

“HTML on Word V1.2” user’s manual

Preface

"HTML on Word" is a tool that converts docx format files edited and saved in Microsoft Word (hereinafter referred to as Word) into simple and easy-to-edit HTML. You can easily create a web page from a document created through the familiar Word interface.

Word has convenient and powerful editing features for documents, such as document review, style setting such as heading, automatic outline numbering settings, advanced drawing, table creation and easy creation of hyperlinks. That's why Word allows you to create high quality documents with high productivity. With "HTML on Word", you can easily convert documents created in Word to HTML, so you can efficiently create web pages with excellent contents.

This manual explains the features of "HTML on Word" and how to use Word as an HTML creation tool.

The composition of this manual is as follows:

“Chapter 1 Overview" explains what you need to understand before using this product, such as an overview of the features, operating environment, restrictions, and user support.

“Chapter 2 Installation and License Settings” explains the installation/uninstallation of this product and the type of license.

“Chapter 3 Command-line Version" explains how to use the command-line version to convert Word documents to HTML.

“Chapter 4 Add-in" explains how to use add-in that are embedded in the Word ribbon.

“Chapter 5 Conversion Specifications” explains the specifications for converting styles set in Word to HTML tags.

“5.9 Word Editing Guidelines” provides guidelines on how to use Word's editing features to create web pages.

Notational Convention

This manual uses the following notation for HTML elements.

1. The HTML element type name is enclosed in `<>`, like a `<` tag.
2. In the conversion specification, only the start tag to be output is indicated, and the end tag is omitted even when the end tag is output.
3. Heading rank tags (`<h1>`, `<h2>`, `<h3>`, `<h4>`, `<h5>`, `<h6>`) are collectively referred to as `<h>` tag.
Note that `<h>` tags are not HTML element type names.

Contact

If you have any questions about the features and operations of this product, contact us by e-mail.

xhw@antenna.co.jp

Table of contents

| | |
|---|----|
| “HTML on Word V1.2” user’s manual | 1 |
| Preface | 1 |
| Chapter 1. Overview | 6 |
| 1.1 Features of command-line version | 6 |
| 1.2 Add-in features..... | 7 |
| Chapter 2. Installation and License Settings | 8 |
| 2.1 Installation procedure..... | 8 |
| 2.1.1 Installation options..... | 16 |
| 2.2 License | 17 |
| 2.2.1 Evaluation version | 17 |
| 2.2.2 Official license | 17 |
| 2.3 Uninstallation..... | 18 |
| Chapter 3. Command-line version | 20 |
| 3.1 Command-line startup message | 20 |
| 3.2 Conversion options | 20 |
| 3.2.1 Specifiable parameters in the add-in menu | 22 |
| 3.2.2 Setting file..... | 23 |

| | | |
|------------|---|----|
| 3.3 | Error messages | 23 |
| Chapter 4. | Add-in | 24 |
| 4.1 | Add-in installation/uninstallation..... | 24 |
| 4.1.1 | Add-in installation | 24 |
| 4.1.2 | Add-in uninstallation..... | 24 |
| 4.2 | “Antenna House” tab | 24 |
| 4.3 | “Convert to HTML” button..... | 25 |
| 4.4 | Convert to HTML | 25 |
| 4.4.1 | Application to display the conversion result..... | 26 |
| 4.5 | Changing the conversion destination folder..... | 30 |
| 4.6 | “Use specified CSS” | 30 |
| 4.7 | “Line break at block tag” | 31 |
| 4.8 | Help..... | 32 |
| 4.9 | Error messages | 32 |
| Chapter 5. | Conversion specifications | 34 |
| 5.1 | Original documents | 34 |
| 5.2 | Version of destination HTML | 34 |
| 5.3 | Root, head and meta-information..... | 34 |
| 5.4 | Block elements..... | 35 |
| 5.4.1 | Heading styles and outline levels..... | 35 |
| 5.4.2 | Heading outline numbers | 36 |
| 5.4.3 | Lists..... | 36 |
| 5.4.4 | Paragraph numbering and ordered lists..... | 36 |
| 5.4.5 | Paragraph style name (optional)..... | 37 |
| 5.5 | Figure and figure arrangements | 37 |
| 5.5.1 | Image and shape formats..... | 37 |
| 5.5.2 | Layout Options..... | 37 |
| 5.5.3 | Position to output the figure with “With Text Wrapping” specified | 38 |

| | | |
|------------|--|----|
| 5.6 | Formula..... | 38 |
| 5.7 | Tables..... | 39 |
| 5.7.1 | Table header row..... | 39 |
| 5.7.2 | Table header column..... | 40 |
| 5.8 | Inline elements..... | 40 |
| 5.8.1 | Font group..... | 40 |
| 5.8.2 | Links and cross-references..... | 42 |
| 5.8.3 | Paragraph text alignment..... | 42 |
| 5.9 | Endnote..... | 43 |
| Chapter 6. | Word Editing Guidelines | 44 |
| 6.1 | Principle of content and style separation..... | 44 |
| 6.1.1 | What is web page content and layout separation?..... | 44 |
| 6.1.2 | Word is a mixture of content and layout..... | 44 |
| 6.1.3 | This product ignores Word layout specification in principle..... | 44 |
| 6.1.4 | Things to avoid when creating Word documents..... | 45 |
| 6.2 | Output the HTML heading rank tag..... | 45 |
| 6.2.1 | Set the Word heading style..... | 45 |
| 6.2.2 | Set the title..... | 46 |
| 6.2.3 | Set the paragraph outline level in Word..... | 46 |
| 6.3 | Bullets and Numbering..... | 47 |
| 6.3.1 | Bullets..... | 47 |
| 6.3.2 | Numbering..... | 49 |
| 6.4 | Layout of shapes..... | 49 |
| 6.4.1 | In line with text..... | 50 |
| 6.4.2 | With text wrapping..... | 50 |
| 6.5 | Blank lines and spaces in Word..... | 52 |
| 6.6 | Grouping of shapes and pictures..... | 53 |
| 6.7 | Reference links..... | 53 |

| | | |
|-----------|--------------------------------------|----|
| 6.7.1 | Link..... | 53 |
| 6.7.2 | Cross-reference | 53 |
| 6.7.3 | Link reference | 54 |
| 6.8 | Tables | 55 |
| 6.9 | Character decoration and fonts | 55 |
| Copyright | 55 | |

Chapter 1. Overview

This product consists of (1) a command-line version program (Word2HTML) that converts a docx file to an HTML file, and (2) an add-in built into the Word ribbon.

The target file format in the command-line version is a docx format file edited and saved with Microsoft Word (hereinafter referred to as Word). Old Word document format (extension “doc”) cannot be converted.

1.1 Features of command-line version

Word2HTML is a converter (conversion engine) that reads docx format files and converts them to HTML format. The conversion engine was originally developed using the technology of "[Office Server Document Converter](#)", which is a product of Antenna House. It hasn't used the "Save As" feature in Word. The conversion engine runs as a command-line version program of Windows.

The command-line version can be operated directly from a Windows DOS window. When converting a document being edited in Word to HTML from the add-in menu, the add-in launches the command-line version.

The command line conversion operation can be specified with options and parameters. For details, refer to "Chapter 3 Command-line Version". You can also set some options and parameters from the add-in. (The features that can be set in the add-in are limited.)

[Notice] The command-line version license is for use on a local PC. It is not permitted to install this product on a PC used as a server and use this product from a client PC connected to that server via a network. If you would like to install and use it on the server, contact our sales staff (sis@antenna.co.jp).

[Notice] The command-line version does not support folders and files on Microsoft's OneDrive.

- Files on OneDrive cannot be selected for conversion.
- Folders on OneDrive cannot be specified as a conversion destination.

1.2 Add-in features

The add-in adds (1) a feature to set conversion options, (2) a feature to convert the contents of the document being edited in Word to a HTML file, and (3) a feature to display the converted HTML file in the associated application as the menu of the Word ribbon.

The conversion process itself in (2) above is performed by the add-in program starting Word2HTML. After converting the docx file to an HTML file, an application such as a browser will open to display the HTML file.

For details, refer to "Chapter 4 Add-in".

The add-in menu has built-in Japanese and English. When the Word language setting is Japanese, the add-in menu will also be Japanese. When the Word language setting is English, the add-in menu will also be English.

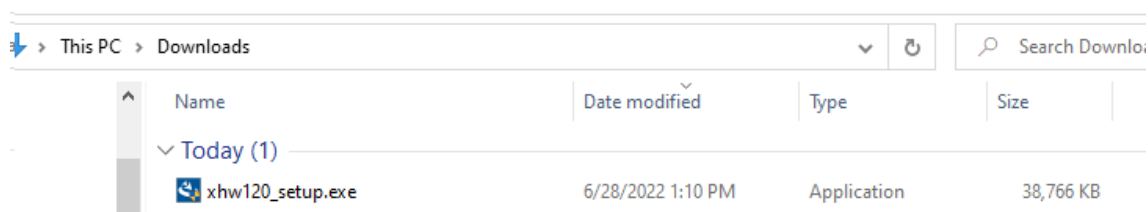
Chapter 2. Installation and License Settings

V1.2 and earlier versions cannot be installed in the same environment. If you have an older version (V1.1 and earlier) installed, uninstall the older version and then install V1.2. See 2.3 for uninstallation.

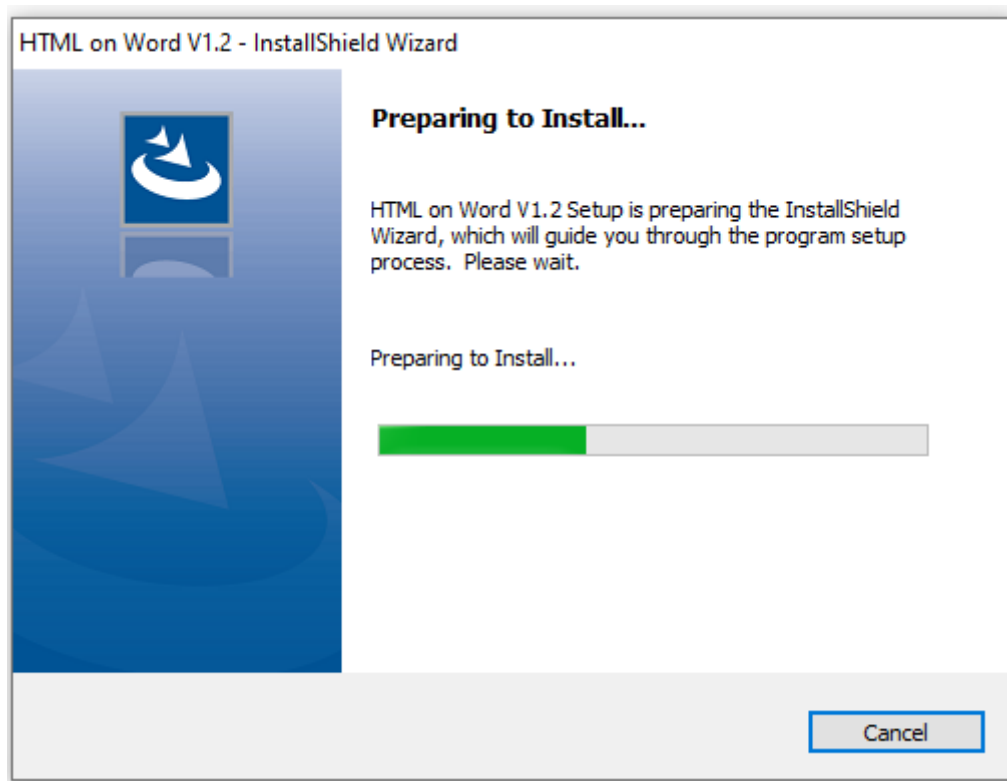
2.1 Installation procedure

When you download this product on your PC, the ZIP format archive file (xhw120_setup.zip) will be saved in the download destination folder.

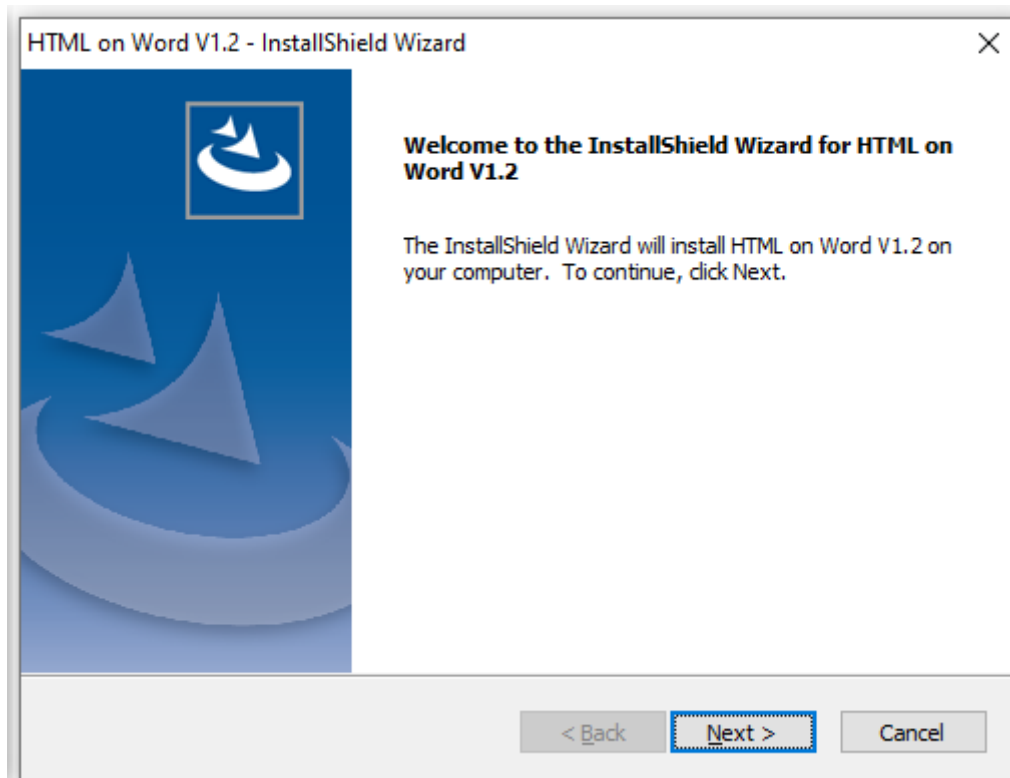
- ① When the ZIP format archive file is decompressed, the installer file (xhw120_setup.exe) of this product will be created in the decompression destination folder.



