

Microsystems
377 E. Butterfield Road, Suite 910
Lombard, IL 60148
Phone 630-598-1100
Fax 630-598-9520
www.microsystems.com



“Are You SEC-Ready?”

*Best practices, processes, techniques & tools
to accelerate your EDGAR filing cycle*

**by Sherry Kappel
Sr. VP, Chief Innovation Officer**

March 18, 2004

Contents

Contents 2

The Dizzying Cycle ‘Gets Real’ 3

So, what’s changed? 3

The Questions, The Indicators — The Answers..... 3

The Objective 4

What steps should we be taking? 5

How much time should we allocate to preparatory cleanup? 5

From EDGAR — Best Practices for Repurposing its Content..... 6

Manual Cleanup from Modernized EDGAR (HTML) 6

Manual Cleanup Process from EDGAR (ASCII) 12

Manual Cleanup Process from EDGAR (Supplemental PDFs) 17

An image-based PDF 17

A Text-Based PDF 18

Verifying Original PDF vs. Resulting OCR'ed & Cleaned Copy 21

To EDGAR — Creating Modernized HTML from Word..... 21

Manual Process to Create Modernized EDGAR (HTML) from Word . 21

Creating the HTML file 22

The DocXtools SEC Accelerator..... 23

When Repurposing from EDGAR:..... 23

When Inserting New Content: 23

When filing to EDGAR: 23

About Microsystems 23

The Dizzying Cycle 'Gets Real'

EDGAR¹ — founding father of all e-filing initiatives — opened its virtual doors way back in 1992. Since then, law firms have invested incalculable amounts of time:

- retrieving previously-posted content from EDGAR
- converting and/or cleaning up the downloaded files
- assembling new content from myriad sources
- editing and adding newly-authored content
- reconverting and preparing for submission back into EDGAR

So, what's changed?

That's exactly the point: nothing much, at all. Everywhere we looked, document producers' skills, the tools made available to them in support of the endeavor, and most startling of all: the internal and external-facing collaboration processes in use remain wedded to a non-electronic era long gone by.

Things were *so much* simpler back when the processes were first established — with organizational designs well-poised to ensure accuracy and accelerated delivery. But here, in 2004, where things are *much more* complicated — looming little initiatives like Sarbanes-Oxley² being just one of them — we've reduced our support staff, denied access to enabling tools, and accepted that our attorneys needn't attend training at all.

The Questions, The Indicators — The Answers

The Questions:

Have you reviewed your processes lately?

So, when *was* the last time you revisited your EDGAR filing processes? How challenging is the intake of your client's financial data? Is the quality of your editable result one you're proud to reveal to your client, or do you find yourself electronically masking its less-than-flattering construction by scanning or PDF-ing? Would your internal workflow withstand external scrutiny? As for the deliverable, are you still filing to EDGAR in ASCII text, or have you moved your 10-K's and 10-Q's forward into HTML instead?

¹ EDGAR stands for the Electronic Data Gathering, Analysis and Retrieval System of the United States Securities and Exchange Commission.

² The Accounting Reform and Investor Protection Act of 2002, often referred to as the Sarbanes-Oxley Act for the legislators who sponsored it, is directed at improving the accuracy of financial statements and ensuring full disclosure of such information. The Sarbanes-Oxley Act seeks to protect the objectivity of securities analysts and to improve the SEC's resources and oversight functions.

The Indicators:
Accelerate and improve your processes and tools

For as lucrative a venture as Corporate and Securities work can be to a law firm, the indicators speak loud and clear: improve efficiencies, reduce both your costs and the costs to your clients, accelerate your ability to deliver, and collaborate *expertly*.

If you no longer employ proofreaders (most firms don't), have you provided readily-available technology people know how — and when — to use? With the amount of content drawn from disparate electronic sources, and the monkey-wrenching required to get it into the EDGAR-delivered document, do you know whether critical statistical data is being retyped?

The Answers:
Skill up, Tool up, and Shape up your internal processes

While the fleet-of-footedness required to 'get the document out the door' still exists, the inherent capacity of a firm to deliver is met with an even more focused eye on *minimizing the content development risks*.

