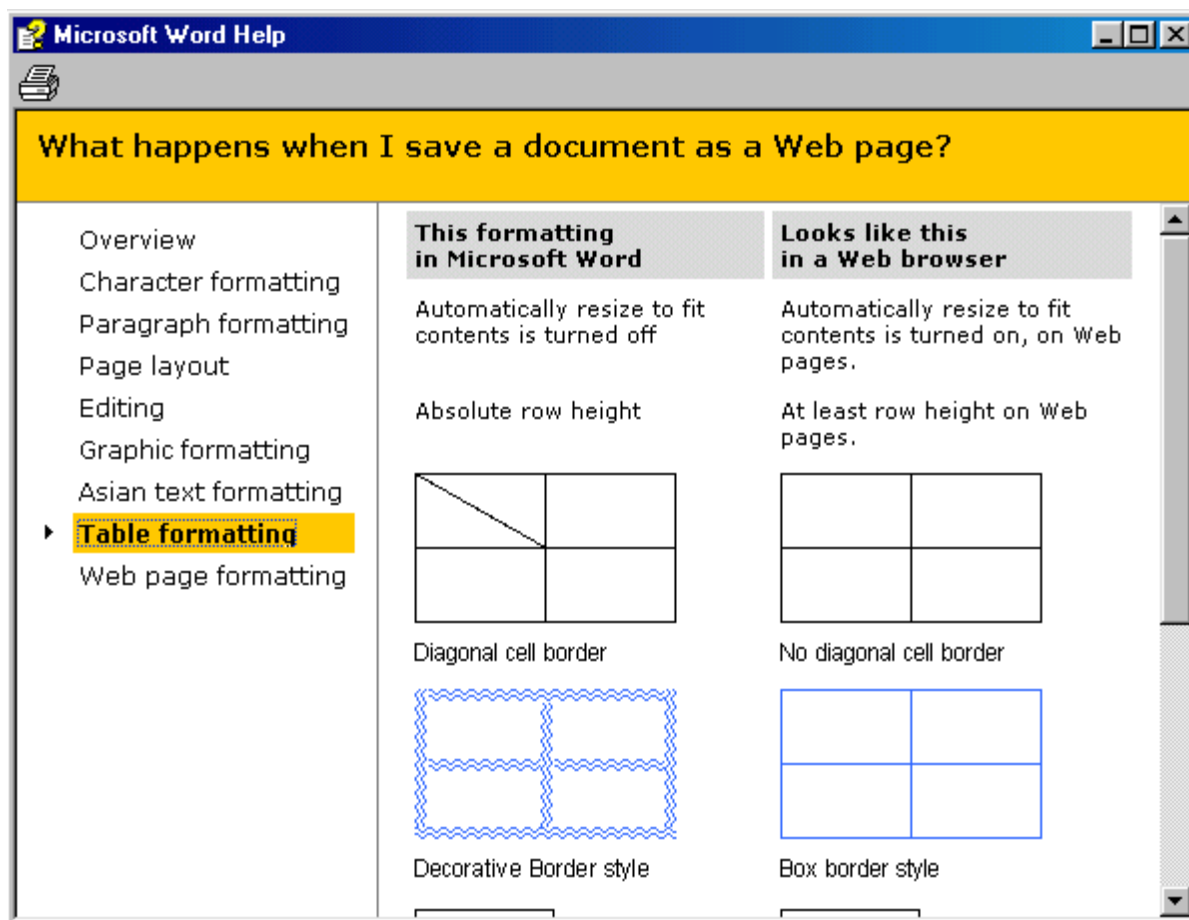
You will get a list of possible topics, including one that should match exactly what you just typed. Click on that article title and, at this point, your screen should look like this:



Next, on the right side of the screen, click the article titled **Learn about what happens when you save a Word 2000 document as a Web page**. This will lead to a help screen prompting you to click on a graphic:



Click on the graphic, as prompted, and you will (finally) get to a screen where you can click on a format option on the left, and *Word* will explain what happens in HTML conversion on the right:



If you used any pictures (such as clip art), you may have to place the graphics within tables to retain their placement on your page (refer to section on tables below). Any drawn objects (text boxes, shapes) will be converted--but only if you enabled the option to use VML (see section below on **How Word 2000 Converts Graphics**). Otherwise, it's safest to convert any text boxes into tables.

Any reference you make to **page numbers** on your document should be eliminated. (Remember that a Web page has no page numbers!) Consider using bookmarks within the document (see section below on **Creating Hyperlinks Between Locations Within Your Document**), or creating a separate page under a different file name (such as CALENDAR.HTM) and creating a hyperlink to that page from your main syllabus page (INDEX.HTM).

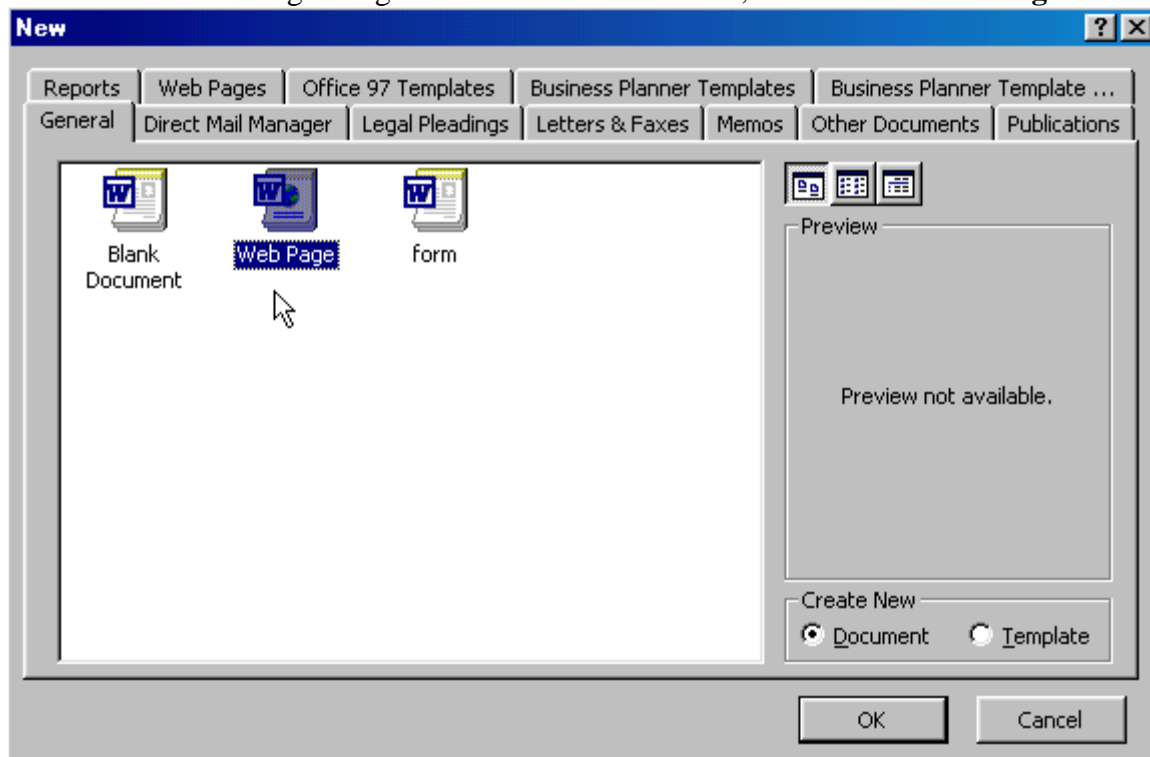
If your document has illustrations or clip art of *any* kind, be sure to read the section **Important Information Regarding Folders** for where this clip art may wind up, and the section **How Word 2000 Converts Graphics** later in this document.