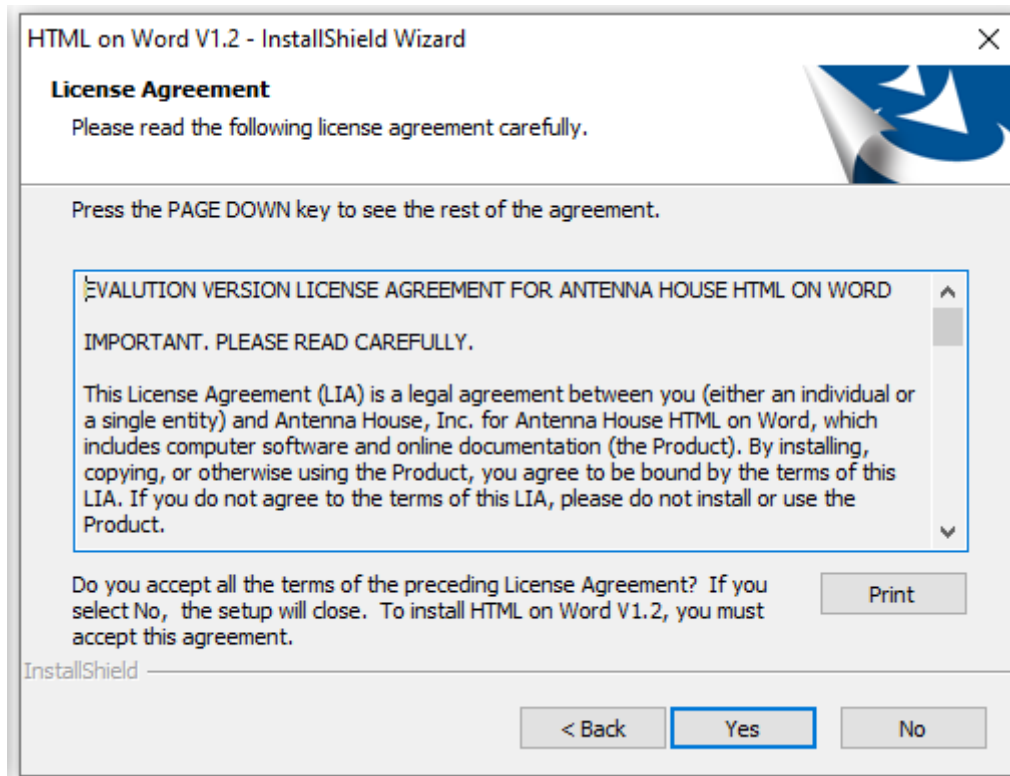
- ② Select the xhw120_setup.exe file with the mouse and double-click it. Windows will display a confirmation dialog asking "**Do you want to allow this app to make changes to the device?**", then click "**Yes**".
- ③ The installation program will start and preparations for the installation will start.



- ④ When the installation preparation is completed, a dialog confirming the start of installation is displayed.

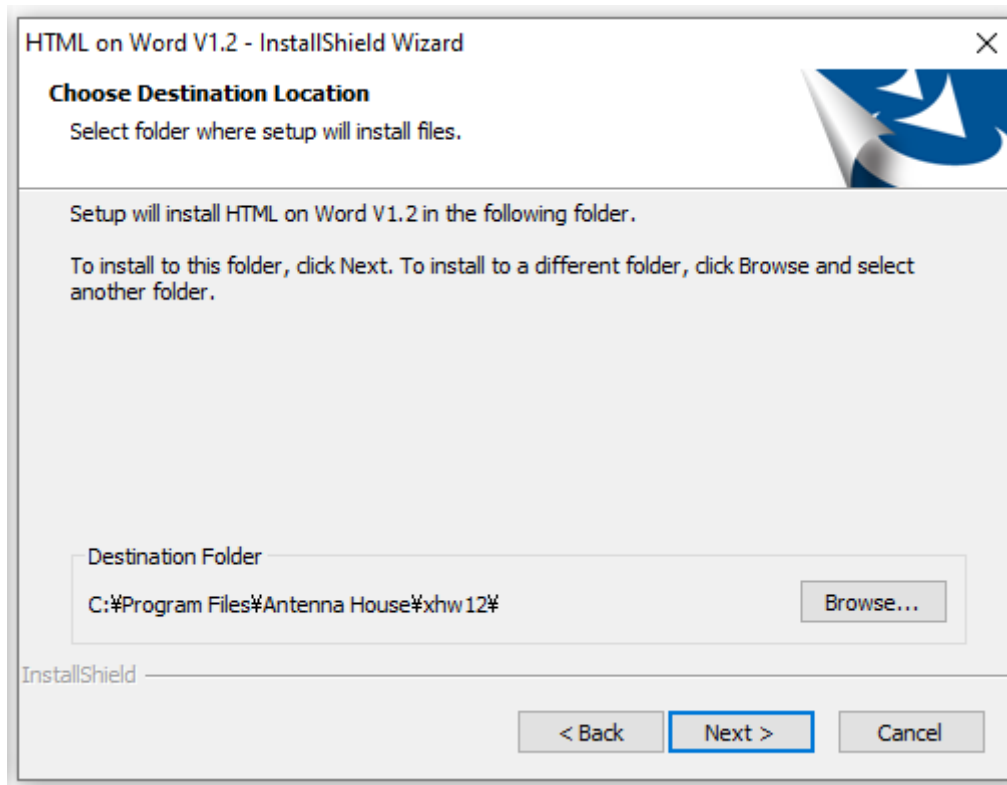


- ⑤ In the next dialog, the License Agreement for this product will be displayed. Confirm the contents, and if you agree it, click **"Yes"**.

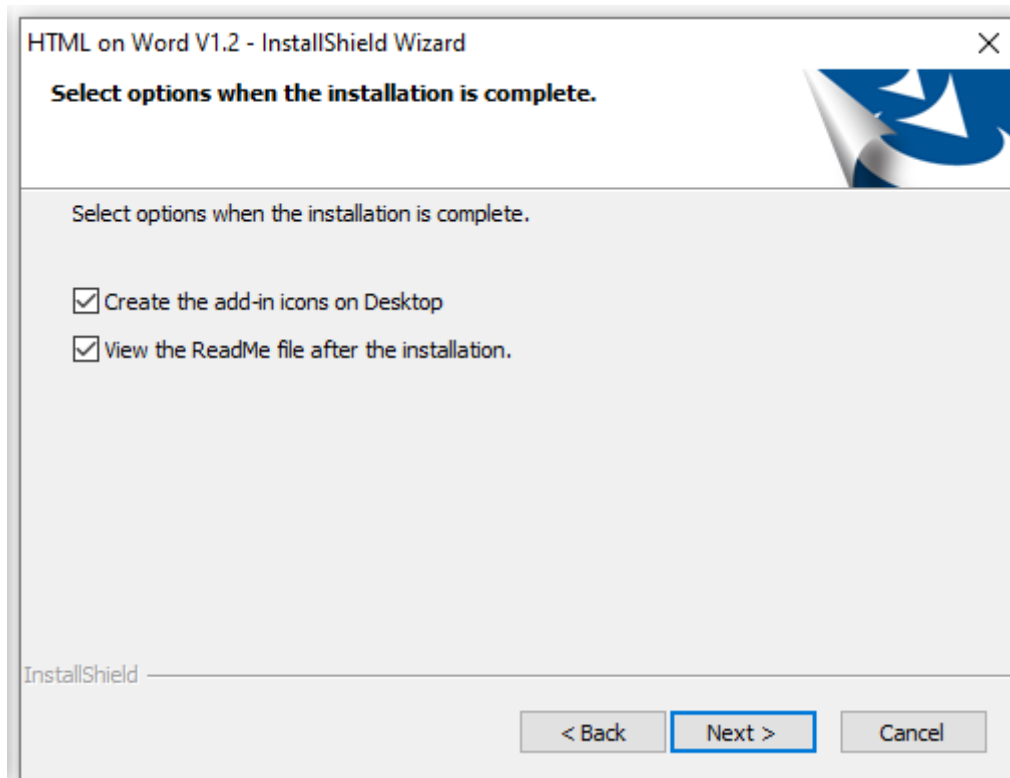


- ⑥ Select the folder to install this product in the next dialog. The default installation location is C:\Program Files\Antenna House\xhw12\ If this is acceptable, click the "Next" button.

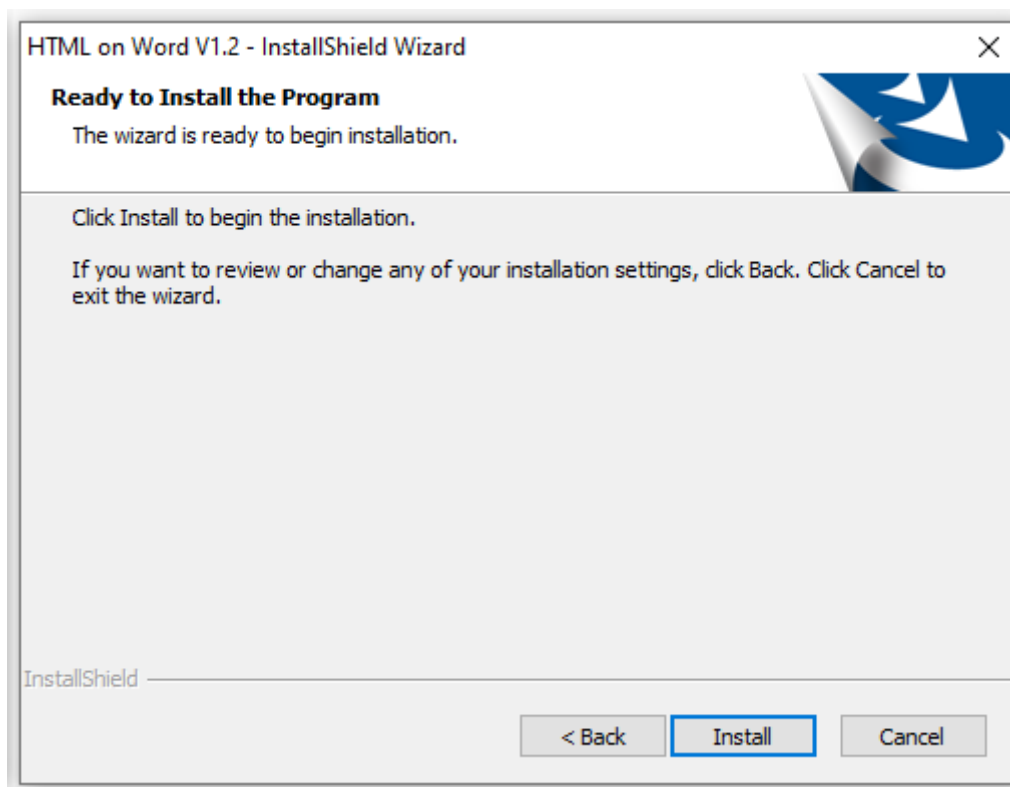
To change the installation folder, select the installation folder from the folder selection dialog that appears when you click the "Browse" button on the right.



