



Word HTML 2 Formatting Objects

WH2FO Application

Table of Content:

[WH2FO Application](#) . 1

[What is it ?](#) . 2

[How does it work ?](#) . 2

[How to ...](#) 3

[Executables](#) . 3

[How to run it using the MS Virtual Machine](#) . 3

[How to run it using the JRE](#) . 3

[How to run in the Linux/Unix Environment](#) 3

[Results](#) . 4

[Execution options](#) . 4

[How to obtain the XSL-FO file](#) . 4

[Implemented Features](#) . 5

[Versions](#) . 6

[Downloads](#) . 7

WH2FO is hosted by [SourceForge](http://www.sourceforge.net) (<http://www.sourceforge.net>)



Word HTML 2 Formatting Objects

What is it ?

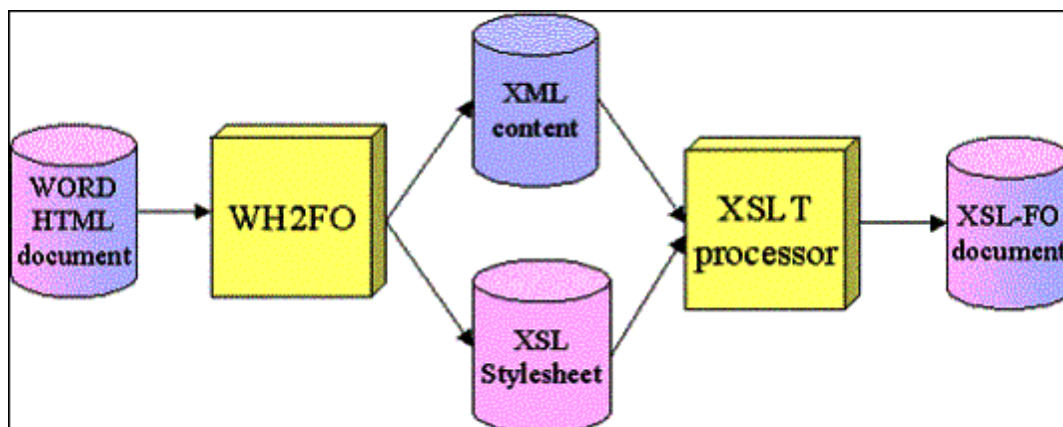
WH2FO is a **Java application** which **processes an HTML output created with Word 2000**, and **transforms it into an XML content file and an XSL style-sheet file**. From these files, a standard XSLT processor may be used to obtain a file containing only XSL-FO mark-up. You can also apply a style-sheet that converts the XML back into HTML discarding all the extra mark-up added by Word. Using an XSL-FO rendering engine, such as [FOP](#), you can also render your document into PDF.

How does it work ?

WH2FO is able to separate content from style into an HTML file generated using Word 2000. When the input has been processed you will get an XML file containing the content and an XSL-T file containing the “rules” to convert it into an XSL-FO document.

The separation between content and style gives you the possibility to re-use the style information and the possibility to re-purpose the content for different uses.

WH2FO also gives a quick way to generate XML content without knowing nothing about tags.





Word Html 2 Formatting Objects

How to ...

Executables

The following executable files are included in the distribution:

✧ **bin/WH2FO_x_y_z.jar** - WH2FO Java code

Microsoft Windows Specific files

✧ **wh2foMSV.bat** - batch file that call the MS virtual machine

✧ **wh2fo.bat** - batch file that call the java virtual machine

✧ **setpath.bat** - batch file that add the wh2fo jar into the java CLASSPATH

Unix Specific files

✧ **wh2fo.sh** - unix script for wh2fo written by Simone Merli

How to run it using the MS Virtual Machine

For Microsoft Windows users it is possible to run the compiled code using a batch file that calls the MS Virtual Machine

The MS VM (jview.exe) is normally installed with IE 5.x and can execute java code without downloading a JRE

✧ open a Shell (or a Dos session)

✧ put the jar file inside the class path using **setpath wh2fo-path**

✧ **translator htm-file-name [options]**

How to run it using the JRE

It is possible to execute the Java code directly on any OS.