If some of your users cannot see your graphics, be sure to read the section **How Word 2000 Converts Graphics** later in this document.

CREATING A NEW HTML DOCUMENT IN WORD

If re-formatting an existing *Word* document as HTML proves more trouble than it's worth, you may wish to consider creating your documents from scratch as HTML documents *first*, and then convert them to *Word* documents afterwards. Particularly with *Word 2000*, this is easily done with minimal appearance changes to your document. (It is, in fact, the main reason why XML and CSS are used so extensively in *Word's* Web documents.)

To start a *Word* document as a blank, HTML formatted document, click **File . . . New . .** and from the following dialog box under the **General** tab, double-click **Web Page**:



Make sure that in the **Create New** section in the lower right-hand corner, that **Document** is clicked before clicking **OK**.

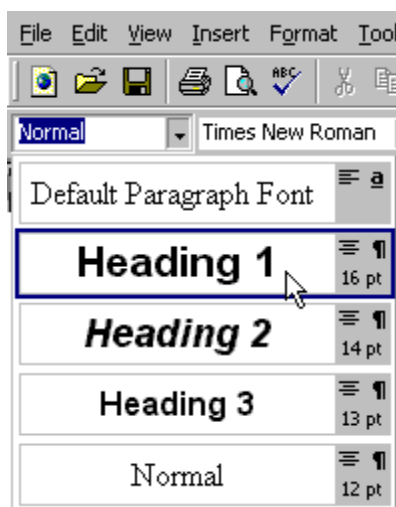
The rest of this document will assume you are using *Word* after doing an initial File..Save as Web page command or have selected Web Page from the dialog box shown above!

TIPS ON TEXT FORMATTING AND PAGE LAYOUT

There are several factors to consider when editing an HTML document. Not all of the elements from regular text editing in *Word* will necessarily translate into HTML. Below are some factors to consider.

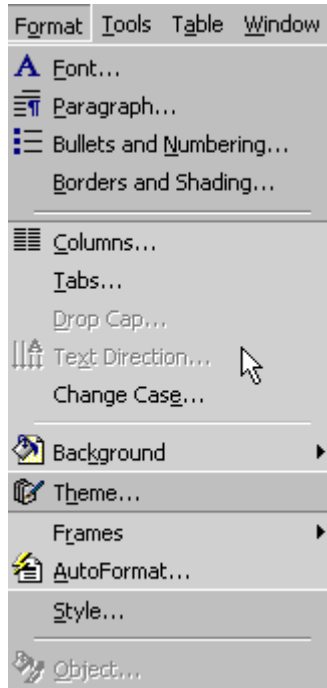
Fonts and Background

In addition to clicking on the traditional font sizes, you may want to take advantage of the **Style** pull-down menus. These insert various HTML headings with pre-determined font sizes into your document, and are listed within your toolbar to the far left, where you usually see the word **Normal**:

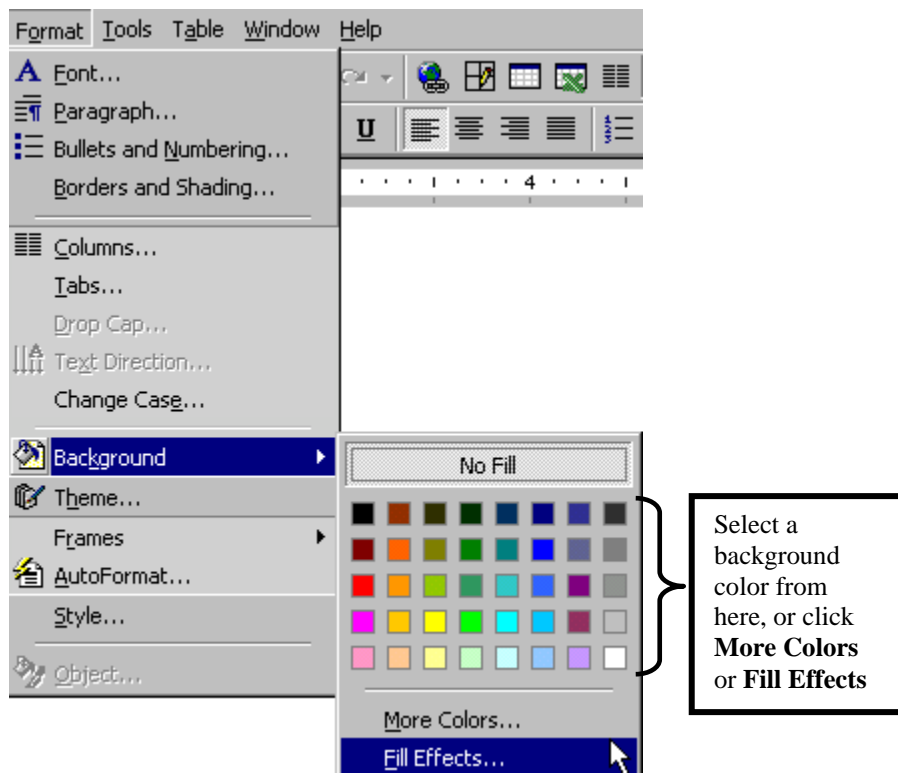


Because hyperlinks on the Web are underlined, it is best to use italicized type or bold face type when citing the titles of books or journals rather than underlining them.

To change the appearance of your page, you may want to change the color of your font or of the page's background color. Both options are available under the **Format** pull-down menu, as shown on the next page:

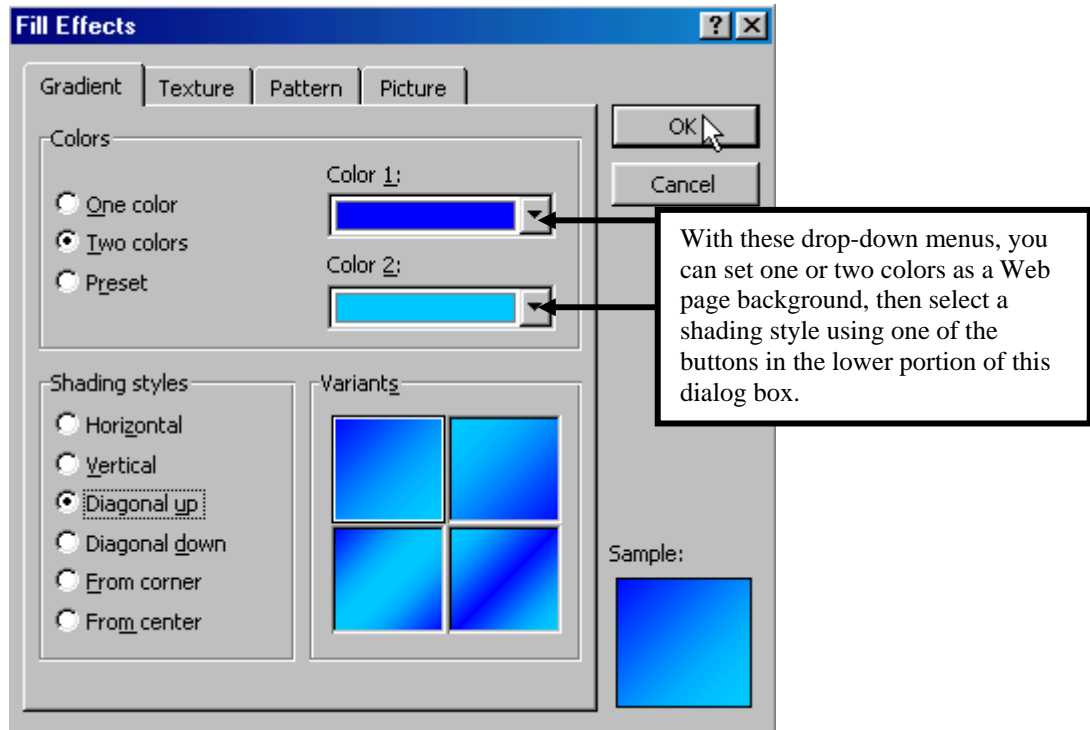


The **Font** choice from the pull-down menu above will change the color of your text (you will have to select the text with your mouse first if you've already typed it), while the **Background** option will enable you to select a color for the background of the document:

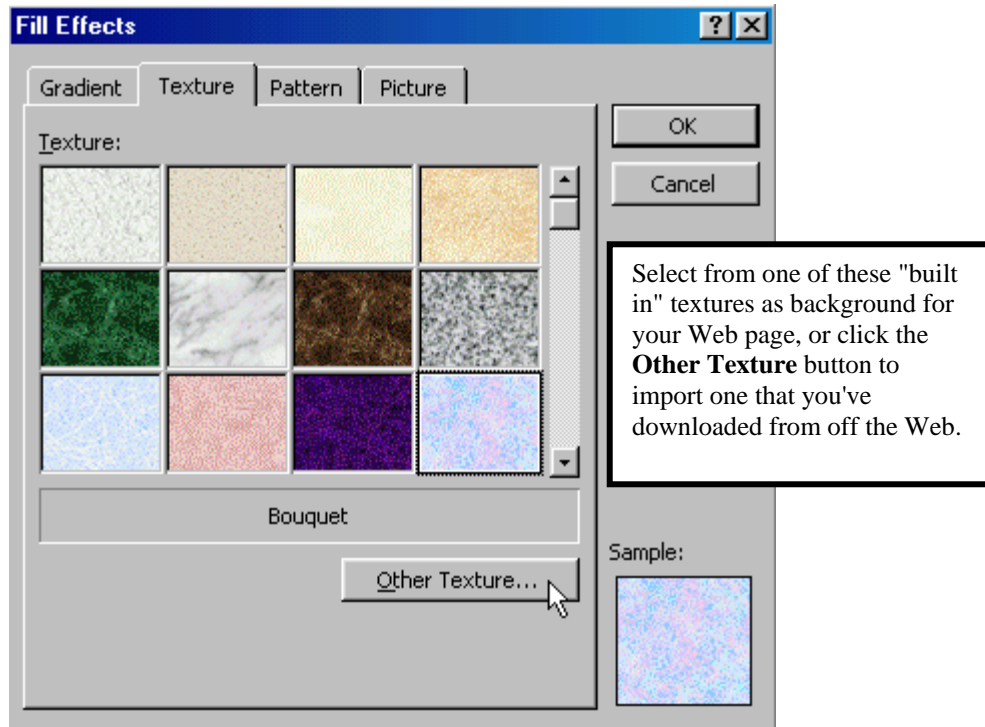


The **More Colors** option from the above menu will open a window featuring additional color choices.

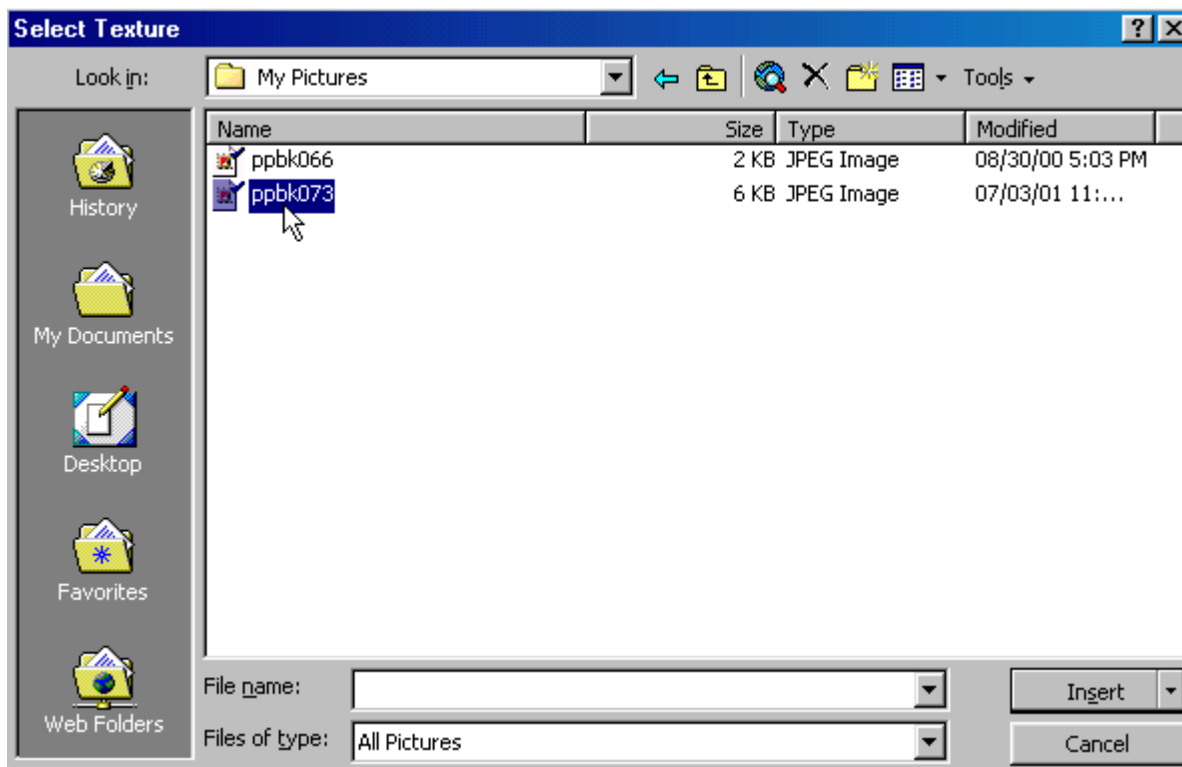
The **Fill Effects** option will lead to this screen:




In the above dialog box, clicking on the **Texture** tab will lead to this screen:



In the above box, to import a different background (such as one that you have downloaded from the Web) click **Other Texture**, and you will see a dialog box where you can select which graphic you would like to use as your background:



You can click the **Up One Level** icon in the box above () to go to a different directory if you have to, or the down arrow next to the **Look in** window. After clicking the **Insert** button, the image you selected should appear as the background.

Tables

Remember these three things when typing text you want to appear in a specific place in your document--such as text you want in columns:

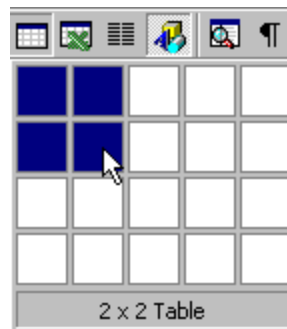
1. Place text in tables
2. Place text in tables, and
3. Place text in tables.