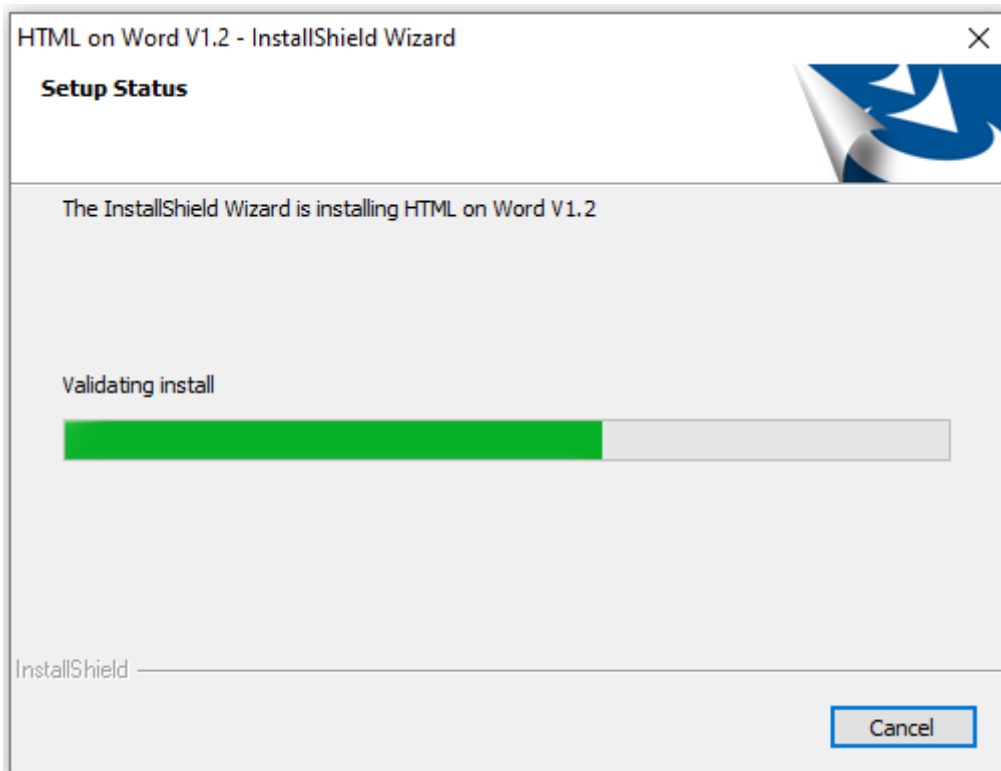
- ⑦ Next is a dialog for selecting options when the installation is completed. You have two options: See 2.1.1 Installation options for a description of the options.
- ☐ Create the add-in icons on Desktop
 - ☐ View the ReadMe file after the installation



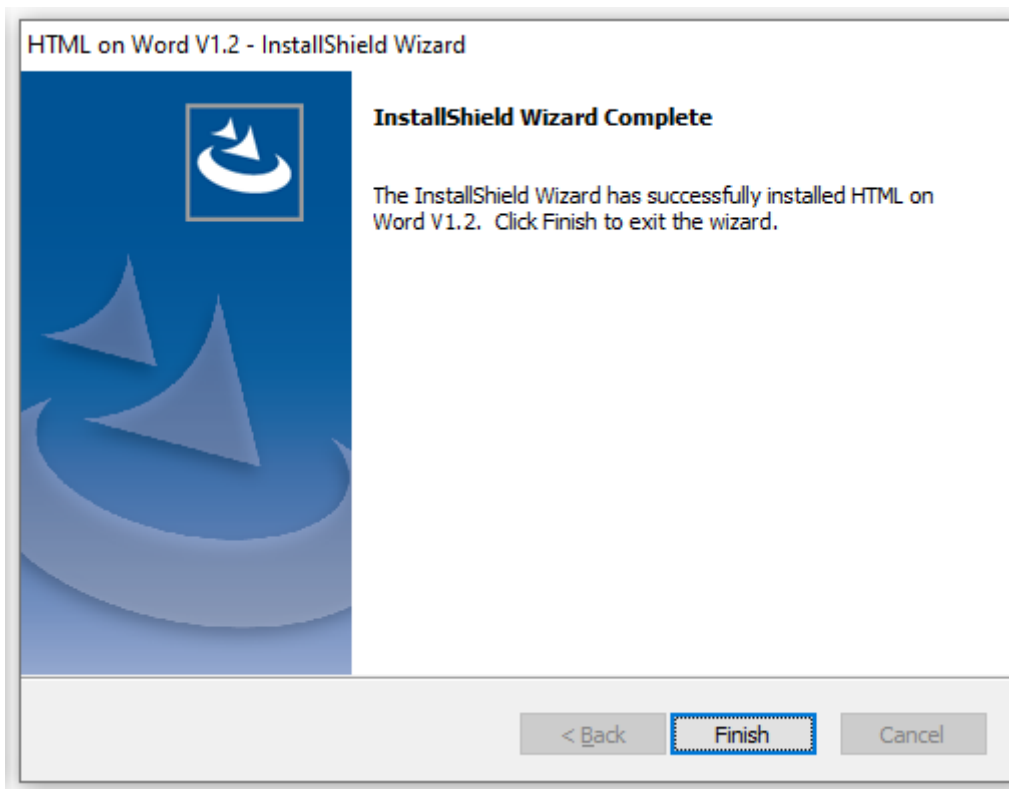
- ⑧ Click "Next" to display the final confirmation dialog asking whether to start the installation.
Click "Install" to start the installation.



- ⑨ The installation process is executed.



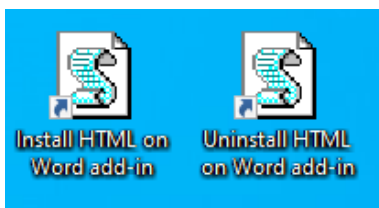
- ⑩ When the installation is completed, the following dialog will be displayed. Click “Finish” to run the options.



2.1.1 Installation options

① Create the add-in icons on the desktop

If you check the checkbox of this option, two icons will be created on the desktop: the program that install the add-in on the ribbon of Microsoft Word and the program that uninstall the add-in.



For the Installation/Uninstallation of add-in, refer to "Chapter 4 Add-in".

② View the ReadMe file after the installation

If you check the checkbox of this option, the ReadMe.txt file included in the installer will be displayed on the screen with Notepad when the installation is completed.

2.2 License

There are two types of HTML on Word licenses: a 30-day evaluation license and an official license. There is no difference in conversion features between these two types of licenses. The license type is switched according to the license file.

2.2.1 Evaluation version

The evaluation version can be obtained from the [product's web page](#).

When the installation of the evaluation version is completed, a license file containing the 30-day evaluation license data will be set in the installation folder.

There are no feature restrictions on the evaluation version, and you can use the same features as the official version. However, the usage period of the evaluation version is limited to 30 days, and you will not be able to start the command-line version 30 days after the installation. To continue using it, you need to purchase the official version from the Antenna House Online Shop.

Antenna House Online Shop

<https://web.antenna.co.jp/shop/html/>

2.2.2 Official license

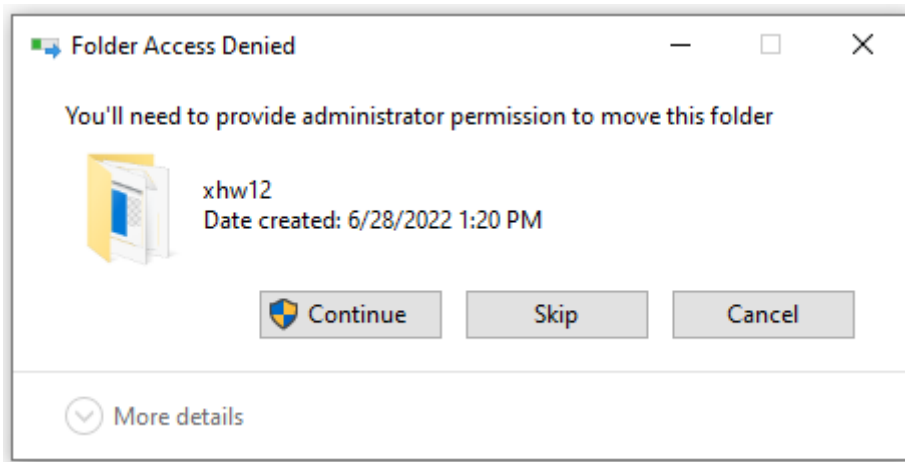
When you purchase this product, the official license data (license file) and license certificate will be provided to you. The license file name is “xhwlic.dat”.

To switch to an official license, copy the license file to the same folder as the command-line version of the program (Word2HTML.exe).

The default installation folder for the command-line program is as follows:

C:\Program Files\Antenna House\xhw12

Writing to this folder requires administrator privileges, so if you try to copy the official license file, you will see the following warning dialog:



If you do not have administrator privileges, ask the administrator to copy it.

2.3 Uninstallation

To uninstall this product, follow the steps below:

① Add-in uninstallation

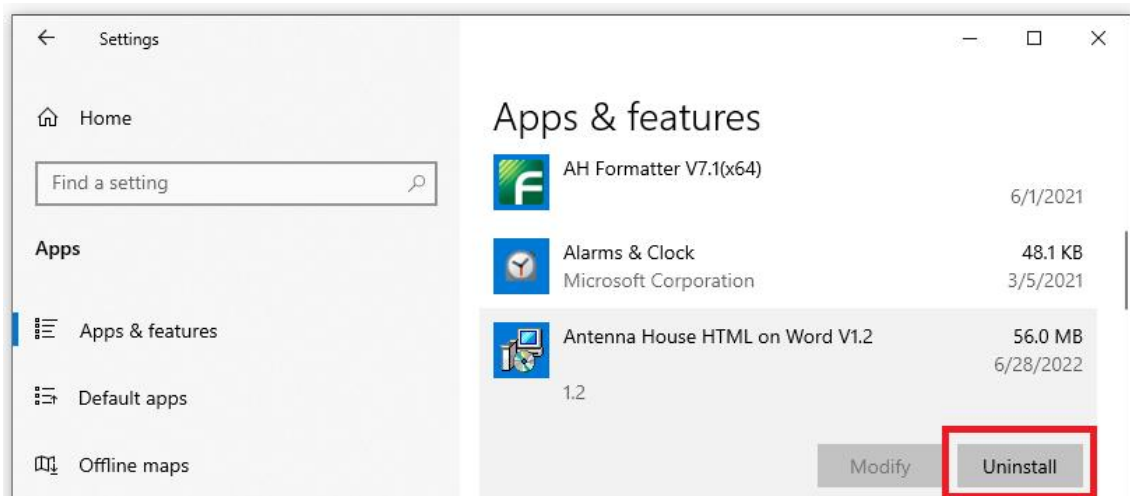
If you have an add-in installed in Word, uninstall the add-in first. For information on uninstalling an add-in, refer to "4.1.2 Add-in uninstallation".

Note that if you uninstall the command-line version without uninstalling the add-in, the program that uninstall the add-in will also be deleted and you will not be able to uninstall the add-in from the Word ribbon.

② Command-line version uninstallation

Uninstall the command-line version from the "Apps and Features" screen of "Settings" in Windows.

"Apps and Features" shows a list of applications installed on Windows. Find "HTML on Word V1.1" and click it to enable the "Uninstall" button as shown below.



Click "Uninstall" to start the installer and perform the uninstallation process.

Chapter 3. Command-line version

The command-line version is a program used from the Windows command prompt. It provides the feature to convert the input docx file to an HTML file.

3.1 Command-line startup message

When you start the command-line version, the following message is displayed:

```
C:\Program Files\Antenna House\xhw12>Word2HTML.exe
Word2HTML : docx to HTML Converter for Windows : V1.2
Copyright (c) 1999-2022 Antenna House, Inc.
Serial: XHW12Serial Number Trial Deadline: Deadline

usage: Word2HTML [-settings <settings-file>] [-xhtml] [-viewport <content>] [-en
dl] [-emptyP] [-nonrefid] [-imgwidth]
               [-hstrong] [-embedding] [-(x|o)math] [-throughimg] [-pstyle] [-c
itation] [-textcolor] [-italic n|t|s]
               [-underline n|t|s] [-linethrough n|t|s]
               <input-file> <output-file> [-css css-file [<media>]]* [-js java
script-path]*
               [-savesettings <settings-file>]
```

The alphanumeric characters beginning with “XHW12” are the serial numbers.

The meaning of the message following the serial number is as follows:

| | |
|-----------------------|------------------------------|
| Maintenance Deadline: | For official version |
| Trial Deadline: | For trial/evaluation version |

3.2 Conversion options

After the Word2HTML message, specify the input file name (required), the output file name (required) and the conversion option.

