**WWN Development Document**

WWN Testing : tests performed, and Word docs used

Word’s Navigation pane shows the table-of-contents (View : Show : Navigation pane).

* Contents:
  + Documentation of testing performed, includes
    - Regression tests that could be used for WWN enhancements
    - Browsers tested

This document was created by the WWN author for his own use in developing WWN. It is included in the WWN repo, as other developers may find it useful.

# VScode Python testing

* See: D:\Documents\Professional-projects\Reference-info\Software-development\Python\python--language--notes.docx

# Cerberus

* undefined key:value specified
* 6/26 code review and test of all schema defininitions and code that processes the parameter-file, esp looking for needed dictionary key checks before accessing key, and processing of required and optional key/value pairs, e.g., omitted keys and/or values

# Testing to-do

* TOC in different locations in the doc
  + multiple TOCs
* Testing: linters for css and html
  + https://nvuillam.github.io/mega-linter/descriptors/css/
  + <https://nvuillam.github.io/mega-linter/descriptors/html/>
* GitHub - mwilliamson/python-mammoth: Convert Word documents (.docx files) to HTML
  + <https://github.com/mwilliamson/python-mammoth>
  + *The following features are currently supported:*
    - Headings.
    - Lists.
    - Customisable mapping from your own docx styles to HTML. For instance, you could convert WarningHeading to h1.warning by providing an appropriate style mapping.
    - Tables. The formatting of the table itself, such as borders, is currently ignored, but the formatting of the text is treated the same as in the rest of the document.
    - Footnotes and endnotes.
    - Images.
    - Bold, italics, underlines, strikethrough, superscript and subscript.
    - Links.
    - Line breaks.
    - Text boxes. The contents of the text box are treated as a separate paragraph that appears after the paragraph containing the text box.
    - Comments.
* HTML validator
  + <https://rocketvalidator.com/html-validation>

# Original - test files

* Chrome web-page downloader:
  + Tab Save - Chrome Web Store
    - <https://chrome.google.com/webstore/detail/tab-save/lkngoeaeclaebmpkgapchgjdbaekacki>

## From Internet search of Word files

### python-mammoth

* Useful set of Word files used for testing Word -> HTML features

## From Internet search of Word-generated HTML files

### Can be processed by BeautifulSoup

* Notes
  + In testing the generated HTML, links that are relative to the web-server don't work, e.g., links to images
  + Testing is simple: just compare source web-page with the generated web-page
* <https://www.bible.ca/canon.htm>
  + <meta name=Generator content="Microsoft Word 15 (filtered)">
  + Uses color:white, but properly
* <https://static.sparemin.com/static/doc/201808291932/media-use-terms.html>

<meta http-equiv=Content-Type content="text/html; charset=windows-1252">

<meta name=Generator content="Microsoft Word 15 (filtered)">

### Cannot be processed by BeautifulSoup (decode errors)

* <http://dandanplay.com/bsintro.htm>
  + <meta name=Generator content="Microsoft Word 15 (filtered)">
  + Asian text, bullets, TOC, tables, hyper links,
  + title:

<p class=MsoTitle style='border:none;padding:0cm'><span style='font-family:

宋体'>弹弹</span><span lang=EN-US>play </span><span style='font-family:宋体'>展会版使用文档</span></p>

* <https://www.bible.in.ua/security/>

<html>

<head>

<meta http-equiv=Content-Type content="text/html; charset=windows-1251">

<meta name=Generator content="Microsoft Word 14 (filtered)">

<style>

<!--

* <http://izapya.com/policy_en.html>
* <https://klein.cz/>
  + Could be a useful test.
  + Two div sections, one nested
  + Uses arrow symbols for list items

<meta http-equiv=Content-Type content="text/html; charset=utf-8">

<meta name=Generator content="Microsoft Word 15 (filtered)">

* <http://nastooh.ir/docs/portfolio/newsagency-analytics-report.htm>
  + Could be a useful test.
  + Multiple div sections, including: WordSection2, WordSection3

<meta name=Generator content="Microsoft Word 15 (filtered)">

# Regresssion-test test-files

* How do I find out what version of WORD was used to create a Word document?
  + <https://answers.microsoft.com/en-us/msoffice/forum/msoffice_word-mso_win10-mso_2013_release/how-do-i-find-out-what-version-of-word-was-used-to/85334915-6881-4efc-8312-3712f268e124>

*Print ActiveDocument.CompatibilityMode*

*A version number will appear on the next line, where*

*11 = Word 2003*

*12 = Word 2007*

*14 = Word 2010*

*15 = Word 2013*

* Google search
  + filetype:docx "table of contents" site:www.microsoft.com/en-us/research/wp-content/uploads/
    - also: filetype:docx

## Summary of features to test

### Document body

* HTML modifications and fixes
  + Unordered lists (multi-level)
    - 2 symbols replaced, to be viewable in Firefox
    - Indentation fixed
    - Spacing after bullet symbol fixed
  + Ordered lists (multi-level)
    - Indentation fixed
    - Spacing after bullet symbol fixed
  + White-colored text
    - Removed style "color:white"
    - Test files that have style "color:white" are noted as such.