The best way to ensure that your text will stay formatted a certain way (in columns, with white space, etc.) is to place the text within a table. If you are an instructor working on a syllabus, it has been found through experience that common syllabus elements such as reading lists or

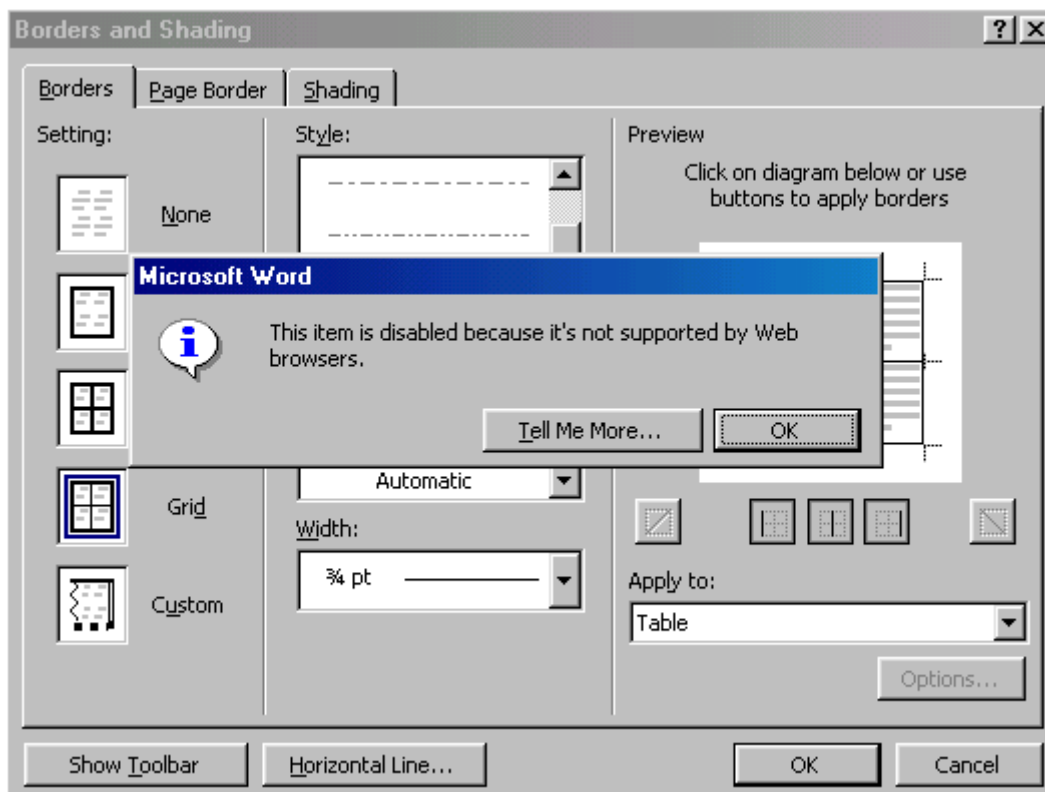
course calendars are best placed within tables. If you want to place text next to a graphic image, it's best to place those within tables, too. Although *Word 2000* uses extensive XML features, text wrapped around an image is always a potential problem--particularly for users with older browsers.

Although the latest versions of Web browsers are now recognizing additional word processing formatting (such as tabs), such features are not standard and will often fail to look right if the user adjusts font size in his/her Web browser.

To create a table, click on the **Insert Table** icon in the standard toolbar:

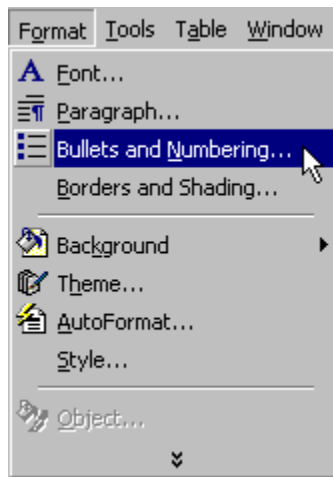


Or, click the **Table** pull-down menu, then **Insert**, and from the following menu click **Table**. You can choose whether or not to show table borders by clicking anywhere within the table, then clicking on the **Format** pull-down menu, then **Borders**. Note that you do not have the same "decorative" elements with table borders while using *Word's* HTML editor as you do with regular *Word* documents, and may find yourself getting the following message box:

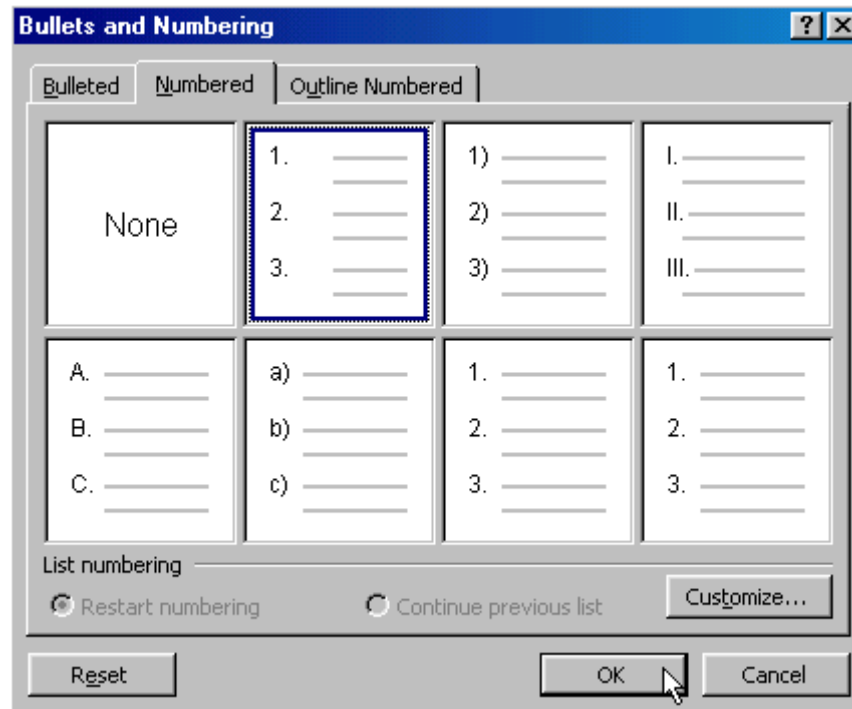


Lists

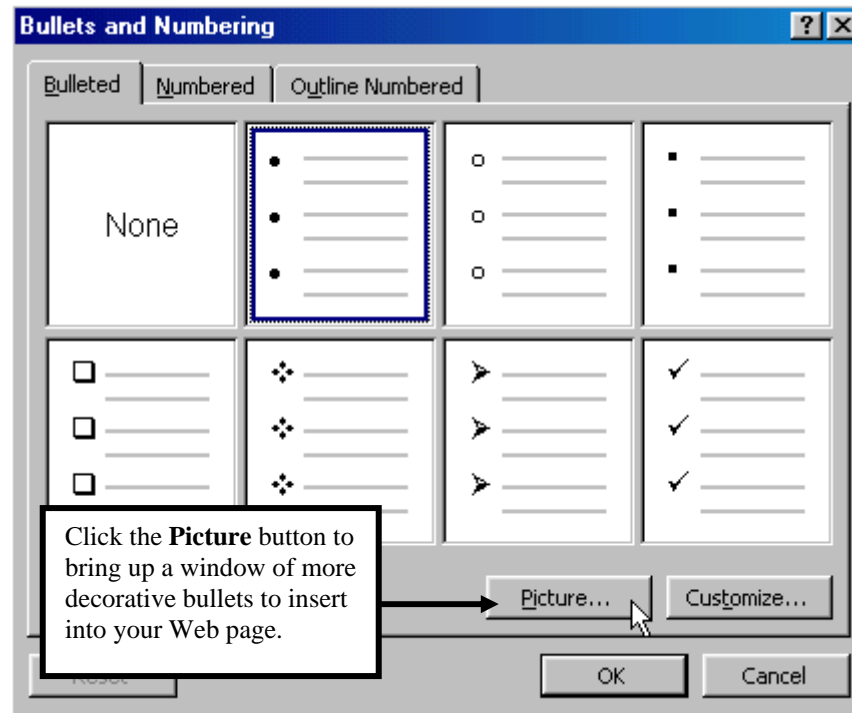
When using any sort of numbered list, it is best to use *Word's* numbered list feature rather than number lines yourself. To start a numbered list, click the **Format** pull-down menu, then **Bullets and Numbering**:



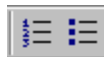
From the next box, click the **Numbered** tab and click in the appropriate box:



You can also click the **Bulleted** tab if you prefer a bulleted list:



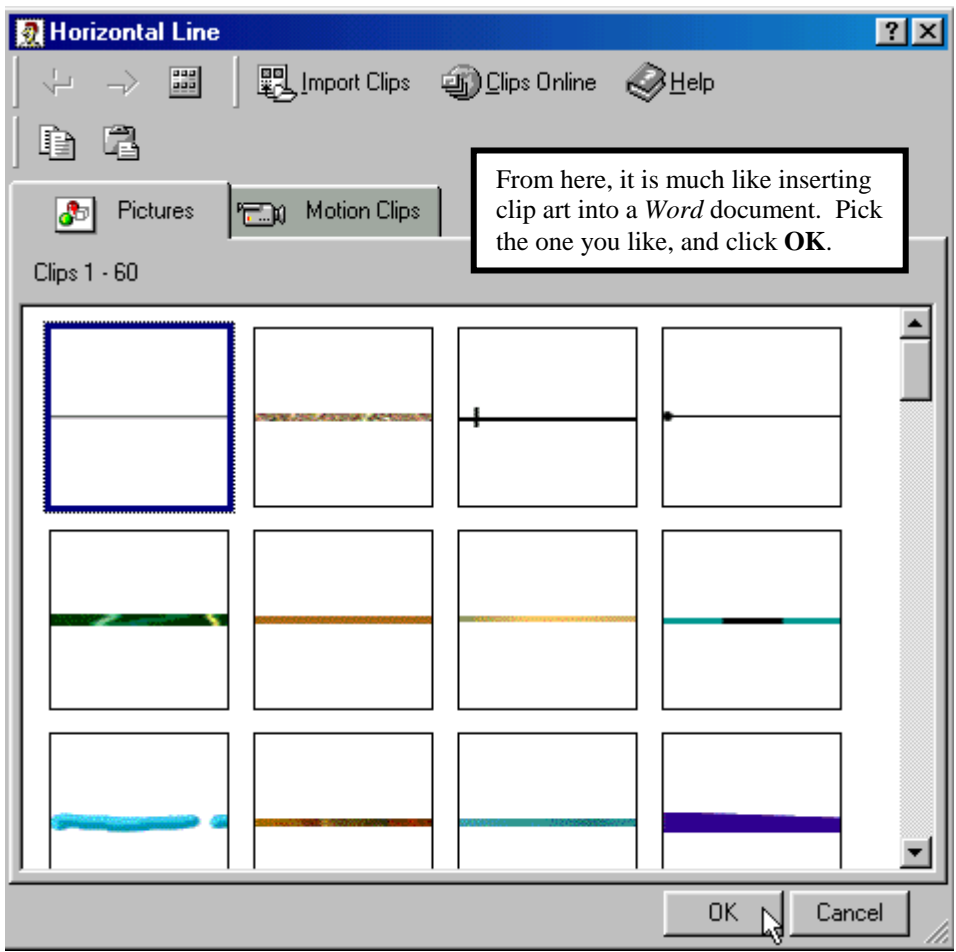
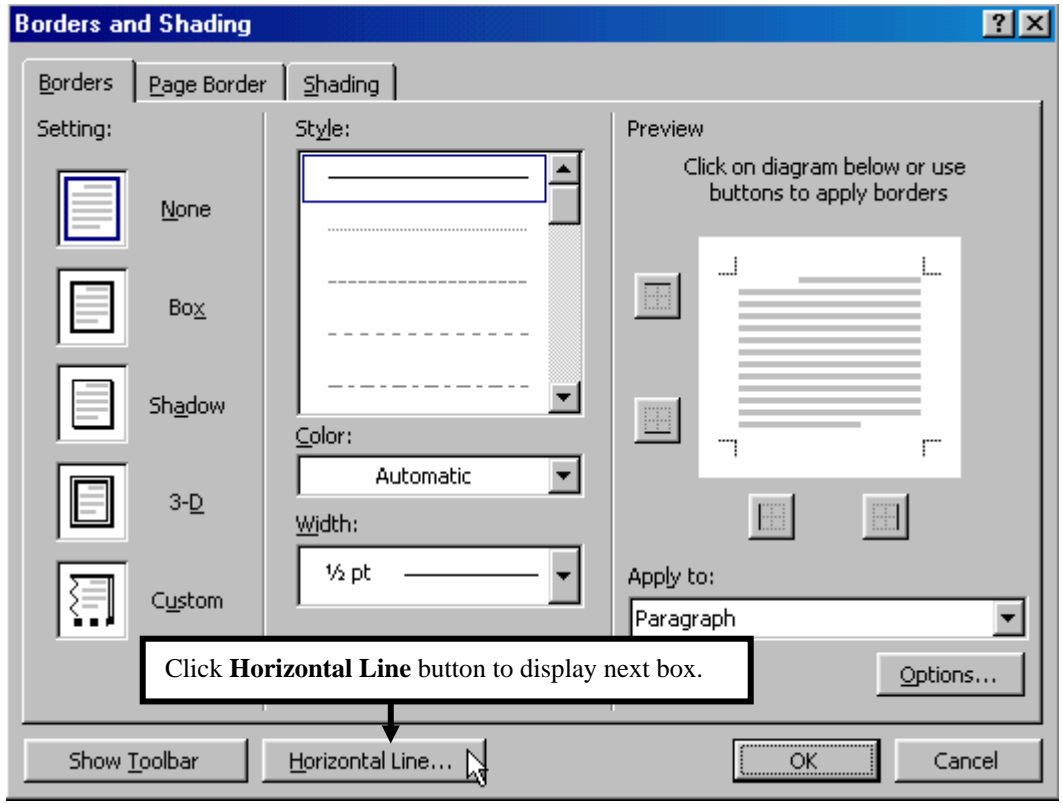
An easier method is just to click the **Numbering** icon on the formatting toolbar for numbers, or the **Bullets** icon just to the right of it (so long as you want plain bullets and not fancy ones):



When you are finished with your numbered (or bulleted) list, just hit the ENTER key twice to stop numbering or bulleting.