The following table shows the parameters that can be specified as conversion options. The two conversion options that can be specified from the add-in are (1) the CSS file to link (-css) and (2) whether to output line breaks at the end of block tags (-endl). All conversion options and parameters can be specified when launching Word2HTML from the command prompt.

| Parameter | Description |
|---------------------------|---|
| <input-file> | (Required) Specify the input file name. |
| <output-file> | (Required) Specify the output file name. |
| -settings <settings-file> | Reads the conversion parameter setting file specified in <settings-file>. |

| | |
|---------------------|--|
| -xhtml | By default, HTML grammar tags are output, but if -xhtml is specified, XML grammar tags are output. |
| -viewport <content> | Outputs a meta tag of the following format to <head>. <meta name="viewport" content=" Content specified in 'content'"> |
| -endl | Outputs a line break at the end of the block tag. |
| -emptyP | By default, blank lines (lines with line breaks only) in Word are ignored when outputting HTML. When this option is specified, empty <p></p> tags are output as many as the number of blank lines. |
| -nonrefiid | While editing in Word, a lot of IDs that are not internally referenced may be created. By default, this converter scans IDs that are not internally referenced and deletes them when outputting HTML. Unreferenced IDs will not be deleted when this option is specified. |
| -imgwidth | Outputs the width of the image. |
| -hstrong | Ignores the emphasis specified in the heading style. |
| -embedimg | Embed the image in the body HTML as a data URL. |
| -(x o)math | Specifies the output format for formulas edited in the Word formula editor. The following four output formats can be specified: Unspecified: Output formulas to tags as files in svg file format. -math: Output formulas to tags as files in MathML format. -xmath: Output formulas in MathML format markup. -omath: Output formulas in Word's own Office Math format. |
| -throughimg | Outputs the image in its original format inserted into Word. |
| -pstyle | Outputs the style name of the paragraph by setting it as the value of the class attribute. |
| -citation | Outputs the value of tag in the Citation field by setting it as the value of the href attribute of the <a> tag. |
| -textcolor | Outputs the color specified for the text as . |
| -italic n t s | Specifies the output method when italics are specified for text: -italic n: Do not output. -italic t: Output as <i>tag. |

| | |
|--------------------------------|--|
| | -italic s: Output as <code><span style="font-style:italic</code> |
| -underline n t s | Specifies the output method when underline is specified for text: -underline n: Do not output. -underline t: Output as <code><u></code> tag. -underline s: Output as <code>style="text-decoration-line:underline;"></code> . |
| -linethrough n t s | Specifies the output method when italic is specified for text: -linethrough n: Do not output. -linethrough t: Outputs as <code></code> tag. -linethrough s: Outputs as <code></code> . |
| -css cssfile | Links the CSS file. Place the CSS file in a folder on Windows and specify its path. An error will occur if the specified CSS file does not exist. You can optionally specify "media". Outputs a link tag of the following format in <code><head></code> . <code><link href="xxx.css" rel="stylesheet" type="text/css" media="print"></code> The specified CSS file is copied to the HTML output destination folder. You can specify multiple pairs of -css and CSS files. |
| -js javascript-path | Place the script tag in <code><head></code> and specify the path (URL) of the JavaScript file in its src attribute. No error will occur even if the specified JavaScript path does not exist. |
| -save settings <settings-file> | Saves the conversion option settings with the file name specified in <code><settings-file></code> . |

3.2.1 Specifiable parameters in the add-in menu

Only the following two parameters can be specified in the add-in menu:

- Use specified CSS
- Line break at block tag

Checking the "Use specified CSS" checkbox corresponds to specifying -css in the command-line version. In the command-line version, you can specify multiple pairs of -css and file name, but you can specify only one in the add-in.

Checking the "Line break at block tag" checkbox corresponds to specifying -endl in the command-line version.

3.2.2 Setting file

You can save the options set on the command-line to a setting file. It is convenient to save the conversion options in a settings file for repeated conversions using the same settings.

From the second time, if you specify the setting file name instead of specifying the same options, you can use the contents of the option setting repeatedly. Since the setting file is in XML format, you can also modify the option settings with a text editor.

3.3 Error messages

The error messages in the command-line version are:

| Error message | Possible cause |
|---|--|
| 'Word2HTML' is not recognized as an internal or external command, operable program or batch file. | ① The command-line version is not installed normally. (Countermeasure) Reinstall. ② The path to the folder where the command-line version is installed is not set. (Countermeasure) In the Windows settings, set the path to the folder in the environment variable. |
| "Cannot Open File" | The conversion destination file cannot be opened. (Countermeasure) It is possible that the conversion destination file has been opened with an editor, etc., and editing is locked. In that case, please finish editing. (Countermeasure) It is considered that the CSS file for which the link is specified does not exist. |
| "Input file not found" | Input file not found. |

Chapter 4. Add-in

4.1 Add-in installation/uninstallation

4.1.1 Add-in installation

To install an add-in, close Word and click the “Install HTML on Word add-in” icon on your desktop.

If you did not create an add-in installation icon on your desktop during installation, do the following:

- ① Find the program for installing the add-in that is copied to the HTML on Word installation folder.
- ② Double-click the add-in installation program file (install.vbs).

4.1.2 Add-in uninstallation

To uninstall an add-in, close Word and click the “Uninstall HTML on Word add-in” icon on your desktop.

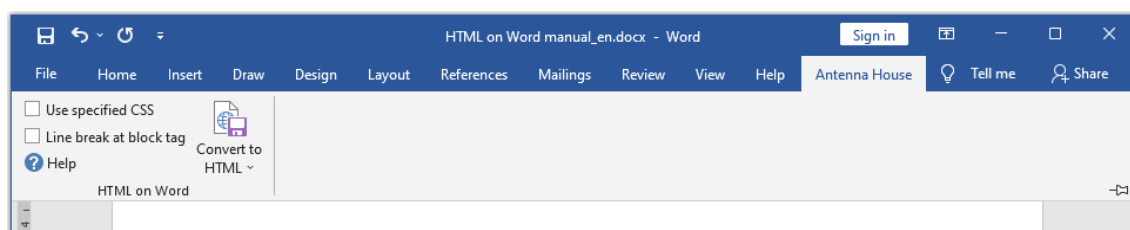
If you did not create an add-in uninstallation icon on your desktop during installation, do the following:

- ① Find the program for uninstalling the add-in that is copied to the HTML on Word installation folder.
- ② Double-click the add-in uninstallation program file (uninstall.vbs) .

4.2 “Antenna House” tab

When the add-in is installed, there will be an "Antenna House" tab on the Word ribbon. The “HTML on Word” group on this tab has the following buttons.

- “Convert to HTML” button
- “Use specified CSS” checkbox
- “Line break at block tag” checkbox
- “Help” button



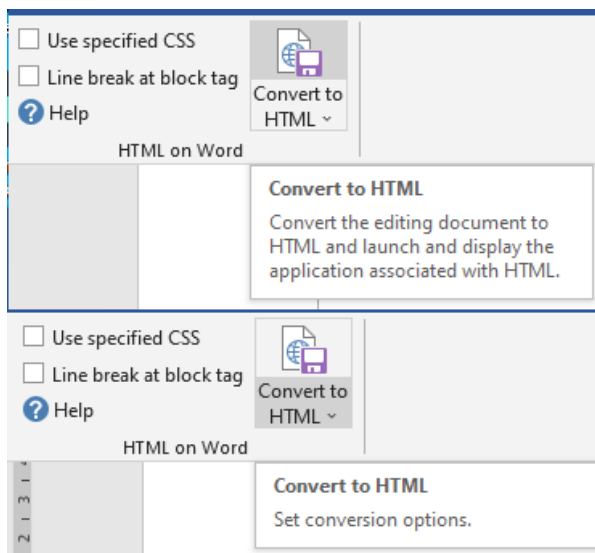
You can use this button and checkbox to convert the document being edited in Word to an HTML file, and check the conversion result with a browser or text editor.

When the "Preferred Languages" is "English" in the "Language" tab of the Windows settings "Time & Language", the tab names and menus in Word will be English as shown in the figure. If this happens, the tab name of the add-in will be “Antenna House”, and the command name and tooltip message are in English.

If you change the "Preferred languages" to "Japanese" in the "Language" tab of the Windows settings "Time & language", the tab names and menus in Word will be in Japanese. If this happens, the tab name of the add-in will be "アンテナハウス", and the command name and tooltip message will also be in Japanese.

4.3 “Convert to HTML” button

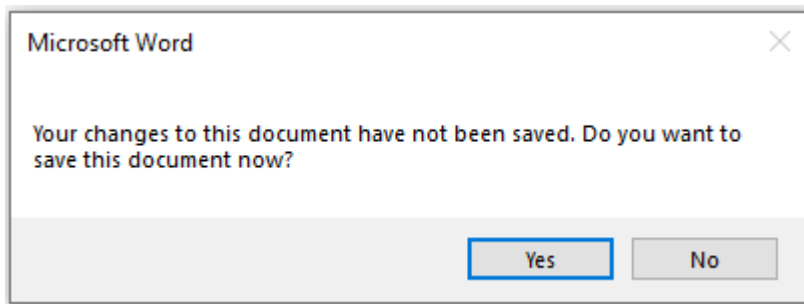
The "Convert to HTML" button is divided into two commands, the "Convert to HTML" command at the top and the "Conversion options" command at the bottom.



4.4 Convert to HTML

Click the top of the "Convert to HTML" button to launch the "Convert to HTML" command. The operation of "Convert to HTML" is as follows:

- ① When the docx document being edited is updated, a dialog prompting you to save the changed document is displayed before the conversion starts.



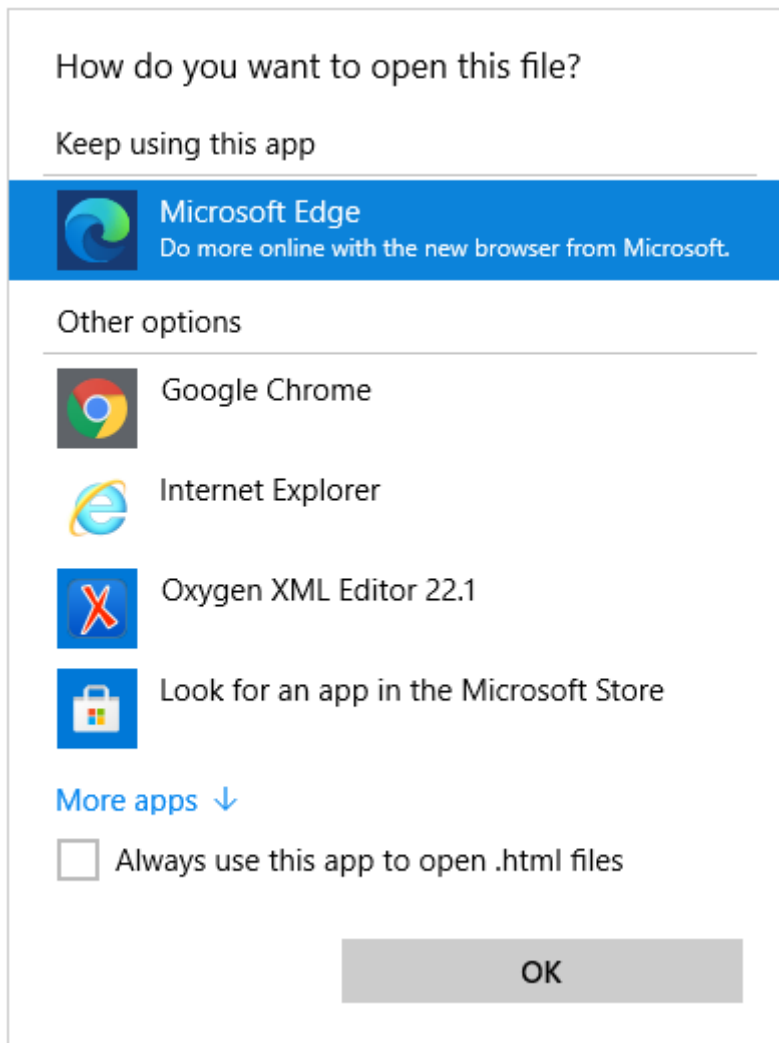
- ② When the HTML save destination folder is not set, the dialog for selecting the save destination folder is displayed. The displayed dialog has the same contents as "4.5 Changing the conversion destination folder", so see also that section.
- ③ Convert the docx document being edited to HTML format.
 - * For the conversion process, start the installed Word2HTML command-line version separately. For the command-line version, refer to "Chapter 3 Command-line version".
- ④ When the conversion is completed normally, the application associated with the extension html is started by the Windows function and the conversion result is displayed.

4.4.1 Application to display the conversion result

When "Convert to HTML" is completed, the HTML file will be displayed in the application associated with the extension html in Windows.

When "Convert to HTML" is executed for the first time, a dialog for selecting the application (browser or editor) to display the file from the applications associated with the extension html in Windows may be displayed.

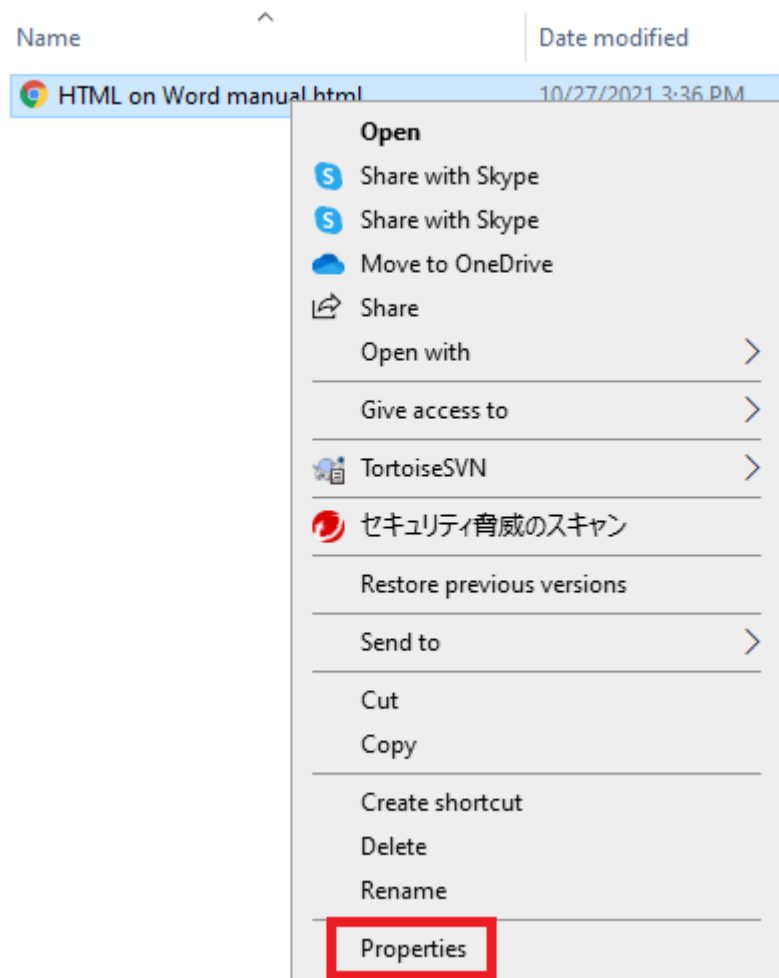
However, depending on the operating environment of Windows, the application selection dialog may not be displayed. This dialog is displayed on Microsoft Windows, and the add-in does not control the display/non-display of this dialog.