These challenges underscore the need for answers, among them following the best practices of:

- rock-solid repurposing (e.g., Edit | Paste Special as Unformatted Text) techniques
- ambidextrous comparison methods (e.g., ability to compare disparate source content such as Excel to Word)
- create standard templates in which all content development occurs
- create boilerplate text and table structures
- develop and assess skills, techniques and tools which minimize re-typing data
- improve the collaboration landscape between you and your client, up to and including providing your client with the same boilerplate content and templates, tools and training resources you enjoy in-house

The Objective

The goal of this document is to enlighten you on ways other firms have reduced not only their "time-from-EDGAR", but their "time-to-EDGAR", while simultaneously maximizing predictability and quality.

This document will:

- identify the framework against which to assess your own processes
- offer techniques to improve the manual cleanup pursuit specific to each type of source
- raise awareness of tools to ease the burden of bringing forth a quality, timely result:
 - creative use of basic Word features:
 - find/replace with wildcards and white space
 - AutoFormat
 - AutoText

- feature sets in later versions of Word (2002/2003):
 - “Highlight all items” in Find
 - “Select Text with Similar Formatting”
 - File|Save As Filtered HTML
- DocXtools (occasions flagged with product logo)
- Adobe Acrobat

What steps should we be taking?

The most successful EDGAR-filing workflows build the following into their pursuit:

1. Preparatory cleanup of springboard documents, most particularly:
 - baseline document taken from EDGAR (if applicable)
 - baseline document or documents used from previously-filed deals
2. Careful and expert placement of newly-authored content through team-wide use of:
 - boilerplate templates and content elements (e.g., commonly used table structures, text frames and other form elements)
 - Edit | Paste Special as Unformatted Text when source was drawn from other Word documents, PDF files or HTML, and
 - appropriately-treated table structures, whether brought in from Excel, or created from tabbed or typewriter-form text
3. Continuing assessment and development of:
 - boilerplate document quality and integrity
 - quality and integrity checks throughout the document lifecycle
 - tips, techniques and tools which accelerate the cleanup and table-structure building endeavors
 - operators’ command of available tools, and
 - expert-level document skills, requiring intimacy with varying formats and tools — Word, RTF, HTML, PDF — and resolving the dynamics created by all-too-frequent use of the Windows Clipboard.
4. Simultaneous-focus on:
 - the collaborative needs of the Client
 - requirements for the EDGAR-filing moment

How much time should we allocate to preparatory cleanup?

The time required to cleanse content largely depends upon :

- source format (e.g., PDF, HTML, ASCII, .DOC, Excel — or all of the above...)
- electronic quality of the source documents
- skills of the operator, and

- the availability of boilerplate table structures and/or reusable content and its formatting

The preparatory cycle alone averages about 20 minutes per tabled-page. This means a properly-prepared 80-page 10-K consumes between 12-14 hours *before* content edits even begin³.

The trail back to EDGAR isn't much better. Among the most time-consuming tasks are:

- adding uniformity to content assembled from disparate sources
- general cleanup (e.g., consistent quotation marks, similarly-formed paragraphs, appropriate font references, consecutive numbering, reliable table formatting, etc.)
- preparations — or reparations! — required to create the properly-formed HTML required by the EDGAR system

All this means you can plan on spending between eight to 10 hours 'fiddling' and tweaking an average-sized 10-K. Add to that all the authoring, data entry and proofing required, and the merit of accelerating the repeatable tasks becomes evident.

From EDGAR — Best Practices for Repurposing its Content

Although EDGAR is a destination, it's the only place you'll find the latest version of the document: as last-minute filing changes work their way into the posted content, your internally-managed source may not be up to date.

As such, and dependent upon the format of that source (PDF, HTML or ASCII text), it can be challenging (read: time-consuming) to whip content and formatting into shape.

What follows are some tips and tricks to assist in the endeavor, regardless of the source of your content. And don't miss the final page of this document if you're a Microsystems Knowledge Partnership client: DocXtools SEC Accelerator hastens the task even further!

Manual Cleanup from Modernized EDGAR (HTML)

The EDGAR cycle can begin when the requisite, modernized EDGAR document (in HTML format) is retrieved from <http://www.sec.gov>.⁴ Cleanup steps required are:

³ It is assumed all cleanup and preparatory work is complete *before* content editing occurs. In situations where content edits are made without due diligence paid to cleanup tasks and forging consistent formatting onto the work product, the filing moment becomes jeopardized, not to mention the firm's relationship with the client!