Horizontal Lines

Inserting a horizontal line is a good way to break up your document into sections. *Word 2000* has a very easy facility for inserting a horizontal line although, oddly, it is not within the **Insert** pull-down menu. Instead, you must click the **Format** pull-down menu, then **Borders and Shading**. You will then see the following dialog box:



Clicking the top choice (the default) will insert a plain looking line like the following:



You can re-size this line by clicking on it directly and dragging the re-sizing handles at either end (shown above) to make the line longer or shorter. Clicking on one of the other lines shown in the dialog box above will insert a graphic--and all rules about ensuring that it's saved to the appropriate folder outlined in the section on **Important Information Regarding Folders** applies!

Line Breaks

Inserting a line break--though not always recommended because of varying screen/font sizes among users--is another way to ensure your text stays formatted. This is different from starting a new paragraph. Hitting the <ENTER> key in HTML formatting ensures that another paragraph is started, though there will be blank space between the two paragraphs. On the other hand, forcing a line break causes the text to separate, but without the blank line that characterizes a paragraph break in HTML. To insert a line break while editing, simply press: <Shift>-<ENTER>. This may come in handy while in a table cell to ensure a clean line break:

To the right is text without any line break; notice the break may not occur where you want it to.	Chaucer, Geoffrey. <i>Canterbury Tales</i> . New York: Macmillan, 1993.
Now to the right is text <i>with</i> a line break, pressing <Shift>-<ENTER> after the title.	Chaucer, Geoffrey. <i>Canterbury Tales</i> . New York: Macmillan, 1993.

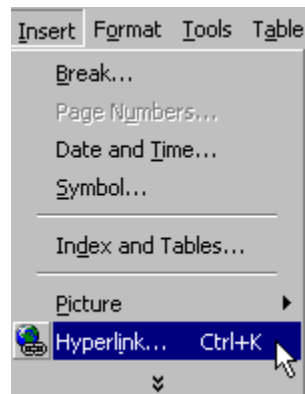
Inserting Hyperlinks

The primary advantage to having a document on the Web is that you can direct your readers to other Web resources using hyperlinks on the page. To create a hyperlink on your page, highlight the text you want hyperlinked (in other words, the material that is typically underlined in blue on a Web page), and do either of the following:

Click the **Insert Hyperlink** icon on the standard toolbar:



Or, click the **Insert** pull-down menu, then **Hyperlink**:



Either option will bring you to the next dialog box:

This box displays the text that will be underlined in blue as the active hyperlink on your Web page.

A **Screen Tip** is the wording that appears in a yellow box on many browsers when the user's mouse hovers over the hyperlink. The wording that appears can be editing by clicking this button.

Type the Web address of where you want your users to go to in the **Type the file or Web page name:** box.

Click the **Bookmark** button to create a hyperlink to another location *within* this Web page (procedure explained below).

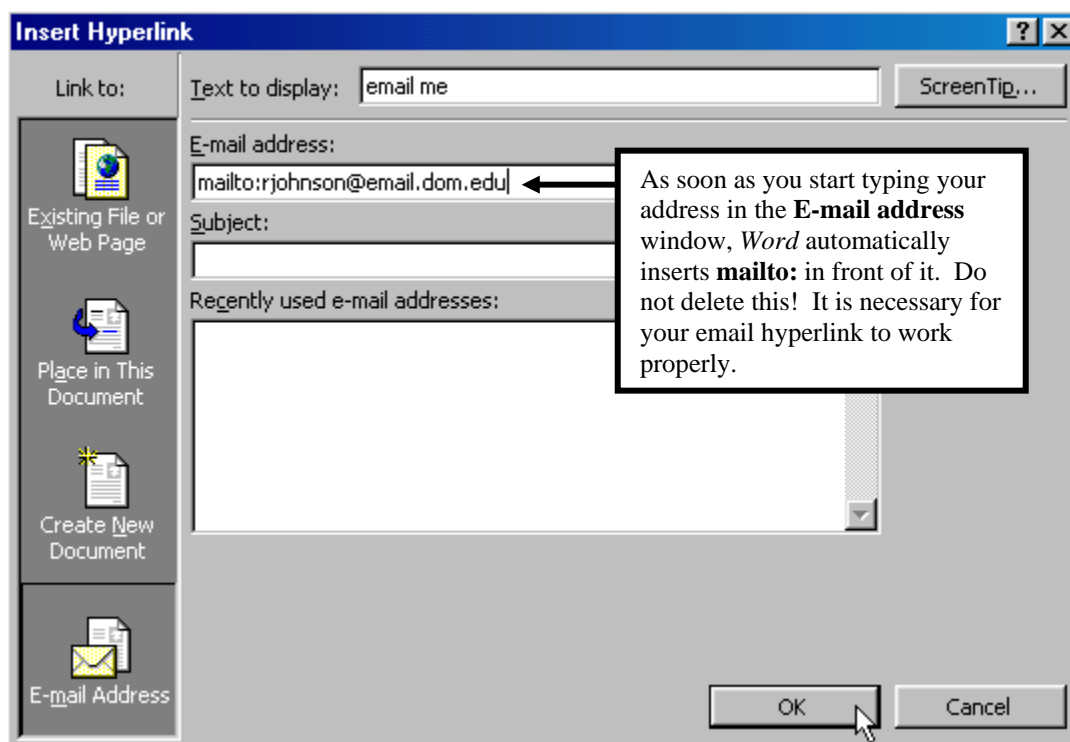
You may find it easier to link to a file on your I: drive by clicking the **File...** button in the box above.

Common Mistake

If you are linking to a **File**, make sure it is already on your I: drive, or it will not be accessible to your readers!

Linking to an Email Address

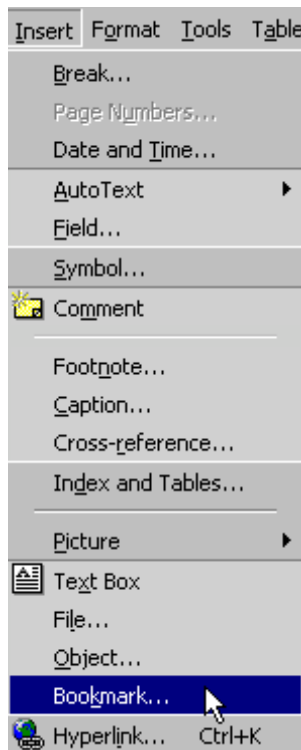
If you want to create a hyperlink to an email address so that the user's email reader automatically opens up with your name in the **To:** field, click the **E-mail Address** button in the Insert Hyperlink box instead:



Creating Hyperlinks Between Locations Within Your Document (Bookmarks)

Internal hyperlinks will navigate your users from one spot on your page to another without having to scroll through the entire document. These are handy if you have your document divided into several sections (such as readings, calendar, class assignments, etc., on a syllabus) but not as separate pages. These internal hyperlinks are also sometimes called targets or, in the case of *Word*, **bookmarks**.