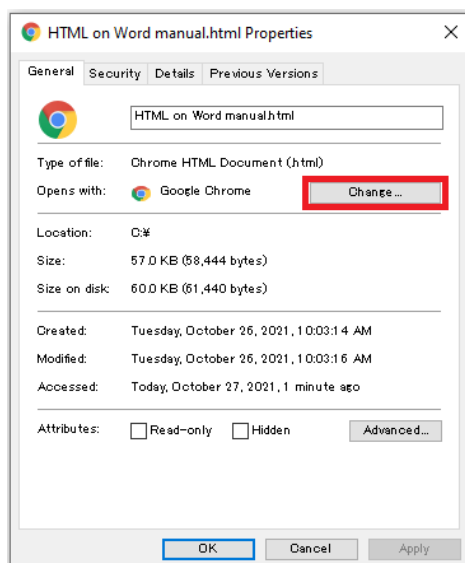
To change the application associated with the html extension on Windows:

1. Select the HTML file in File Explorer.

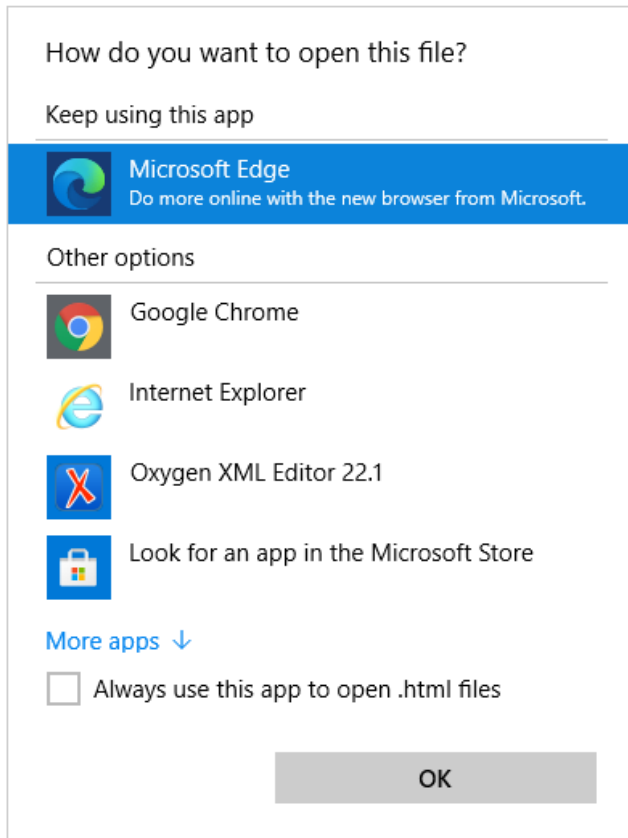
2. Select Properties from the right-click menu



3. From the Properties dialog, click the "Change" button in the “Opens with” section.



4. Select the application in the "How do you want to open .html files from now on?"

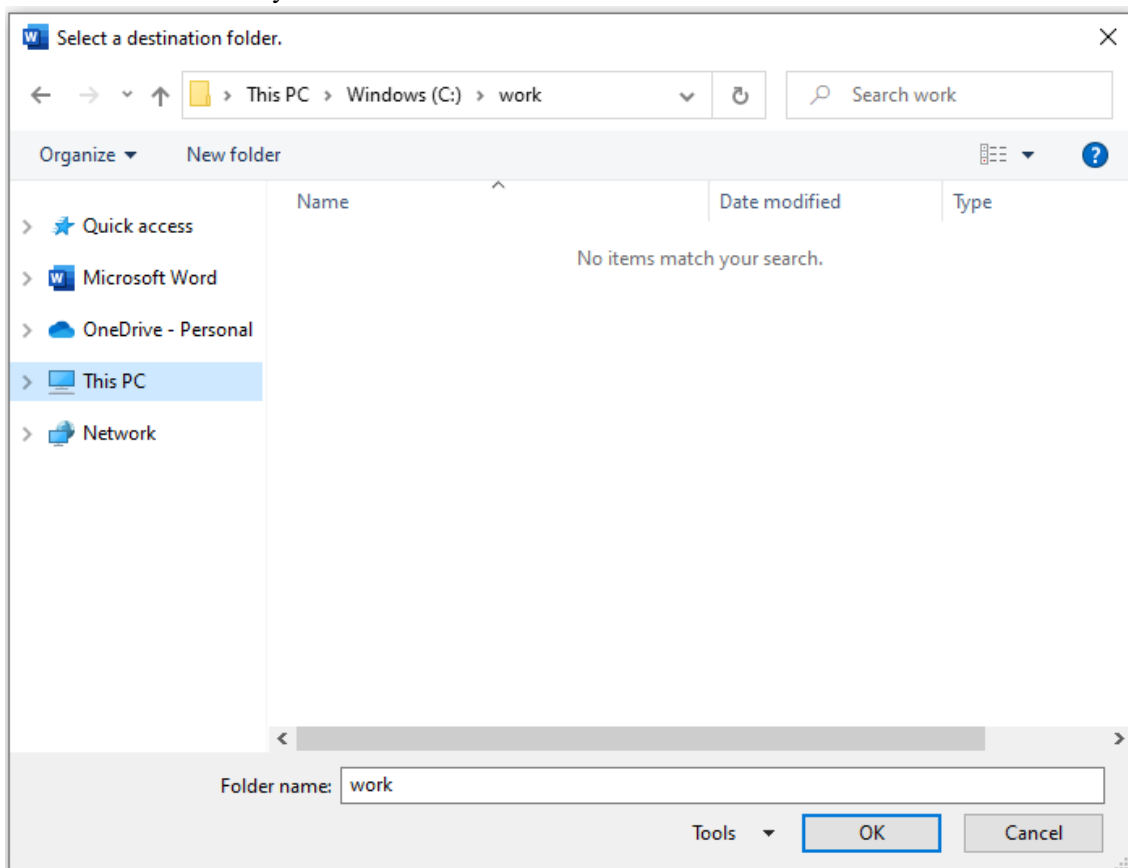


5. Click "OK" to close the dialog.

4.5 Changing the conversion destination folder

Click the bottom of the "Convert to HTML" button to display the "Select destination folder" command.

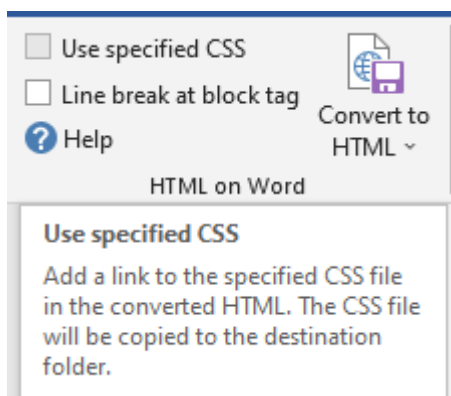
Select the folder where you want to view the conversion results and click "OK".



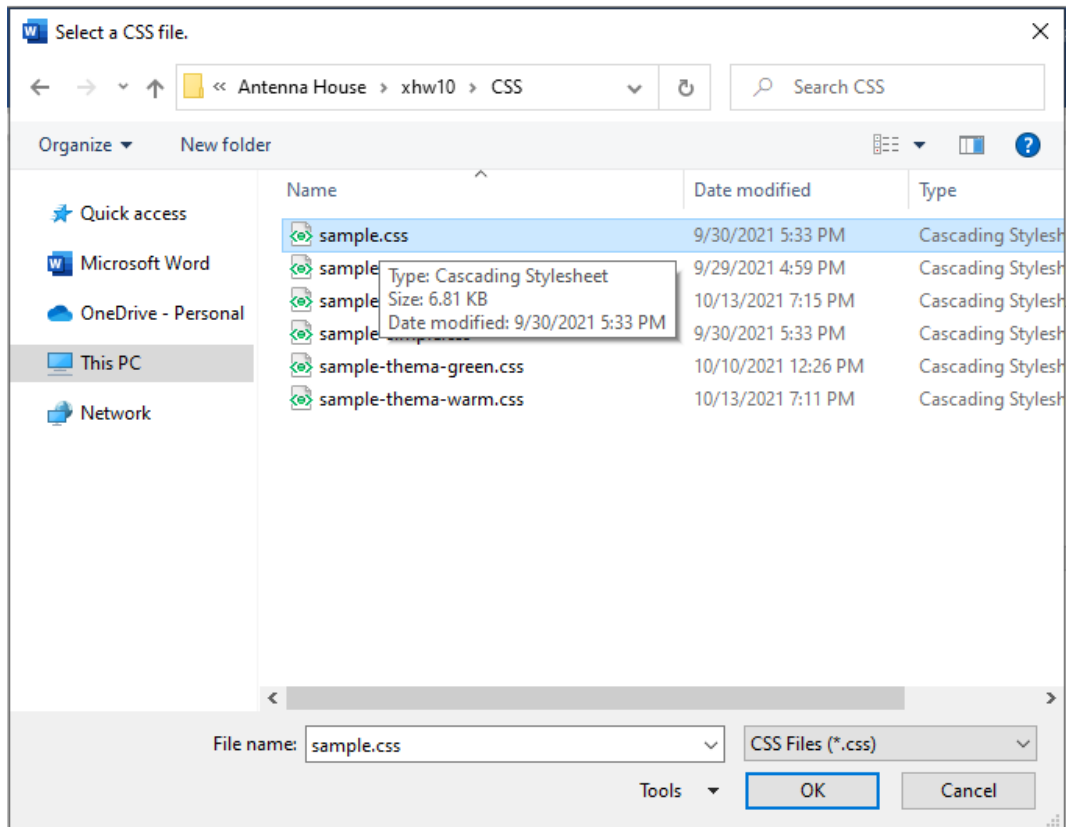
From the next time, the conversion result will be saved in the selected folder.

4.6 "Use specified CSS"

You can change the layout of the HTML file with CSS. Set the link to the specified CSS file with this option to the converted HTML file.



If you check the "Use specified CSS" checkbox, a dialog for selecting a CSS file will open. Select the CSS file you want to link to.

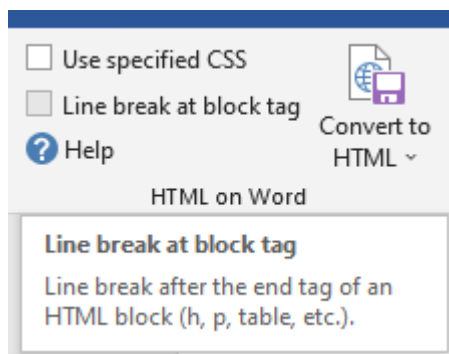


A sample CSS file is included with this product. The sample CSS file is copied to the CSS folder in the folder where this product is installed. In addition to the CSS file included with this product, you can link the CSS file you prepared.

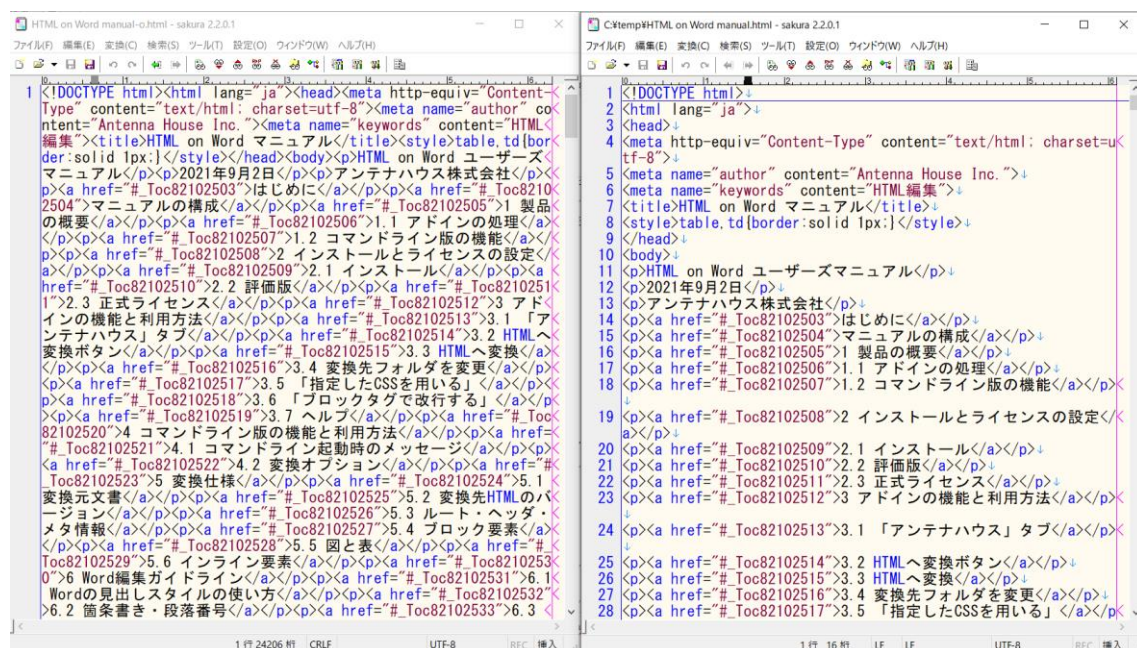
The linked CSS file will be copied to the same folder as the converted HTML file.