The Java code must be executed from a **Java 1.x** runtime environment (JRE) as follows:

Windows users just use the wh2fo.bat : **wh2fo htm-file-name [options]**

All users:

✧ open a Shell (or a Dos session)

✧ put the jar file inside the class path

✧ **java wh2fo.apps.Translator htm-file-name [options]**

How to run in the Linux/Unix Environment

For Linux/Unix users it is possible to run the compiled code using a shell script file that calls the java run-time engine (jdk1.1)

1. run **dosunix wh2fo.sh**

2. run **chmod 744 wh2fo.sh**

3. **./wh2fo.sh htm-file-name [options]**

Results

The application will automatically generate 3 files using the same name as the input file:

1. a file for the content **htm-file-name.xml**

2. a file for the stylesheet **htm-file-name.xsl**

3. a file for the attributes (style-sets) **htm-file-nameAtts.xsl**

3



Word HTML 2 Formatting Objects

Execution options

The possible WH2FO execution options are:

- ⌘ **/INFO** - display the disclaimer
- ⌘ **/TOC** - creates the TOC into the PDF bookmarks (only with the FOP extensions)
- ⌘ **/DEBUG** - shows the parsing process (useful for the debug)

How to obtain the XSL-FO file

To create a XSL-FO file you must process the XML and XSL file using a XSLT processor, for instance you can use [XT](#) (by James Clark) that provides both a jar and a Windows executable



Word HTML 2 Formatting Objects

Implemented Features

Pagination and Layout

- ✧ page dimensions
- ✧ margins
- ✧ header and footer representation
- ✧ page orientation
- ✧ page breaks
- ✧ complex page sequences (first, odd and even; first and rest)
- ✧ page sequence breaks

Text

- ✧ headings
- ✧ all types of text defined by the user
- ✧ inline sequences (bold, italic, underline and change of font size and dimension)

Lists

- ✧ ordered (basic implementation)
- ✧ unordered (only with disc, circle and square)

Images

- ✧ imported images with shape dimensions (also cut&paste from others applications e.g. PowerPoint)

Tables

- ✧ various table layouts and sizes (with column span)
- ✧ nesting

Links

- ✧ external and internal links

Footnotes

- ✧ footnotes/endnotes (all treated like endnotes)

Table Of Content

- ✧ Table of content inside PDF bookmarks (only with FOP extensions)

Versions

- ✧ [WH2FO 0.3.0](#)
- ✧ New Tokenizer parser
- ✧ Correct empty line representation
- ✧ Complex page sequences (first+rest, first+even&odd)
- ✧ Header and footer full representation
- ✧ [WH2FO 0.2.0](#)
- ✧ Tables compatible with FOA
- ✧ More robust parsing (some bug fixed)
- ✧ [WH2FO 0.1.9 \(hosted on http://www-uk.hpl.hp.com/people/fabgia/wh2fo\)](http://www-uk.hpl.hp.com/people/fabgia/wh2fo)
- ✧ Added support for nested tables



Word HTML 2 Formatting Objects

- XSL Files compatible with FOA
- More robust parsing
- [WH2FO 0.1.7 \(hosted on http://www-uk.hpl.hp.com/people/fabgia/wh2fo\)](http://www-uk.hpl.hp.com/people/fabgia/wh2fo)
- Support for the TOC (also PDF bookmarks with FOP extensions)
- Support for footnotes/endnotes (footnotes became endnotes in HTML ...)
- Support for page numbers inside the footer
- Complete support for links (external/internal)
- Added support for column span in tables
- Added a new namespace to resolve conflict with parsers
- More robust parsing
- [WH2FO 0.1.1 \(hosted on http://www-uk.hpl.hp.com/people/fabgia/wh2fo\)](http://www-uk.hpl.hp.com/people/fabgia/wh2fo)
- Fairly complete support for table (imported tables from Excel are treated by Word like images !)
- Support for links (only external)
- [WH2FO 0.1.0 \(hosted on http://www-uk.hpl.hp.com/people/fabgia/wh2fo\)](http://www-uk.hpl.hp.com/people/fabgia/wh2fo)
- More robust parsing (support document created with different tools and imported into Word 2000)
- Support for lists expressed like paragraphs (only style)
- Images improvement with re-dimensioning inside the shapes
- Master names changed to be compatible with FOP
- [WH2FO 0.0.9 \(hosted on http://www-uk.hpl.hp.com/people/fabgia/wh2fo\)](http://www-uk.hpl.hp.com/people/fabgia/wh2fo)
- Support for pagination
- Support for images
- Support for lists
- Support for text and headings

Downloads

File	Version	Type	Size
WH2FO Distribution	0.3.0	Zip	435KB
Examples	0.2.0	Zip	4.4MB
Java Docs	0.3.0	Zip	506KB
UML Diagrams	0.3.0	Zip	214KB
Parser Sequence State	0.3.0	Zip	120KB