Advanced Remediation of Ancient Word Documents:
Your Worst Fears Relieved[1]

Recognizing an Ancient Document

- Often, in an old file format: `.doc` instead of `.docx`.
- Incomplete heading structure.
- Styles misused, if used at all.
- Lots of formatting.
- Lots of empty paragraphs, lines, tabs, and spaces used to position text.

Definitions (for Microsoft Word)

format (n) the appearance of text. *Also* formatting.
style (n) a label for a role for text, to which we can associate formatting, outline level, list formats, language, and perhaps other attributes.

Two Basic Techniques

The Nuclear Option

Demonstrated in How to Make an Old Document Accessible[2]:

1. Print out the document.
2. Select everything in the Word file.
3. Clear all the formatting.
4. Apply the most commonly used style (usually Body Text)
5. Going through the document page by page, and even line by line, apply the correct style to each paragraph, phrase, word, or character based on its role.

The Archeological Dig

1. Adjust the **Styles** pane to display formatting in use.
2. Open **Reveal Formatting** pane
3. Format by format:
   a. Determine the role of the text.
   b. Clear the formatting.

---

[1] Presented at Knowbility's John Slatin Access U May 15, 2018, 3:45 p.m., Room 116, Trustee Hall. By Cliff Tyllick, EIR Accessibility Coordinator, Texas HHS.
[2] URL of the tutorial itself:
https://gov.texas.gov/organization/disabilities/accessibledocs2013-16#wordolddocuments

c. Apply the correct style.

Build Your Toolkit

Check File Options

1. **File** > **Options** > **Advanced** > **Editing options**.
2. Be sure "Keep track of formatting" (11th box) is checked.

Determine Your Basic List of Styles

- Don't worry about their formatting (yet).
- Don't agonize over making it complete. (Can modify as you go.)
- Cover the basic roles:
  ‣ Headings
  ‣ Body Text
  ‣ List styles
  ‣ Character styles
    ◊ Emphasis
    ◊ Strong
    ◊ Book Title
    ◊ Any specific role a phrase, role, or character can have in your document

Make the Styles Pane Show Formatting instead of Styles

1. Click **Options…**.
2. Set **Select styles to show:** to **In use**.
3. For **Select how list is sorted:** try them all; pick your favorite.

Open Reveal Formatting

Through the **Styles** pane:

1. Open **Style Inspector** (middle of the 3 icons at bottom).
2. Open **Reveal Formatting** (icon at bottom left).
3. Your choice: keep **Style Inspector** open, or close it.

Or by improving your dashboard:

1. **File** > **Options** > **Quick Access Toolbar**.
2. Set **Choose commands from:** to **All Commands**.
3. Choose **Reveal Formatting**.
4. Click **Add**.
5. **OK**.

6. In the **Quick Access Toolbar**, click the new command.

Start the Archaeology

To direct our efforts, print out a copy of the original document (or save as a different file).

Surface Survey: What Can We Learn Quickly?

To focus our activities, use **Select formatting to show as styles:**

Word lets us choose three levels of formatting to review:

- Paragraph level formatting
- Font formatting
- Bullet and Numbering formatting

Each can give overwhelming levels of detail. Let's see what each shows us:

Font Formatting: Individual Characters up to Short Blocks of Text

- Less than a paragraph in scope
- Mainly can be thought of as affecting individual characters:
  ‣ Font face (Verdana, Arial, Georgia, Times New Roman…)
  ‣ Font size (in points)
  ‣ Character spacing
- When the Word document was extracted from a PDF, there can be dozens of variations of font size and character spacing. (Extraction isn't perfect.)

Paragraph Formatting:

- Applied to an entire paragraph.
- Typical features of a paragraph:
  ‣ Alignment
  ‣ Indentation (both right and left)
  ‣ Space before and after
  ‣ Line spacing
  ‣ Outline level (structure!)
- **Plus font formatting,** when applied to a whole paragraph.
- When the Word document was extracted from a PDF, there can be dozens of variations of each of these. (Extraction isn't perfect.)

Bullet and Numbering formatting

- Surprising how many variations produce bullets or numbers that look alike— variations in font face, point size and even the character itself.
- If you use list styles (List Bullet, List Number, etc.) instead of list formatting (from buttons on the ribbon), lists are easily fixed and managed.

- Usually fewer instances than Font or Paragraph formatting.

Dig In: One Layer at a Time

I have found a specific approach works well. As you get more comfortable with Word's features, you might try your own variations.

First Layer: Clear the Debris.

1. Remove the weirdest of font formatting:
   a. Select all text in the document.
   b. Right-click on the selected text.
   c. From the context menu, choose Font.
   d. In the Font dialog box, on the Font tab, choose "+Body." (It will be easy to change the headings back later.)
   e. Click the Advanced tab.
   f. For Character Spacing, set these values:
      i. Scale = 100%
      ii. Spacing = Normal
      iii. If there should be no superscript or subscript, Position = Normal
2. Fix Bullet & Numbering formatting:
   a. Set Styles pane Options:
      i. Select styles to show: In use.
      ii. Select formatting to show as styles: Bullet and numbering formatting (only).
   b. Choose formats one by one. For each format:
      i. Decide what list style it should have.
      ii. Set it to that style.
      iii. If necessary, demote it to the appropriate level.
3. Remove the weirdest of all Paragraph formatting:
   a. In Styles pane Options, change to Paragraph formatting.
   b. Skim through formats displayed.
   c. Right-click on any that are way off base—for example:
      i. Anything + Justified
      ii. Anything + unreasonable indentation—for example, "Left indent: 0.4 pt.")
   d. Choose "Delete [this format]."

Second Layer: Find the Structure

1. Start with Paragraph formatting:
   a. One by one, select each remaining format:
      i. Right-click on the format.

  ii. Choose "Select All [however many] Instances."
 b. Decide whether this is the correct format for a specific role:
  i. If so:
   (1) Right-click on the appropriate style.
   (2) Choose "Update [this style] to Match Selection."
    Note: You might realize that you need to create a new style. That's OK.
  ii. If not:
   (1) Clear formatting.
   (2) Apply the correct style, if there is one.
    Note: You will almost certainly be applying one style to two or more formats. In other words, the original formatting is likely to be inconsistent.

2. Check for correct outline in Navigation pane.
3. Check each Heading style in Style pane:
 a. Need to change back to "(Headings)" font?
 b. If so, right-click, choose Modify, and do it.

Curate for Display: Font formatting

Change Style pane Options to show Font formatting.

- If it should be a paragraph style:
  - ‣ Clear formatting.
  - ‣ Apply correct paragraph style.
- If should be a character style:
  - ‣ Either update that style to match selection.
  - ‣ Or create a new style for that purpose.
  - ‣ Or clear formatting and apply the correct style—for example:
    - ◊ Switch *Italic* to *Emphasis*.
    - ◊ Switch **Bold** to **Strong.**

Finishing Up: Compare the Document to the Original

Make any final tweaks you need to. Especially look for:

- Paragraphs with unusual (but correct) indentation.
- Font formatting that should be there, but was lost in cleanup.