4.7 "Line break at block tag"

If you check the "Line break at block tag" checkbox, a line break will be output after each block end tag. It makes no difference when viewing the converted HTML file in a browser, but it is useful when viewing the HTML file in a text editor to view and edit tags.



The following figure compares the display of an HTML file without line breaks in block tags (default) and the display of an output HTML file with the "Line break at block tag" checkbox checked, when read in a text editor..



4.8 Help

Click the “Help” button to display web help. Help can be found on the Antenna House web page. The URL is:

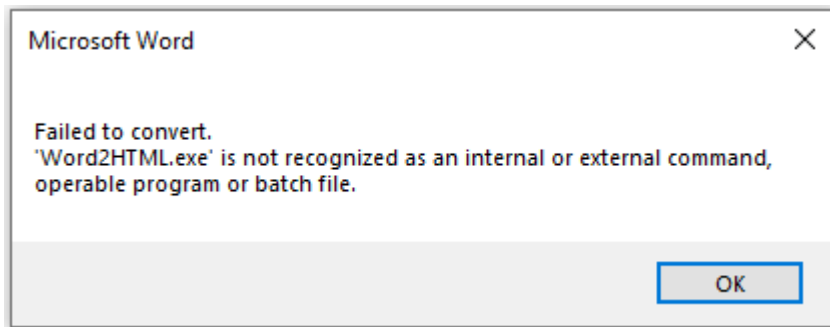
- A) (Japanese)<https://www.antenna.co.jp/xhw/help/ja/>
- B) (English)<https://www.antenna.co.jp/xhw/help/en/>

4.9 Error messages

The add-in "Convert to HTML" launches the command-line version (Word2HTML).

If an error occurs during conversion, the error message output by the command-line version will be displayed in the dialog.

For example, the following error message is a message that the Word2HTML program cannot be found in Windows. The cause may be that the command-line version is not installed normally, or the path to the folder where the command-line version is installed is not set.



For the error message of the command-line version, see “3.3 Error messages”.

Chapter 5. Conversion specifications

This section describes the conversion specifications when converting from Word to HTML on the command-line version.

5.1 Original documents

The original document file format of the conversion source is docx file only. doc format files saved in old Microsoft Word are not subject to conversion processing.

5.2 Version of destination HTML

By default, tags that conform to the HTML specifications are output.

©[HTML specification reference](#)

When converting with an add-in, the version of the destination HTML is HTML only.

If you specify XHTML conversion as an option on the command-line, XHTML 1.0 compliant tags will be output.

In addition, the tag samples of the following conversion specifications explain the state of conformance to the HTML specifications.

5.3 Root, head and meta-information

| Conversion source | Conversion destination (HTML tag) | Remarks |
|-------------------|--|---|
| Root | <!DOCTYPE html> <html lang=""> | Japanese ver.: lang="ja" English ver.: lang="en" See Note 1 for language judgment |
| Encoding | <meta http-equiv="Content-Type" content="text/html; charset=utf-8"> | |
| Info: Title | <head> <title>~</title> </head> | Get the title information from the contents of the property "Title" on the Word "Info" tab. |
| Meta-information | <head> <meta name="" content=""> author, description, keywords | Set the Author, Comments and Tags content of the properties on the Word "Info" tab. |
| CSS link | <link href="xxx.css" rel="stylesheet" type="text/css" media="print"> | xxx.css is the specified CSS file name. The media attribute is optional. |
| Style | <head> <style>CSS style</style> </head> | Set the border attribute of table. However, it is not output when linking external CSS. |

| | | |
|--------------------------|--|---------------------------------|
| JavaScript specification | <pre><head> <script src="xx/yy.js"></script> </head></pre> | xx/yy.js is the JavaScript path |
|--------------------------|--|---------------------------------|

Note 1 Language judgement

Estimated from the percentage of full-width characters in a Word document and the default style language setting. Note that estimates may not be correct.

5.4 Block elements

| Conversion source | | Conversion destination (HTML tag) | Remarks |
|------------------------------------|--|---|---|
| Body text | | <code><body>-</body></code> | |
| Title | When outline level 1 is set for the style. | <code><h1>-</h1></code> | Depending on the style set, some titles have outline level 1 and some do not. |
| | When the style does not have an outline level set. | <code><p>-</p></code> | |
| Paragraph | | <code><p>content</p></code> | By default, lines with only line breaks are ignored. The -emptyP option outputs lines with only line breaks as empty <code><p></p></code> . |
| Forced line break | | <code> </code> | |
| Forced page break and column break | | Ignored. | |
| Section | | When the <code><h></code> start tag is at the beginning or only the <code><h></code> tag with a lower rank before it, the <code><section></code> start tag is output before the <code><h></code> start tag. When there is a <code><h></code> with a higher rank before it, output the <code></section></code> . | Create a tree structure with the <code><section></code> tag before <code><h></code> . |

5.4.1 Heading styles and outline levels

| | | |
|--------------------------------|------------------------------------|--|
| Heading 1 to Heading 6 (Style) | <code><h1>-<h6></code> | Set the heading style outline level to the heading rank tag. |
|--------------------------------|------------------------------------|--|

| | | |
|--|-----------------------------------|--|
| Heading 7 to Heading 9 | <p class="l7">~ <p class="l9"> | Heading style outline levels 7 and 8 are set as class attributes of a paragraph. |
| Paragraph outline levels 1 to 6 | <h1>-<h6> | Set the paragraph outline level to the heading rank tag. |
| Paragraph outline levels 7 to 9 | <p class="l7">~ <class="l9"> | Paragraph outline levels 7 and 8 are set as class attributes of a paragraph. |
| Bullets (with bullet symbol) | / | Remove bullet symbols. |
| When there is one paragraph with numbering. | <p> convert numbering to text. | |
| When paragraphs with numbering are consecutive. | / | Remove numberings. |
| The same applies to the second and lower layers of bullets or paragraph numbering as in the first layer. | | |

5.4.2 Heading outline numbers

When an outline number is added to a paragraph for which a heading style is specified, the outline number is converted to the content string of the <h> tag after specifying the class attribute value number for the outline number. If there is a space between the outline number and the heading text, the space is output as a single-byte space, or if there is a tab, the tab is deleted and a single-byte space is inserted instead.

5.4.3 Lists

Paragraphs with Word lists are converted to HTML lists (unordered lists) (/). At this time, the bullet symbols in Word paragraphs are removed.

5.4.4 Paragraph numbering and ordered lists

Paragraphs that have been numbered at the beginning of a paragraph using Word's paragraph numbering feature (numbered paragraphs) are converted as follows:

- (1) When a numbered paragraph is preceded or followed by an unnumbered paragraph or line break, the numbered paragraph is output as an HTML paragraph (<p> tag). In this case, the paragraph number is output as a normal text with the class attribute value specified as number.

- (2) When two or more numbered paragraphs are consecutive, they are output to an HTML ordered list (/ tags):
- (a) If numbered paragraphs are arranged in a hierarchy and the first and next paragraphs are adjacent to each other, even if they are at different levels, they are considered to be consecutive.
 - (b) Sets the type of numbering specified in the Word document as the value of the class attribute of the tag.

5.4.5 Paragraph style name (optional)

By default, paragraph style names are not output.

When the -pstyle option is specified, the name of the paragraph style is output as the value of the class attribute of the <p> or <h> tag when a paragraph style is specified in a Word paragraph. When paragraph formatting is specified without using the style feature, the value of the class attribute is not set.

5.5 Figure and figure arrangements


| Conversion source | HTML element |
|---------------------|--------------------------------------|
| Pictures and Shapes | |
| | Save the image in the image folder. |

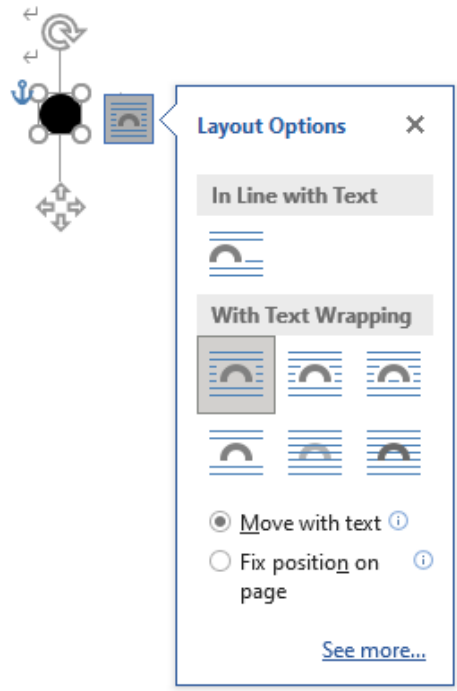






5.5.1 Image and shape formats

By default, images are converted to PNG or JPEG format, and AutoShape, line shapes inserted in Word, and shape files in EMF and WMF formats are converted to SVG format for output. If the -throughimg option is specified, images and shapes inserted into Word in GIF, EMF or WMF formats are output to the image folder in their original formats without file format conversion.

5.5.2 Layout Options

Saves the layout option type specified in Word format as the class attribute.

| Conversion source | Options | class attribute |
|-------------------|---|-----------------|
| In Line with Text |  | class="inline" |

| | | |
|--|--|---------------------------------|
| <p>With Text Wrapping</p>  | <p>Common for “With Text Wrapping”</p> | <p>class="block"</p> |
| | <p> Square</p> | <p>class="block square"</p> |
| | <p> Tight</p> | <p>class="block tight"</p> |
| | <p> Through</p> | <p>class="block through"</p> |
| | <p> Top and Bottom</p> | <p>class="block top-bottom"</p> |
| | <p> Behind Text</p> | <p>class="block behind"</p> |
| | <p> In Front of Text</p> | <p>class="block front"</p> |

[Notice]

In CSS, the display property specifies whether the figure layout is inline or block. Since the default value of the display property is inline, even if you set “With Text Wrapping” in the Layout Options in Word, it may be displayed as “In Line with Text” in the browser. In such a case, specify as follows in CSS:

```
img.block {
display: block
}
```

5.5.3 Position to output the figure with “With Text Wrapping” specified

In headings and paragraphs, the position to output the converted img tag from the figure with “With Text Wrapping” specified in the Layout Options in Word is, after the end tag of the anchored block. However, in bulleted items, it is just before the end tag. For details, refer to “6.4 Layout of shapes”.

5.6 Formula

Formulas edited in Word's formula editor are output as SVG format files using tags by default.

Other options include conversion to MathML format or output as Office Math markup, which is Word's proprietary Office Open XML representation of math formula expressions.

| Option | Output format |
|-------------|---|
| Unspecified | Output formulas to tags as svg format files. |
| -math | Output formulas to tags as MathML format files. |
| -xmath | Output formulas as mathML format markups. |
| -omath | Output formulas in Word's own Office Math format. |

5.7 Tables

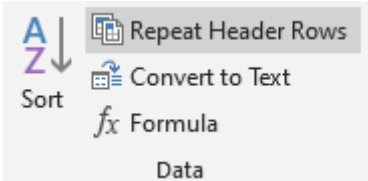
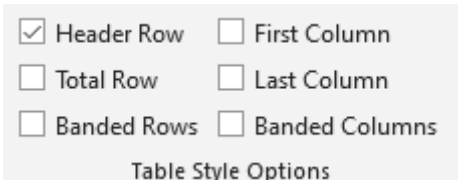
| Conversion source | | HTML element | Example |
|-------------------|------------|-------------------------------------|--|
| Table | | <table> <tbody> <tr> <td> | |
| Merge | Cell merge | <td colspan="n"> | “n” is the number of horizontally merged cells. |
| | Row merge | <td rowspan="n"> | “n” is the number of vertically merged cells. |
| Text Box | | alt=""> | The text box is converted to a line drawing (SVG) image. |

5.7.1 Table header row

To output the table header tag (table header: thead), set either of the following in the first row of the table.