### Navigation pane

* HTML modification to table of contents:
  + Color does not change when clicked
  + Not underlined

### Header

### Trailer

## Word files from the Internet

### Example formatting for Kindle

* Document description: Provides an example of Word formatting for a Kindle book. Created in the UK, using Word 2003.
* Source:
  + <https://ebookconversion.paulbrookes.net/wp-content/uploads/2011/09/basicstyles.doc>
    - That Word document's link is on this web-page. The link text is: "basicstyles.doc".
      * <https://ebookconversion.paulbrookes.net/tag/formatting-2/>
  + Local file: example-book-formatting--original.doc
* Bibliographic info:
  + Title: Sequential File Programming Patterns and Performance with .NET
  + Date: 2004
  + Pages: 15
* Testing-related features:
  + Testing relevance: example of the typical layout for a book that will be converted to Kindle format; created with Word 2003
  + Word-file features (.doc)
    - Created by Word 2003
    - TOC
      * No existing TOC
      * Current TOC-format setting: Word's default TOC styles
    - Other features: heading-levels 1 to 3, ordered and unordered lists
  + Word HTML features (.html)
    - Some paragraphs have span-tags with attributes "lang=DE" and "lang=EN-GB".
      * In fixing the Word HTML, the HTML parsing should be resiliant enough to ignore these tags and make needed corrections.

<p class="MsoToc1"><span lang="DE"><span class="MsoHyperlink tocAnchor"><a class="tocAnchor" href="#\_Toc75212936"><span lang="EN-GB">CHAPTER

I</span></a></span></span></p>

* Variations of the document that I created, for testing:
  + Local file-name: example-book-formatting--added-features.doc
    - Added a multi-level ordered-list. This is for testing the code that fixes the Word HTML, for ordered-lists.
* Program tests, using file: example-book-formatting--added-features.doc
  + Tests for the program: generate\_word\_html.docm
    - Verify: TOC created
  + Tests for the program: create\_web\_page.py
    - Inspect created web-page for the following. All should be OK.
      * TOC levels and formatting
      * HTML fixes
      * Text formatting vs Word Doc

### Microsoft technical-report on programming

* Description: Microsoft programming technical-report, using Word 2003
* Source:
  + <https://www.microsoft.com/en-us/research/wp-content/uploads/2004/12/tr-2004-136.doc>
  + Local file-name: MS-tech-report--Sequential File--original.doc
* Bibliographic info:
  + Title: Sequential File Programming Patterns and Performance with .NET
  + Date: 2004
  + Pages: 15
* Testing-related features:
  + Testing-relevance: Microsoft-published technical-report on programming; 14-pages; Word 2003
  + Word-file features
    - Created by Word 2003
    - TOC
      * Existing TOC after title
      * Current TOC-format setting: Word's default TOC styles
    - Other features: tables, text-boxes, figures, source-code, API specs, only heading-level one is used
  + Word HTML-file features
    - Document formatting problems due to problems with Word's HTML-generation
      * Blurred images, and problems with page-layout of text-boxes
      * These problems can be fixed by changing the Word document. This is described elsewhere in the system documentation.
      * These problems can be ignored in the system function-test
* Variations of the document that I created, for testing:
  + Local file-name: MS-tech-report--Sequential File--fixed.doc
    - Contains proof-of-concept fixes for the document-formatting problems due to problems with Word's HTML-generation
* Program tests, using file: MS-tech-report--Sequential File--original.doc
  + generate\_word\_html.docm
    - Verify: TOC created
    - Note: the Word-file has Track Changes on, and the VBA code will turn it off
  + create\_web\_page.py
    - Inspect created web-page for the following. All should be OK:
      * TOC levels and formatting
      * HTML fixes
      * Text formatting vs Word Doc (except for problems due to Word's HTML-generation)