⁴ If you're working from the ASCII or supplemental PDF postings, the manual cleanup steps are similar, but the process is likely to be elongated. Those distinctions can be found in the section titled "Manual Cleanup Process from EDGAR (Supplemental PDFs)" which begins on Page 17.

1. File | New in Word to appropriate SEC template (e.g., 10-Q, 10-K or other), then Insert | File the downloaded HTML document. File | Save into Word format.

NOTE: Insert | File is the preferred method since File | Open of HTML means content will continue to be saved as HTML, not as a Word document.⁵ You'll need to perform a File | Save As | (Word document) to make it a binary .DOC. file.

2. As long as there are no other graphics in the document, you can remove the spurious graphic lines in the table using Edit | Replace. In the “Find What?” box, enter “^g” without the quotes, make sure the “With What?” box is empty, and click “Replace All”.



Figure 1 – Replace All Graphic Lines

If there are other graphic objects you need to preserve, you can use the Browse Object method (see Figure 2 — Browse by Object below), or **Shift-F4** to find next.

3. Table cleanup comes next. Select “Table” as your “Browse” object type using the icon at the lower right corner of the Word window. Navigate to the first table.



Figure 2 — Browse by Object

4. Select the entire table (**Alt-5** on the keypad, NumLock off), then **Ctrl-Shift-N** (applies “Normal” to the selected text) to ensure any obscured text becomes visible, and all applied styles are removed.

⁵ As we tested our process against several EDGAR-posted HTML files, we found at least one file which hung Word 2003 upon an Insert|File. Should you encounter this issue, use File|Open, instead, then copy and paste content into your SEC template.

5. Copy the table, and paste it into a new blank document. File | Save as a separate, temporary document.
6. With the cursor anywhere in the table, clear away residual table formatting with **Ctrl-Alt-U**.



7. To ensure the table wraps properly, select the table (**Alt-5**) and alter the font size to 10 pt.: **Ctrl-Shift-P, 10, Enter**. File | Save, if desired.

Using DocXtools, click the Cleanup drop-down on the Expert toolbar and choose Manual Cleanup, choose the following tools from this menu:

- beneath Character: “Remove extraneous spaces and tabs”
- beneath Tables: “Remove leading & trailing ¶ markers from cells”

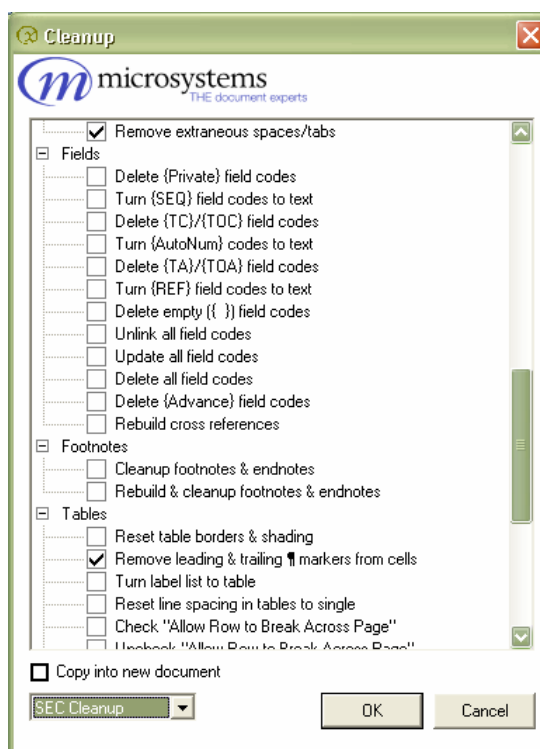


Figure 3 – DocXtools Manual Cleanup

Click OK, then OK again to confirm completion.

NOTE: If you don’t have DocXtools, a *careful* series of replaces will be necessary.

8. Zoom up to 200% or more, then simplify the table structure:
 - use Table | Merge cells to eliminate any superfluous empty cells
 - delete any pointless empty columns
9. Copy the treated table, return to the originating document. Select the ‘source table’, and delete it. Paste the clean table in place.