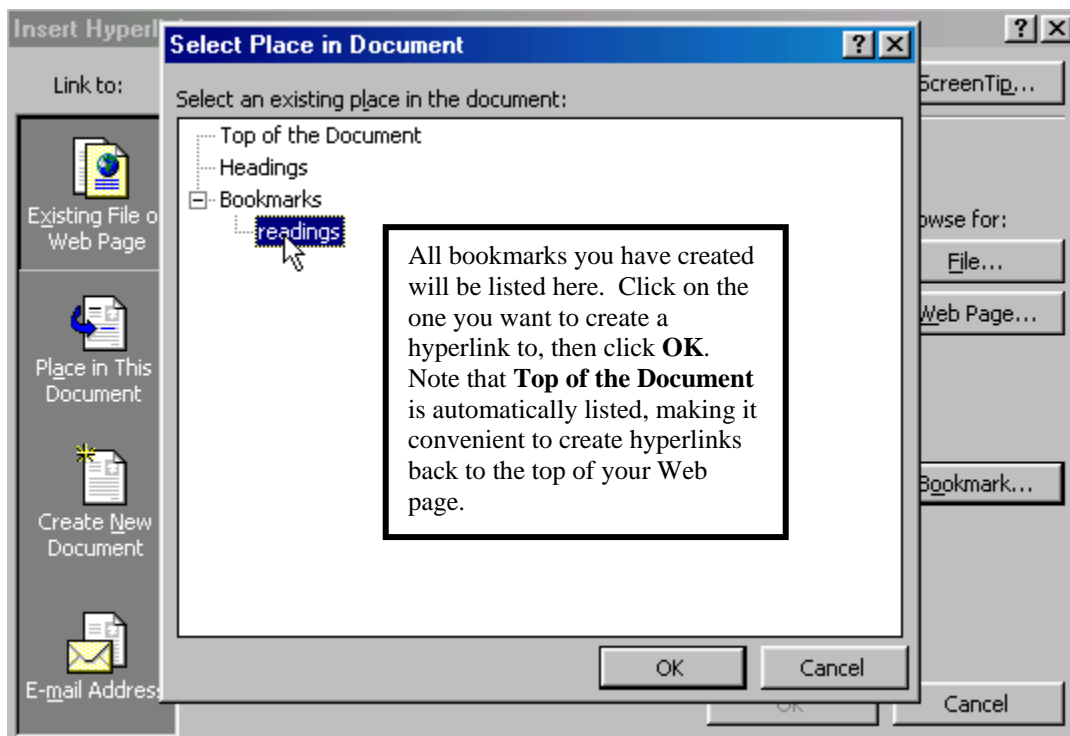
To create an internal hyperlink, first click on a place in the document you would like to establish as a bookmark, such as the heading for COURSE CALENDAR or REQUIRED READINGS. (You may also just highlight the text with your mouse.) Click on the **Insert** pull-down menu, then **Bookmark**:



You will see a box prompting for a name for your bookmark. You may type whatever name seems appropriate, though you should *not* have spaces in it, ideally:

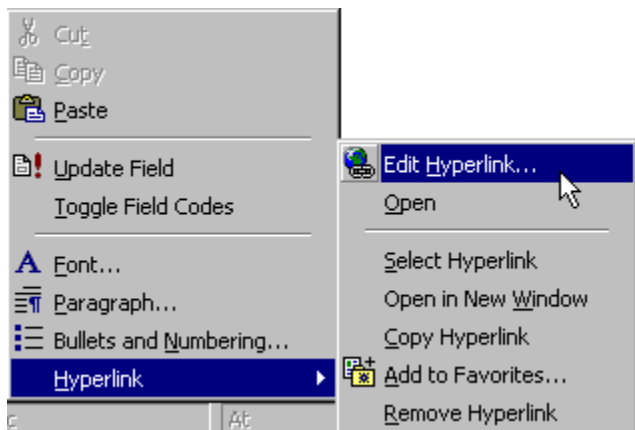


The final step is creating a hyperlink somewhere on your document to this bookmark. You may, for example, have a listing of sections at the very beginning of your Web page--essentially a "table of contents". Each item can be a hyperlink to that section of the document. This hyperlink is created the same way a regular hyperlink is created (**Insert . . . Hyperlink**). For example, you could highlight REQUIRED READINGS at the beginning of a course syllabus, then click the **Insert Hyperlink** icon (as explained in the previous section). This time, however, you should click on the button labeled **Bookmark**. You will then see the following dialog box:



Editing Existing Hyperlinks

Web sites change their addresses from time to time. Even if the *name* of a Web site remains the same, you may discover that the address has not. To change the hyperlink address of something on your Web page, **right-click** the hyperlink on your page, then choose **Hyperlink . . .**
Edit Hyperlink:



The previously shown **Insert Hyperlink** box then appears, where you can change the address of the page.

On the other hand, if the *name* of your site has changed, you will have to change the wording of the underlined link on your page itself. The process is the same, only this time click within the **Text to Display:** section of the Insert Hyperlink dialog box. This is *much* easier than trying to select your hyperlinked text with a mouse--which often results in surfing *to* that hyperlink rather than selecting the text!

IMAGES

You can add clip art and other images to your Web page just as you do when editing a regular *Word 2000* document (click the **I**nsert pull-down menu, then select **P**icture, then choosing either **C**lip Art for the *Office* clip art or **F**rom File for an image you have stored elsewhere). When copying your document on the **I:** drive (DOMIN server), it is critical that you have all image files loaded in the **I:** drive, too. This means *all* images--including those you may have used for lines, bullets, or as a page background! Be sure to check the section **Important Information Regarding Folders** near the beginning of this document for where these images may wind up.

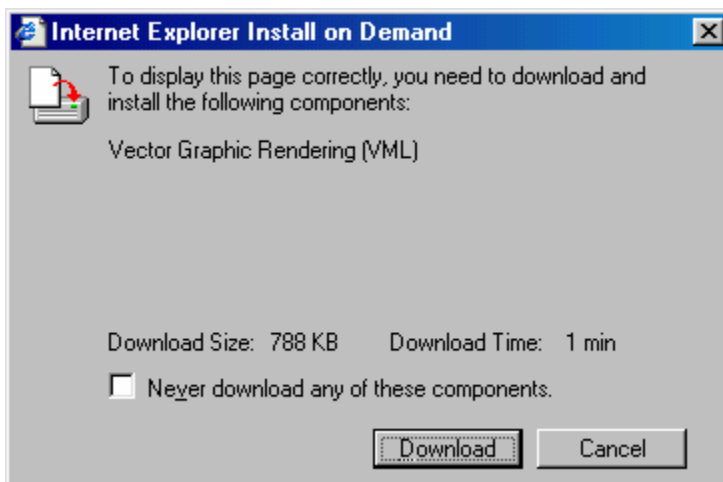


INSERTING CLIP ART OR OTHER GRAPHIC IMAGES?

Don't even *THINK* of proceeding any further before reading the section **How Word 2000 Converts Images** below! It may explain for you why some of your users can't even *see* your images.