### Microsoft paper on programming

* Description: Microsoft publication-quality paper on programming, Word 2013
* Source:
  + <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Final-DengYu-NOW-Book-DeepLearn2013-ForLecturesJuly2.docx>
  + Local file-name: MS-tutorial--Deep Learning --original.docx
* Bibliographic info:
  + Title: Deep Learning For Signal And Information Processing
  + Date: 2013
  + Pages: 97
* Testing-releated features
  + Testing-relevance: long programming paper for publication, Word 2013
  + Word-file features
    - Created by Word 2013
    - TOC
      * Existing TOC after title
      * Current TOC-format setting: Word's default TOC styles
    - Other features: heading levels 1-2, lots of equations
  + Word HTML-file features
    - Document formatting problems due to problems with Word's HTML-generation
      * Equations that are centered in the Word-file might not be centered in the HTML
      * Excessive spacing between some lines (e.g., a paragraph after one of the equations)
      * For some bullets, there is too much space between the bullet-symbol and the text. This appears to be caused by the text format-setting that forces the word at the end of the line to be at the right margin.
      * Figure at the end of the document does not appear in the HTML.
    - It appears that these problems can be fixed by changing the Word document. This is described elsewhere in the system documentation.
    - These problems can be ignored in the system function-test
* Variations of the document that I created, for testing:
  + Local file-name: MS-tutorial--Deep Learning--fixed.docx
    - Contains proof-of-concept fixes for the document-formatting problems due to problems with Word's HTML-generation
* Program tests, using file: MS-tutorial--Deep Learning --original.docx
  + generate\_word\_html.docm
    - Verify: TOC created
  + create\_web\_page.py
    - Inspect created web-page for the following. All should be OK:
      * TOC levels and formatting
      * HTML fixes
      * Text formatting vs Word Doc (except for problems due to Word's HTML-generation)

### University computer-science lab instructions

* Description: university computer-lab instruction manual
* Source:
  + <http://virgil.azwestern.edu/~cvb/CIS120/Book%20Notes/Chapter.04.docx>
  + Local file-name: computer-concepts--original.html
* Bibliographic info:
  + Title: Computer Concepts : Chapter Four: Operating Systems and File Management
  + Date: 2011
  + Pages: 20
* Testing-related features:
  + Testing-relevance: university computer-lab instruction manual
  + Word-file features
    - Created by Word 2003
    - TOC
      * Document has a TOC, but it's not a Word-generated TOC. It was made by the author.
      * Current TOC-format setting: Word's default TOC styles
    - Other features:
      * Uses headings 1 and 8; lots of: tables, lists, highlighting;
      * The unconventional formatting indicates the author had limited understanding of proper Word use
  + Word HTML-file features
    - Body has 2 instances of sytle "color:white". They are intentional, and in the top H1 headers, with gray background.
    - In the Word-file, much of the text starts left of the left-margin. In the generated HTML, text that is left of the left-margin is not visible. This formatting problem appears to be fixable, by changing the Word styles that specify the left-alignment incorrectly.
    - In the Word-file, there are two spurious section-breaks that cause formatting problems in the HTML. In particular, they created extra blank lines for some of the questions. These problems are fixable, by deleting the spurious section-breaks.
* Variations of the document that I created, for testing:
  + Local file-name: computer-concepts--fixed.docx
    - Removed the 2 spurious section breaks that were causing formatting problems.
* Program tests, using file: computer-concepts--original.html
  + generate\_word\_html.docm
    - Verify: TOC creation and formatting
    - The TOC formatting has problems, but due to the document's flawed formatting
      * Heading levels used are 1 and 8
      * Many of the heading instances have one-off formatting changes made to them
    - When TOCs are generated by Word, one-off heading changes are incorporated into the TOC
      * Described here:
        + <http://wordfaqs.ssbarnhill.com/TOCTips.htm>
        + Section: "Effect of direct formatting"
      * This appears to be why the TOC is bold
  + create\_web\_page.py
    - Inspect created web-page for the following. All should be OK:
      * TOC levels and formatting
        + Each TOC entry has a bold tag (<b>). The document's headings have custom formatting, which Word incorporates in the TOC, i.e., bold

<p class=MsoToc1><span class=MsoHyperlink><a href="#\_Toc74846524"><b><span

style='font-family:"Sylfaen",serif'>Computer Concepts</span></b></a></span></p>

* + - * HTML fixes
      * Text formatting vs Word Doc (except for the formatting problems described above)