10. Repeat steps until all tables are done.



11. Using DocXtools “Apply”:

- Outlines
- Titles
- Body Text
- Single Lists
- Bullets

NOTE: If you don’t have DocXtools, follow these guidelines:

- Restyle body text elements in document. Quickly do so using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise a series of ‘replace’ operations will also accomplish this.
- Restyle title text elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise, a series of ‘replace’ operations will also accomplish this.
- Restyle bulleted elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise a series of ‘replace’ operations will also accomplish this.
- Restyle any paragraphs which indicate that outline or single-list numbering should be applied, if necessary (e.g., if a Table of Contents, navigation using Word’s Document Map, or Adobe Reader’s Bookmarks pane when producing a PDF are desired). Additionally, automatic numbering may be required to facilitate any further editing.

12. Revisit each table, ensuring that fonts and font size are consistently applied throughout.

⁶ This feature only enabled when Tools|Options|Edit| “Keep Track of Formatting” and “Mark formatting inconsistencies” are checked. Also, Tools|Options|Save| “Disable features introduced after:” must remain *unchecked*.

- Visit each table, resetting borders using Table | Table Autoformat, then choose “Table Normal” style in Word 2002 or Word 2003.

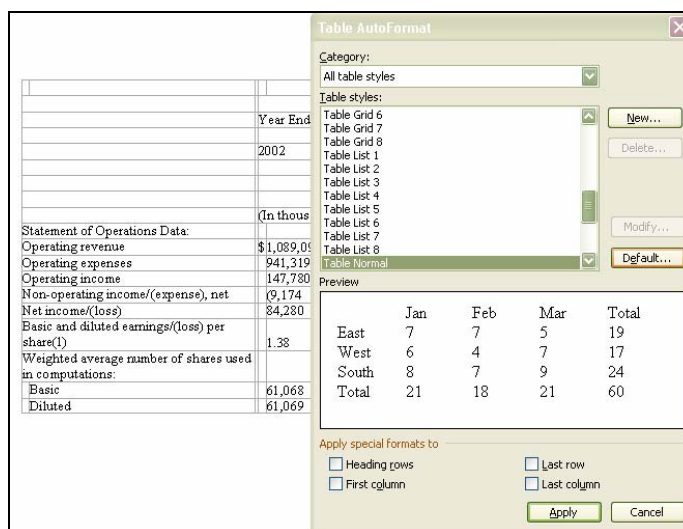


Figure 4 – Table Autoformat

- Visit each table, ensuring tables do not extend into page margins. If so, resize them. (In Word 2003, a Tools | Options | Compatibility setting can help to safeguard against this issue.)
- Revisit each table, ensuring that the properties of merged and split cells are appropriate.
- Revisit each table, ensuring that any graphical elements (e.g., separator lines and/or borders used within a cell) appear as desired.

Normal (Web)	PART I
Normal (Web)	Item 1. Business
Normal (Web)	Our website address is www.expressjet.com. All of our filings with the U.S. Securities and Exchange Commission (“SEC”) are available free of charge through our website on the same day, or as soon as reasonably practicable after we file them with, or furnish them to, the SEC.
Normal (Web)	Forward-Looking Statements
Normal (Web)	This Annual Report on Form 10-K contains forward-looking statements as defined in the Private

Figure 5 – HTML’s Spurious Use of Hard Spaces

- ‘Convert’ hard spaces which lead the paragraph into proper paragraph indents — see Figure 5 above. This requires a two-pronged series of search/replace operations. Assuming a pattern of five hard spaces should become a first line indent of .5 inches:



Figure 6 – Replace #1: Hard Spaces to First Line Indent



Figure 7 – Replace #2: Eliminate the Hard Spaces

18. Clean up random inserted hard spaces throughout document using replace, e.g., those which fall *within* or *internal* to the paragraph. Reinsert required hard spaces throughout the document where line endings and page endings necessitate.



Figure 8 – Eliminate Remaining Hard Spaces

19. Locate and handle spurious manual line breaks (e.g., [Shift-Enter]’s) throughout document: Edit|Replace ^1 with ^p.
20. Optional, but based on your firm standards: replace straight quotes with smart or curly quotes.
21. Locate and replace any inappropriate symbols and (e.g., “- -” becomes “—”).