HOW WORD 2000 CONVERTS IMAGES

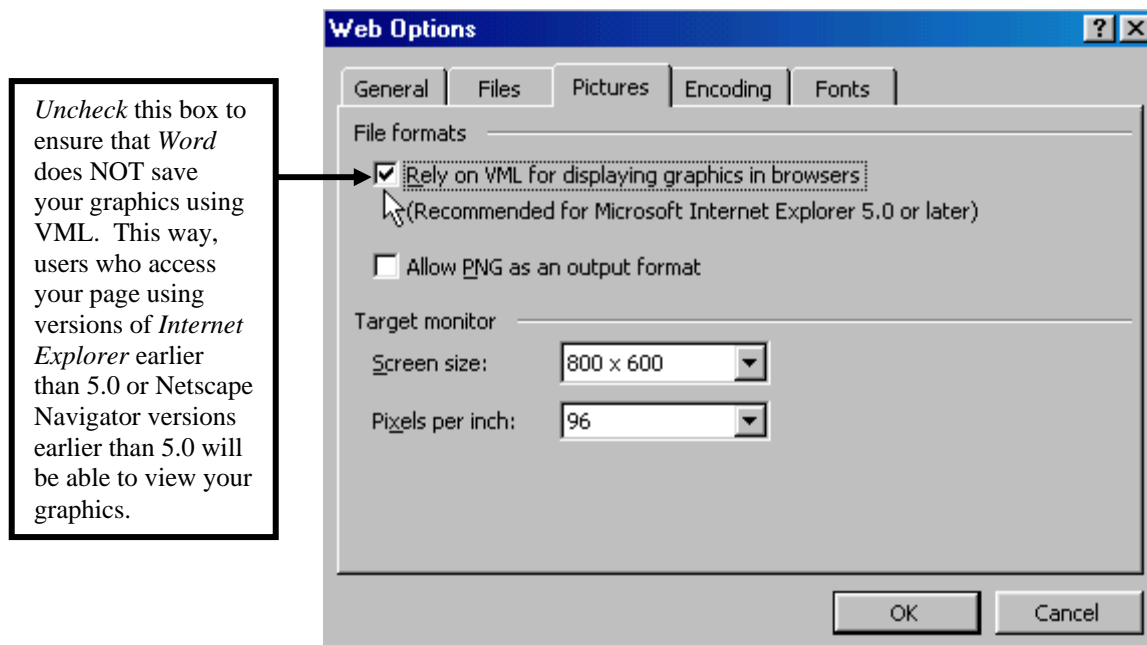
Rather than display images as traditional .GIF or .JPG images (likely used on the majority of Web pages you have seen), *Word 2000* converts these images to VML-based images instead. This speeds up download times, but at the cost of many users perhaps not being able to see your images. What may happen to some of your users--particularly from off-campus--is the following message may appear in *Internet Explorer* when they attempt to open your page:



Essentially, this dialog box is telling the user that they do not have the necessary components loaded in order to view VML images on a Web page. All they need to do is click the **Download** button, and this problem should be taken care of. (Fortunately, this is a relatively small file at 788 KB and should not take too long to download.)

Unfortunately, any users of *Internet Explorer* or *Netscape* who have versions earlier than 5.0 will not be able to view your graphics material at all!

A way around *Word 2000*'s default of saving graphics in the VML format is to click on the **Tools** pull-down menu, then **Options**. Click on the **General** tab, then click on the button labeled **Web Options**. (See pages 3-4 for screenshots of this process.) Finally, click on the **Pictures** tab, and you will see the following:



Unfortunately, this will not solve all of your problems. Regardless of whether you clear the above checkbox or not, *Internet Explorer* versions 5.0 and above will *still* attempt to *read* the

VML markup tags that *Word 2000* creates! (In other words, although the graphics will no longer be rendered using VML, all of the coding associated with it is still created by *Word*. Honestly, this is not our fault. Don't shoot the messenger!) The end result is, your *Internet Explorer* version 5.x users will still need to download the necessary VML tools to read your page. Netscape users (ironically) will read the pages just fine, as will all users with older versions of *Internet Explorer*.



Boring Technical Talk (Skip if you want to)

VML stands for Vector Markup Language and was designed by Microsoft to be used in concert with XML and Cascading Style Sheets. The advantage of VML is that it uses vector-based graphics rather than the traditional bitmap-based graphics that the more common JPG and GIF formats use. While bitmapped images store information as a series of individual pixels, VML stores the images as a description of its shapes, colors, lines, and curves. The end result for a user is that VML images download into a browser much, much faster than the usual GIF and JPG images.

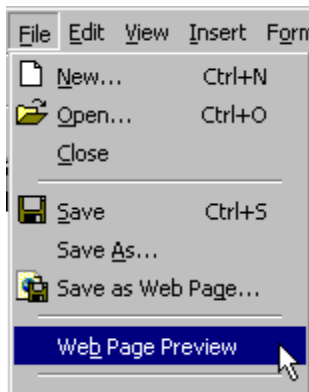
VML is a proposed standard for Web graphics submitted by Microsoft to the World Wide Web Consortium (W3C) in 1998, and which they decided to use in the entire *Office 2000* suite for Web publishing. A competing standard, submitted by Adobe, is the Precision Graphics Markup Language (PGML). At this point, the W3C is considering a standard that *both* Microsoft and Adobe have agreed upon, called the Scalable Vector Graphics (SVG) specification, which looks likely to ultimately be the agreed upon implementation for vector graphics on the Web. As with all things Web-related, stay tuned!

WORD 2003 AND IMAGES

The above dialog box (with VML images) is not seen in *Word 2003*.

CHECKING YOUR DOCUMENT

Once you've completed editing your document, check it in a Web browser to ensure it will look the same as you think it will. Even with *Word's* ability to convert and save in HTML, not everything will look the same on the Web as it does in *Word*. The easiest way to get a "sneak preview" of your Web page is to click on the **File** pull-down menu, then select **Web Page Preview**:



TIP

It is a good idea to do a **Web Page Preview** to see your document as your users will see it. Once you have done a **Web Page Preview** your Web browser will remain open. You can easily switch back and forth between *Word* and your Web browser by clicking on their respective icons on your Windows task bar with both open. **However**, make sure you constantly click the **REFRESH** button on your browser, or else you will not see the most recently saved version of your Web page!

NOTHING LOOKS THE SAME!!!

Sometimes, even the best preparation results in a document that still looks nothing like you intended when you start previewing it in a Web browser. If this happens (and the more complex the original document, the more likely it will), don't worry. Although *Word* does a tolerable job converting documents to HTML, it is not a true HTML editor. If you want a *real* Web page editing tool, consider using *FrontPage 2000* instead of *Word 2000*--particularly if you constantly run into problems with the XML tags *Word* routinely tosses into the document (see first page of this document).

LOADING YOUR DOCUMENT

Once you have completed your document, if you have not already been editing it directly on the academic (or DOMIN) server (usually your **I:** drive) you can copy it into the appropriate directories. Consult the *Steps For Loading Your Document Onto the DOMIN Server* handout prepared by Information Services for additional information on this topic.

Thanks to *Word 2000's* extensive use of XML, there will likely be a few extra files you are not used to seeing. Typically, you may see a file called **[your file name]_filelist.xml** . Be sure to copy this file, too! In addition, if you used any graphics, such file names may be called **[your graphic].wmz** . These too, should be copied!

Remember to Copy ALL Files!

If you are copying your Web page from one computer to another, when in doubt copy ALL files. They are likely necessary to properly display your page. (All the better reason to create a folder on your computer *just* for your Web page to avoid mixing the files up with others.)

If you are faculty member loading a syllabus or personal home page, be sure to notify Ken Black, Director of Teaching and Learning Technology, at 708-524-6512 (email: kblack@email.dom.edu) so that a link can be made from the academic server's Course Syllabi page or Directory of Faculty Pages to your page. Students who wish to be listed in the Directory of Student Pages should also contact Ken Black.

Created July 2001/ minor updates 2005