### University Python computer-lab instructions

* Description: University Python computer-lab instructions, with related formatting
* Source:
  + <https://www.usna.edu/Users/cs/bilzor/ic411/calendar.php?key=1bbdbf5bf81ce676f97fe0b6135d391fbbef1daa&type=class&event=17>
  + Local file-name: python-lab-instructions--original.docx
* Bibliographic info:
  + Title: Python Lab
  + Date: 2020
  + Pages: 2
* Testing-related features:
  + Testing-relevance: standard formatting for a programming-related doc
  + Word-file features
    - Created by Word 2010
    - TOC
      * There's no headings, so no TOC is possible
    - Other features:
      * Standard simple-features used, no lists
      * Spurious underlines
        + Before and after "Part 0: Overview" is a paragraph with one space, and underline formatting on.
        + The underline does not show-up in Word, but it does in the HTML
  + Word HTML-file features
    - Nothing noteworthy
* Program tests, using file: python-lab-instructions--original.docx
  + generate\_word\_html.docm
    - Verify: No TOC created
  + create\_web\_page.py
    - Inspect created web-page for the following. Should be OK.
      * Text formatting vs Word Doc
        + Note: the 2 spurious underlines are OK, in the HTML

### University physics-lab instructions

* Description: University physics-lab instructions. Contains python code, equations, ordered-list.
* Source:
  + <https://www.niu.edu/brown/_pdf/physics374_spring2021/l4-1-21.docx>
  + Local file-name: physics-tutorial--original.docx
* Bibliographic info:
  + Title: Physics 374 – Junior Physics Lab
  + Date: Spring 2021
  + Pages: 5
* Testing-related features:
  + Testing-relevance:
  + Word-file features
    - Created by Word 2013
    - TOC
      * There's no headings, so no TOC is possible
    - Other features:
      * Lots of equations; ordered-list and its list-symbols use letters
      * Some of the blank lines have underline formatting turned-on. It's not visible in Word, but is visible in HTML
  + Word HTML-file features
    - One misasligned equation.
      * It uses the old Word equation editor. The "(3)" for the equation was additional text
      * I fixed it by using Word's feature to upgrade the equation for use with the current Word equation-editor (right click on equation). Also, I edited the equation to put the (3) in with the equation.
    - For the ordered list, the text after the list-symbol is slightly misaligned. It is indented too much, by about 1 character-width. This is a tolerable error, for now.
* Variations of the document that I created, for testing:
  + Local file-name: physics-tutorial--fixed.docx
    - Has the fix for the misaligned equation
* Program tests, using file: physics-tutorial--original.docx
  + generate\_word\_html.docm
    - Verify: No TOC created
  + create\_web\_page.py
    - Inspect created web-page for the following. All should be OK, with the exceptions described above.
      * HTML fixes
      * Text formatting vs Word Doc

### University Python tutorial

* Description: University Python tutorial, with related formatting. 23 pages
* Source:
  + <http://brahe.canisius.edu/~meyer/PYTHONICTEMPLE/Very%20quick%20tutorial%20on%20Python.docx>
  + Local file-name: python-tutorial--original.docx
* Bibliographic info:
  + Title: Very quick tutorial on Python
  + Date: 2014
  + Pages: 23
* Testing-related features:
  + Testing-relevance: standard formatting for a programming-related doc
  + Word-file features
    - Created by Word 2010
    - TOC
      * There's no headings, so no TOC is possible
    - Other features:
      * Spurious underlines
        + There are some paragraphs with one space, and underline formatting on. The underline does not show-up in Word, but it does in the HTML
  + Word HTML-file features
    - Minor formatting problems in the spacing between header and paragraph, but it appears they can be fixed in the Word file.
* Program tests, using file: python-tutorial--original.docx
  + generate\_word\_html.docm
    - Verify: No TOC created
  + create\_web\_page.py
    - Inspect created web-page for the following. All should be OK, with the exceptions described above.
      * HTML fixes
      * Text formatting vs Word Doc