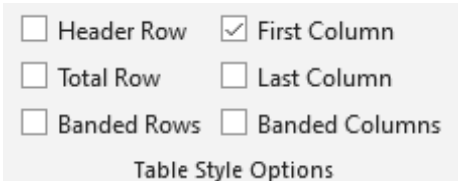
- ① Select the first row of the table and turn on "Repeat Header Rows" in "Table Tools: Layout" on the Word ribbon.
- ② Check **only** "Header Row" in "Table Style Options" in "Table Tools: Table Design" on the Word ribbon.

| Conversion source | HTML element | Description |
|-------------------|--------------|-------------|
|-------------------|--------------|-------------|

| | | |
|--|--|---|
|  <p>“Table Tools: Layout”</p>  <p>“Table Tools: Table Design”: “Table Style Options”</p> | <pre><thead><tr><td>...</td></tr> <></thead></pre> | <p>The first row of the table is enclosed with <thead>.</p> <p>If you turn on “Repeat Header Rows”, the header rows will be repeated on each page when the table spans pages. If you want to avoid this, turn off "Repeat Header Rows" and check "Header Row" in “Table Style Options” in “Table Design”.</p> |
|--|--|---|

5.7.2 Table header column

Select the first column of the table and check **only** "First Column" in "Table Style Options" in "Table Tools: Table Design" on the Word ribbon to set the cell of the first column as the header cell.

| Conversion source | HTML element | Description |
|---|--|--|
|  <p>“Table Tools: Table Design”: “Table Style Options”</p> | <pre><tr><th>...</th></tr></pre> | <p>The cells in the first column of the table are marked up with the header cell tags.</p> |

5.8 Inline elements

5.8.1 Font group

| Font group | HTML element | Example |
|------------|---|---|
| Bold | strong | Note that the bold set in the heading style is ignored. |
| Italic | Ignored by default. Optionally set the <i> tag or the following CSS style specification for output. | |

| | | |
|-----------------------------|---|---|
| | | |
| Underline | <p>Ignored by default. Optionally set the <u> tag or the following CSS style specification for output.</p> <p></p> <p>Note that the anchor text of the link is not underlined.</p> | |
| Strikethrough | <p>Ignored by default. Optionally set the tag or the following CSS style specification for output.</p> <p></p> | |
| Subscript | sub | |
| Superscript | sup | |
| Text Effects and Typography | Ignored. | |
| Text Highlight Color | Ignored. | |
| Font Color | <p>Ignored by default. Optionally set the following CSS style specification for output.</p> <p></p> | <p>text color red, text color green</p> |
| Character Shading | Ignored. | |
| Enclose Characters | Ignored. | |
| Font | Ignored. | |
| Font Size | Ignored. | |
| Case | Ignored. | |
| Phonetic Guide | ruby rp rt | <p><ruby>紫陽花<rt>あじさい</ruby></p> <p><ruby>漢<rp> (</rp><rt>か</rt><rp>) </rp> 字<rp> (</rp><rt> じ </rt><rp>) </rp></ruby></p> |
| Character Border | Ignored. | |

5.8.2 Links and cross-references

| References | HTML element | Example |
|---------------------|---|---|
| Link (external URL) | <code>label</code> | "Link" on the "Insert" tab on the ribbon. |
| Link (id) | <code>label</code> | |
| Cross-reference | <code> label</code> | References in Word documents by "Cross-references" in the "References" tab on the ribbon. |
| | <code></code> | |
| id value | <code></code> | Link to bookmark " here " |

5.8.3 Paragraph text alignment

Set the paragraph alignment set to the "Normal" style in the style gallery on the "Home" tab of the Microsoft Word ribbon to the `<style>` element of the `<head>`. However, when left alignment is set in the "Normal" style, `text-align:start` is the default value in CSS, and it is not necessary to specify the alignment, so it is not set.

| References | | Elements and class attributes | Example |
|-----------------------------|-------------|---|---|
| Alignment of "Normal" style | Align Left | No settings. | <code><style><style></code> |
| | Center | <code>text-align:center</code> | <code><style>html{text-align:center;}</code> |
| | Align Right | <code>text-align:end</code> | <code><style>html{text-align:end;}</style></code> |
| | Justify | <code>text-align:justify</code> | <code><style>html{text-align:justify;}</style></code> |
| | Distributed | <code>text-align:justify;</code> <code>text-justify:auto;</code> | <code><style>html{text-align:justify;text-justify:auto;}</style></code> |

If you specify the paragraph alignment other than "Normal" in the "Paragraph" group on the "Home" tab of the ribbon, the following class attributes will be set in the heading rank tag (h1 to h6) or p tag.

| References | Elements and class attributes | Example |
|------------|-------------------------------|---|
| Align Left | <code>class="start"</code> | <code><p class="start">...</p></code> |

| | | |
|-------------|--------------------|-------------------------------|
| Center | class="center" | <p class="center">...</p> |
| Align Right | class="end" | <p class="end">...</p> |
| Justify | class="justify" | <p>...</p> |
| Distributed | class="distribute" | <p class="distribute">...</p> |

5.9 Endnote

An anchor tag is set to an endnote symbol indicating the location of the endnote in the body text, and the id of the endnote is set to the value of the href attribute of the anchor tag.

The text of the endnote is output at the end of the document, at the same level as the paragraphs at the end of the document except for the endnote. The number of the endnote is set to id="endnote-n" (n is a number).

Chapter 6. Word Editing Guidelines

6.1 Principle of content and style separation

6.1.1 What is web page content and layout separation?

The actual contents of a web page, such as text, images, tables, etc., is called “contents”. In addition, the “layout” is to specify the layout of the block, the margin around the block, whether to surround it with a border, its color, the font to be displayed, the size of the characters and other appearances.

When creating a web page, the contents are marked up with the corresponding tags in HTML and the layout is specified by CSS. In latest HTML, **the basic principle is to separate content from layout.**

6.1.2 Word is a mixture of content and layout

On the other hand, when editing a document in Word, the text format and image layout are specified directly on the text or image while editing on the screen. Word takes a method called "WYSIWYG" that follows the layout on the screen when the document is printed, and the way of thinking about the document layout is fundamentally different from HTML.

This makes it very difficult to create a web page from a document created in Word. In Microsoft Word, if you select "Web Page" as the file type to save in "Save As" of "File" on the ribbon, you can save it in a Web format that can be displayed in a browser at first glance. Unfortunately, the web page format saved in Word is completely useless as it is.

The reason for this is that Word attempts to reproduce the print layout specified on the screen during editing on the web page.

6.1.3 This product ignores Word layout specification in principle

In order to solve these problems, this product discards all layout specifications of documents created in Word and expresses the contents with pure HTML tags.

To master this product, you will need to first understand this basic.

As a general premise, you don't write HTML tags directly in Word, but understanding HTML tags is essential for the result of conversion from a Word document to be proper HTML. Then, you need to edit the Word document, keeping in mind that the Word styling you are currently editing will be converted to what HTML tags.

From this perspective, this chapter describes what you should be aware of when editing a Word document.

6.1.4 Things to avoid when creating Word documents

Avoid the following editing operations on the Word editing screen:

1. Adjust the start position of the line with a space character.
2. Start a new line in the middle of the line where the sentence continues.

For example, suppose you want to edit a bulleted item that spans two lines by entering a line break at the end of the first line and inserting a space at the beginning of the second line to align the beginning of the line. In this case, even if there is no problem when printing on paper or converting to PDF, but the connection of sentences will be broken when converting to HTML.

6.2 Output the HTML heading rank tag

HTML heading rank tags (h1 to h6) are tags for representing headings. From the SEO point of view, it is sometimes explained that the h1 tag that represents heading rank 1 usually represents the heading as the entire title only once at the beginning of the document. In that case, use "Heading 1" (converted to h1 tag) only once at the beginning of the Word document.

However, as HTML, there is no problem even if the h1 tag appears multiple times in the document. When creating such HTML, you can specify "Heading 1" as many times as you like.


In this product, sections are hierarchized according to the rank of the heading rank tag. When using h1 as a large heading, h2 as a middle heading, and h3 as a subheading, specify that heading 1, heading 2 and heading 3 appear in this order in a Word document. You can repeat heading 2 under heading 1 and heading 3 under heading 2.


6.2.1 Set the Word heading style

When adding headings in Word, apply the "Heading Style" built into Word. Word's "Heading Style" is available from Heading 1 to Heading 9.

The Word2HTML converter associates HTML heading rank tag h1 with heading style 1. Set heading rank tags h2 to h6 for heading styles 2 to 6.

[Notice] Depending on the theme of Word, the outline level (described later) may not be set in the heading style. When you use such a heading style in a Word document, setting the heading style does not set a heading rank tag for that paragraph.

You can determine if a paragraph has an outline level by hovering the cursor over the paragraph. Paragraphs with an outline level will have  mark on the left side of the paragraph when you hover over the cursor.

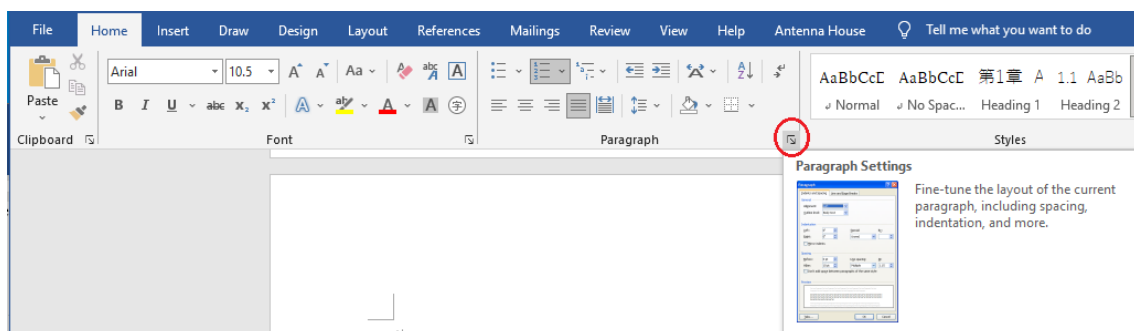
-  6.2.1 Set the Word heading style
When adding headings in Word, apply the "Heading Style" built into Word. Word's "Heading Style" is available from Heading 1 to Heading 9.

6.2.2 Set the title

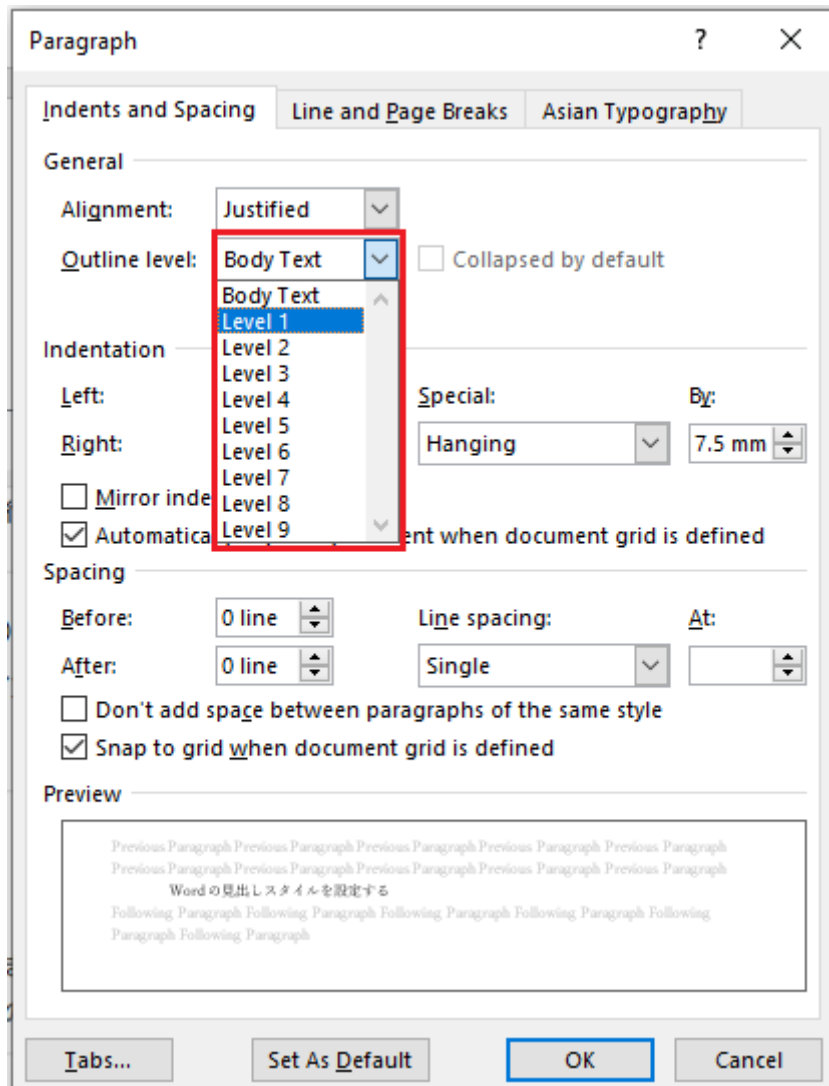
One of Word's built-in styles is "Title". The title style may have the setting of "Outline Level 1". If you apply such a title style to a paragraph in a Word document, the Word2HTML converter will set the h1 tag on that paragraph.

6.2.3 Set the paragraph outline level in Word

Word has a feature called Paragraph Outline Level, which allows you to set paragraphs in 9 levels. Also, the outline level of the outline edit paragraph is set in the "Paragraph" dialog of the "Paragraph" group on the "Home" tab of the ribbon.



The "Paragraph" dialog is displayed by clicking the arrow mark at the bottom right of the paragraph group.

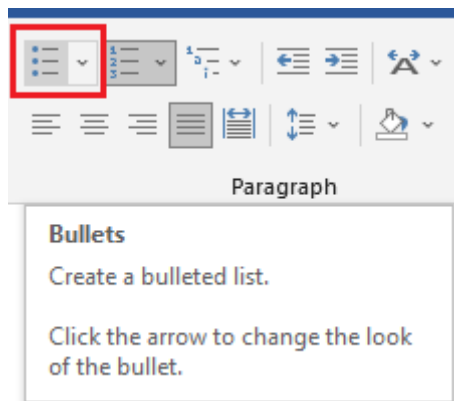


The Word2HTML converter maps outline levels 1 to 6 to HTML heading rank tags h1 to h6. In other words, a paragraph for which outline level 1 is set in Word will have HTML heading rank 1 (h1).