### Agricultural technical-report

* Description: agricultural technical-report from Cornell Univ.
* Source:
  + <http://sbdocs.psur.cornell.edu/download/attachments/4784135/OptionsDAndOUserGuide.docx?version=1&modificationDate=1449157135050&api=v2>
  + Local file-name: pesticide-report--original.docx
* Bibliographic info:
  + Title: Options D and O User Guide
  + Date: 2015
  + Pages: 12
* Testing-related features:
  + Testing relevance: tech-report with creative and unconventional hacked formatting
  + Word-file features
    - Created by Word 2010
    - TOC
      * Current TOC-format setting: Word's default TOC styles
      * There's an existing TOC. It does not include the spurious heading at the top of the file, hidden under the graphic.
    - Other features:
      * Has table of contents, figures, unconventional hacked formatting
      * When opening the Word file, the formatting for first page initially looks flawed, as it opens in Web Layout, but it looks OK when changed to Print Layout
  + Word HTML-file features
    - The TOC gets its formatting from the author-modified headings, e.g., color
      * This is normal for Word's HTML TOC-generation
    - The graphic's formatting becomes flawed, but this problem can be ignored. It should be possible to fix the formatting.
* Variations of the document that I created, for testing:
  + Local file-name: pesticide-report--fixed.docx
    - Deleted spurious heading at the top of the file, under the graphic.
    - Updated existing table-of-contents
* Program tests, using file: pesticide-report--fixed.docx
  + Tests for the program: generate\_word\_html.docm
    - Verify: TOC created OK
  + Tests for the program: create\_web\_page.py
    - Inspect created web-page for the following. All should be OK, except for the problems described above.
      * TOC levels and formatting
      * HTML fixes
      * Text formatting vs Word Doc

### Agricultural software-installation guide

* Description: Agricultural software-installation guide from Cornell Univ.
* Bibliographic info:
* Source:
  + <http://sbdocs.psur.cornell.edu/download/attachments/4784135/PRLSoftwareInstallationGuide.docx?version=1&modificationDate=1449156951423&api=v2>
  + Local file-name: software-install-guide--original.docx
* Bibliographic info:
  + Title: PRL Software Installation Guide
  + Date: 2015
  + Pages: 5
* Description:
  + Testing-relevance: software documentation from an organization whose mission is not software development
  + Notes:
    - Apparently produced by the same organization that created the test-file "Agricultural technical-report from Cornell Univ."
    - Both documents have similar formatting, and formatting bugs, e.g., unused header under the graphic at the top of the document.
  + Word-file features
    - Created by Word 2010
    - TOC
      * Current TOC-format setting: Word's default TOC styles
      * There's an existing TOC. It does not include the spurious heading at the top of the file, hidden under the graphic.
    - Other features:
      * Has table of contents, figures, unconventional hacked formatting
      * When opening the Word file, the formatting for first page initially looks flawed, as it opens in Web Layout, but it looks OK when changed to Print Layout
      * Spurious bullets can be seen in the HTML view. They can be found in the Word file, via the Word Style Manager.
  + Word HTML-file features
    - The TOC gets its formatting from the author-modified headings, e.g., color
      * This is normal for Word's HTML TOC-generation
    - The graphic's formatting becomes flawed, but this problem can be ignored. It should be possible to fix the formatting.
* Variations of the document that I created, for testing:
  + Local file-name: software-install-guide--fixed.docx
    - Deleted spurious heading at the top of the file, under the graphic.
    - Updated existing table-of-contents
    - Deleted supurious bullets
* Program tests, using file: software-install-guide--fixed.docx
  + Tests for the program: generate\_word\_html.docm
    - Verify: TOC created OK
      * The TOC gets formatting from modified headings, e.g., color
      * The graphic's formatting becomes flawed, but this problem can be ignored. The graphics can be reformatted so that TOC can be added without problem.
  + Tests for the program: create\_web\_page.py
    - Inspect created web-page for the following. All should be OK, except for the problems described above.
      * TOC levels and formatting
      * HTML fixes
      * Text formatting vs Word Doc

# Browser testing

## CSS libraries

Browser testing (on 3/2021):

\* Windows 10 with:

\* Firefox 87.0

\* IE version 2004, build 19041

\* Edge version 89.0.774.54

\* Chrome version 89.0.4389.82

\* Ubuntu

\* Firefox 86.0.1

\* Mac OS (the testing service just showed a screenshot)

\* Firefox 82.0

\* Chrome 87.0

\* jQuery

https://jquery.com/browser-support/

#Current Active Support

Desktop

Chrome: (Current - 1) and Current

Edge: (Current - 1) and Current

Firefox: (Current - 1) and Current, ESR

Internet Explorer: 9+

Safari: (Current - 1) and Current

Opera: Current

#Mobile

Stock browser on Android 4.0+[1]

Safari on iOS 7+

\* jQuery UI

https://jqueryui.com/browser-support/#current-active-support

jQuery UI 1.12.x supports the following browsers:

Chrome: (Current - 1) or Current

Firefox: (Current - 1) or Current

Safari: (Current - 1) or Current

Opera: (Current - 1) or Current

IE: 11

Edge: (Current - 1) or Current

https://jqueryui.com/browser-support/#current-active-support