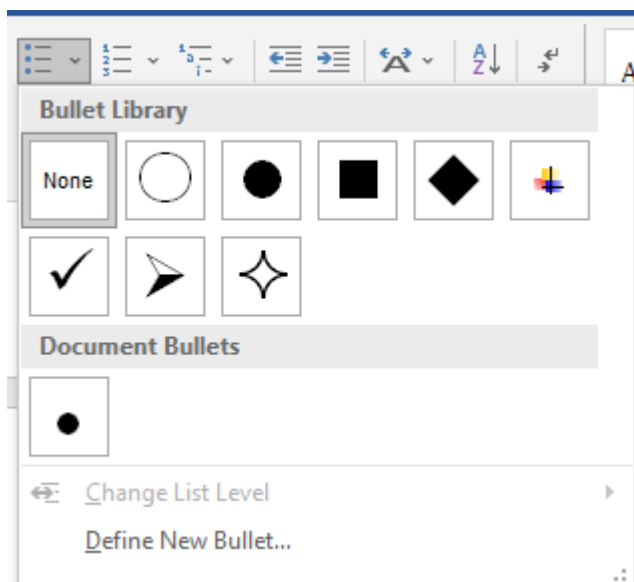
6.3 Bullets and Numbering

6.3.1 Bullets

”Bullets” in the “Home” tab on the Word ribbon creates paragraphs with symbols at the beginning of the line.



In Word, the Bullet Library allows you to change the look of the bullets.

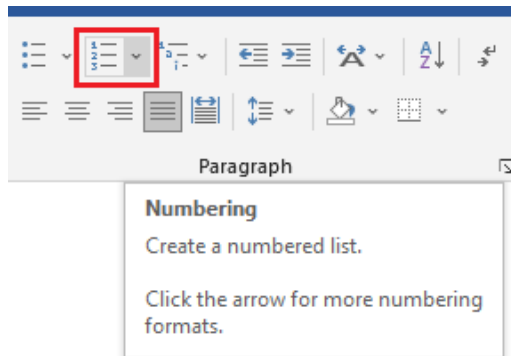


Many of these bullets are displayed using a special Word font and may not be displayed correctly in HTML. For this reason, the Word2HTML converter removes the bullets and then converts them to HTML unordered bullets. In HTML, use CSS to set bullets.

Note that blocks with bullets specified in Word look like HTML unordered bullets. However, inside Word, each item is formatted as a paragraph with a bullet. The Word2HTML converter analyzes the parts where "bullets" are set when editing Word and converts them to HTML unordered bullets. Depending on how you specify the paragraph format, bullets may not be converted to HTML unordered bullets. In this case, try changing the paragraph format.

6.3.2 Numbering

"Numbering" in the "Home" tab on the Word ribbon is a function that automatically numbers the beginning of a paragraph in the selected numbering format.



Blocks with numbering look like HTML ordered bullets.

However, Word does not have a style of numbering bullets. Paragraphs with Word numbering are saved as paragraphs with numbers at the beginning of each paragraph item.

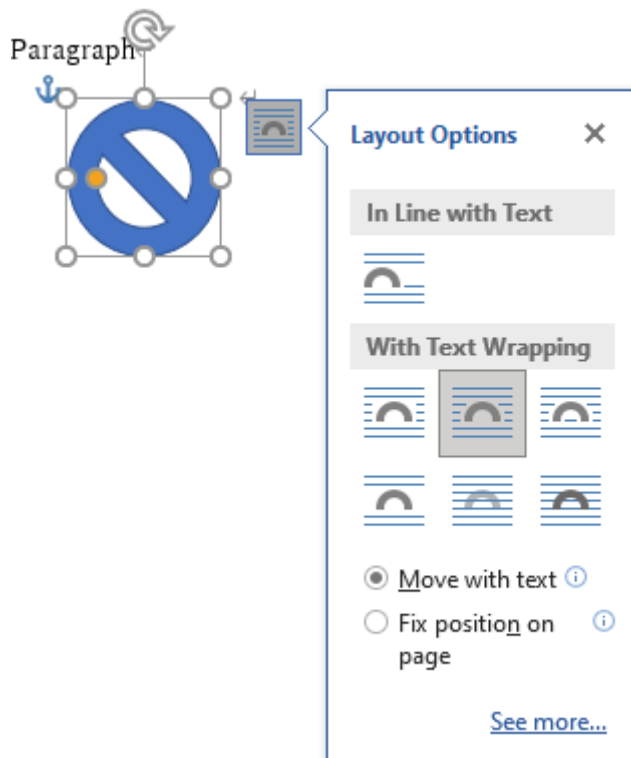
The Word2HTML converter programmatically determines whether to convert the specified "numbering paragraphs" to ordered bullets in HTML or numbering paragraphs. The judgment criteria are as follows:

- When consecutive two paragraphs with numbering are set in a Word document, the ordered bullets are set.
- When a single paragraph with numbering is set in a Word document, the number is converted to the normal character at the beginning of the paragraph.

This judgement may not always work, so if it doesn't work as expected, try editing the paragraph in Word.

6.4 Layout of shapes

The option to layout a figure in Word (Layout Options) can be selected by right-clicking the target figure to display options (see the following figure).



6.4.1 In line with text

“In line with text” is to layout an image between characters as if it were single character, and the position moves together with the characters before and after. The image with “In line with text” specified is placed between characters like (●) in HTML.

As a result of converting to HTML, the attribute of class=”inline” is set in the img tag.


6.4.2 With text wrapping

Images with “With text wrapping” specified will have an anchor mark when edited in Word.

- A. Images with anchor marks in headings and paragraphs are output just after the end tag of headings and paragraphs, and the attribute of class=”block” is set in the img tag.

In the following example, the image anchor mark is at the beginning of the heading.

1.1.1 Specification of the application to display the conversion result



When "Convert to HTML" finishes, the HTML file is displayed in the application associated with the extension html on Windows.

Converting this to HTML will output an img tag between the end tag of the heading rank and the start tag of the next paragraph, as shown below.

```
<section>
<h3>1.1.1 Specification of the application to display the conversion result</h3>

<p>When "Convert to HTML" finishes, the HTML file is displayed in the application associated with the extension html on Windows.</p>
```

In the following example, the image anchor mark is at the beginning of the paragraph.



In HTML, the image that specifies the "With Text Wrapping" is output after the end tag of the paragraph that has the anchor.

Converting this to HTML will output an img tag just after the end tag of the paragraph, as shown below.

```
<p>In HTML, the image that specifies the "With Text Wrapping" is output after the end tag of the paragraph that has the anchor.</p>

```

In this example, even though the paragraph text is at the back of the figure on the Word display, when converted to HTML, the img tag is output after the paragraph. Word places images on paper, so if an image doesn't fit well on one page, it may be placed on the next page. Even in such a case, when converting to an HTML file, the img tag will be placed after the paragraph with the anchor mark.

- B. For images with an anchor mark on a bulleted item, the img tag is output just before the end tag of the bulleted item.

In the following example, the image anchor mark is at the beginning of the first item in the bulleted list.



- ① When the docx document being edited is updated, a dialog prompting you to save the changed document is displayed before the conversion starts.
- ② When the HTML save destination folder is not set, the dialog for selecting the save destination folder is displayed. The displayed dialog has the same contents as "4.5 Changing the conversion destination folder", so see also that section.

Converting this to HTML will output the img tag **just before** the end tag of the first bulleted item, as shown below. (This is because if you put an img tag after the end tag of a bulleted item and between the start tag of the next bulleted item, an HTML syntax error will occur.)

```
<ol>
<li>When the docx document being edited is updated, a dialog prompting you to save the changed document is displayed before the conversion starts. 
</li>
<li>When the HTML save destination folder is not set, the dialog for selecting the save destination folder is displayed. The displayed dialog has the same contents as &quot;4.5 Changing the conversion destination folder&quot;;, so see also that section. </li>
</ol>
```

6.5 Blank lines and spaces in Word

The default conversion of “HTML on Word” ignores blank lines (lines with only line breaks at the beginning of lines) and page breaks in Word documents. Also, Word expects to print on paper, so if a figure or table does not fit on the page, it will be sent to the next page, leaving a large space. These spaces are ignored when converting to HTML.

You don’t need to worry about the spaces and blank lines that will be created on the layout on the Word edit screen.

6.6 Grouping of shapes and pictures

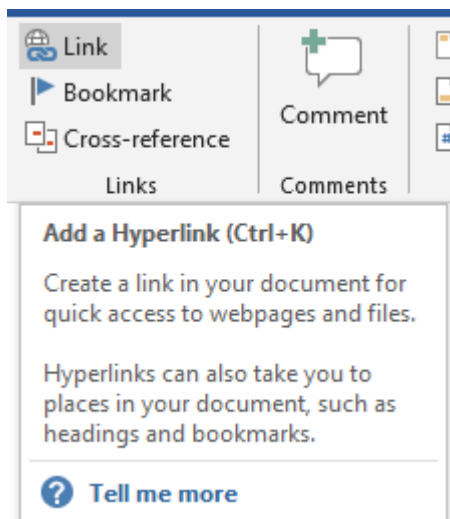
In Word, you can place shapes and images on paper. If you want to combine these shapes and images into one in HTML, group them in Word.

If you just place multiple figures in the same position on the screen of the Word document you are editing, the figures will be disjointed when converted to HTML.

6.7 Reference links

There are two ways to set a reference link in Word: "Link" and "Cross-reference" in the ribbon "Insert".

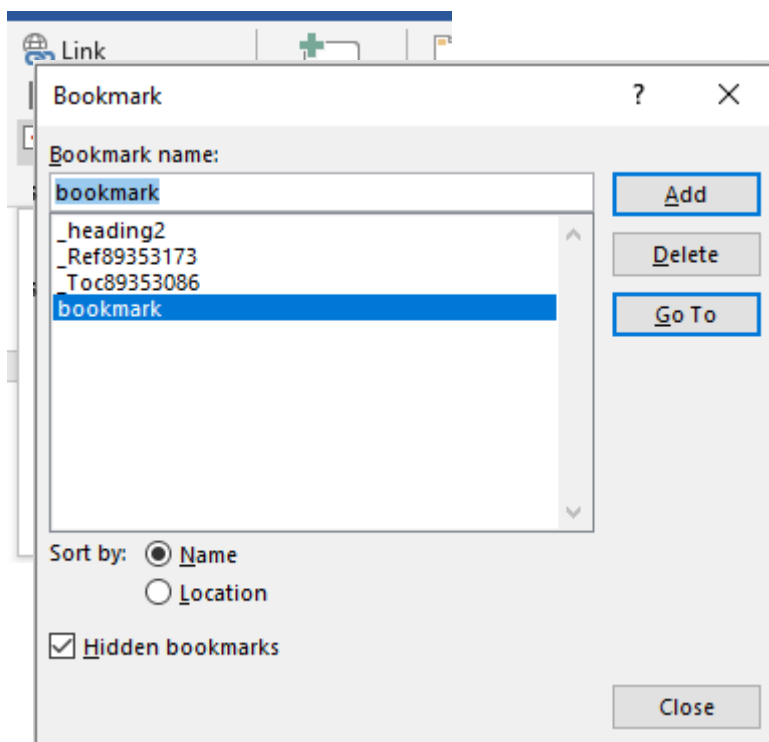
6.7.1 Link



You can set an external URL or a link to the inside of Word in "Link" on the "Insert" tab on the Word ribbon.

6.7.2 Cross-reference

“Cross-references” on the “Insert” tab on the Word ribbon allow you to set links to headings, diagrams, tables, and paragraphs with numbering inside Word.



6.7.3 Link reference

Links can be referenced to external URLs or bookmarks inside the document. Bookmarks can be added, deleted, etc. in the “Bookmark” on the "Insert" tab of the Word ribbon. The following is an example of a bookmark.

Four bookmarks are displayed in the above dialog, and the bookmark types are as follows.

- 1) Items in the table of contents created by automatic generation on Word are bookmarks starting with "_Toc".
- 2) References set in “Cross-reference” start with "_Ref" bookmarks.
- 3) "_heading 2" at the reference of the document internal link.
- 4) "bookmark" is the bookmark added in the bookmark dialog

When converted to HTML, the bookmark will be converted to ``.

[Example] A bookmark named “here” is set here.

6.8 Tables

Converts a table created by the table function of Word to an HTML table tag.

When converting a table to HTML, it does convert table cell merging, but ignores settings such as table width, border thickness, background and text alignment in the table.

You will need to specify these settings with CSS for the output HTML.

6.9 Character decoration and fonts

Of the features that can be set in the "Home" font group on the Word ribbon, "Bold", "Superscript" and "Subscript" are converted to the , <sup> and <sub> HTML tags.

Other than that, "Italic", "Underline", "Strikethrough", "Font Color", "Font" and "Font Size" are ignored during conversion.

Copyright

HTML on Word user's manual

Version: 1.2

Publishing date: July 8, 2022

Publisher: Antenna House, Inc. 2-1-6, Higashi-nihonbashi, Chuo-ku, Tokyo

Copyright ©2022 Antenna House, Inc.

This manual was edited with Microsoft Word 2019 and converted to HTML with "[HTML on Word](#)".