22. Resolve Page Margins, Headers and Footers and page numbering as these are not features inherent in HTML, thus must be assigned during the cleanup process. You should also watch for the inevitable insertion of:
 - Landscape-oriented tables which will require insertion of an altered section above and below the table.
 - Introduction of unique page numbering, per section.

HINT: It is best to work from the back of the document forward during this process.
23. Once all pagination challenges (e.g., line/page endings and section break insertions) are met, rebuild the Table of Contents, most often inserted as a table, at least in the case of a 10-K. If Word's table of contents feature will not be used, it is recommended that you use Insert | Bookmark, cross-referencing to the appropriate Heading style in the document. This ensures further edits to the document will require no further modification to this TOC.
24. If you have Adobe Acrobat, you can verify your result matches that of your source using the technique described in the "Verifying Original PDF vs. Resulting OCR'ed & Cleaned Copy" segment on Page 21.
25. Once cleanup is complete, the document is ready for the content edits required, namely the insertion of new table columns to reflect the current annual or quarterly numbers, update applicable company information and revise or author new content for the coming year.
26. Ensure that throughout the authoring process, and you attempt to post the document back up to EDGAR, on-going quality checks are made. Because content will be assembled from myriad sources, what had previously been a stable document can easily become unstable, requiring risky interventions at the final hour.

Manual Cleanup Process from EDGAR (ASCII)

Many have toiled over the task of turning plain text EDGAR submissions into editable Word content. While text cleanup is assuredly tedious, the real test is spinning typewriter-esque financial content into properly-formed Word tables. That's where skills, ingenuity and command of a few of Word's new features come into play.

Please note: Modernized HTML EDGAR postings already house financial content into discernable — albeit messy — table structures. Refer to Manual Cleanup *from* Modernized EDGAR (HTML) above for best practice methodologies to cleanup these structures.

1. Download a plain text edition of an EDGAR 10-K or 10-Q submission (<http://www.sec.gov/>).

2. Insert|File into the template which contains appropriate margins and styles used to govern your resulting SEC document. File|Save the document.



3. If you have DocXtools, select the following options on the Manual Cleanup dialogue:
 - o Reflow end of line markers into paragraphs
 - o Convert straight quotes to curly quotes

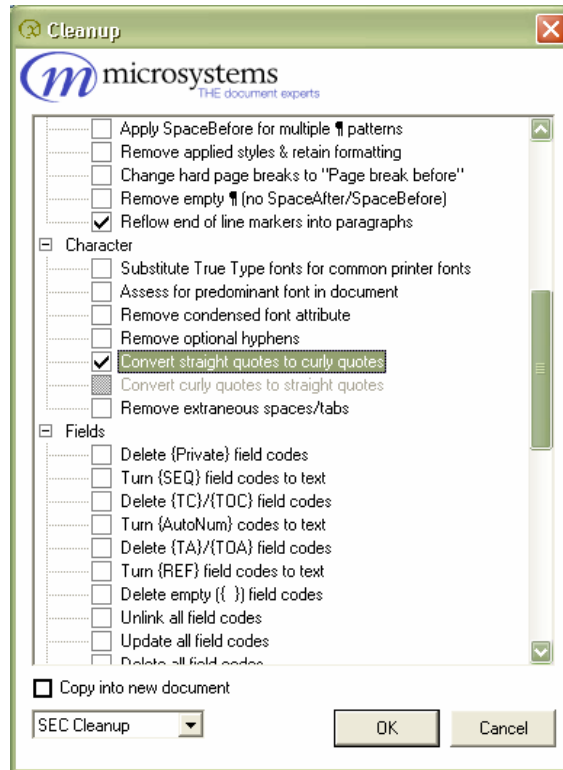


Figure 9 – Manual Cleanup Options for ASCII EDGAR

If you don't have DocXtools, you'll need first to confirm your settings for Word's AutoFormat by visiting Format|AutoFormat|Options ensuring the following options are selected:

- o Straight quotes with "smart quotes"
- o Plain text WordMail documents

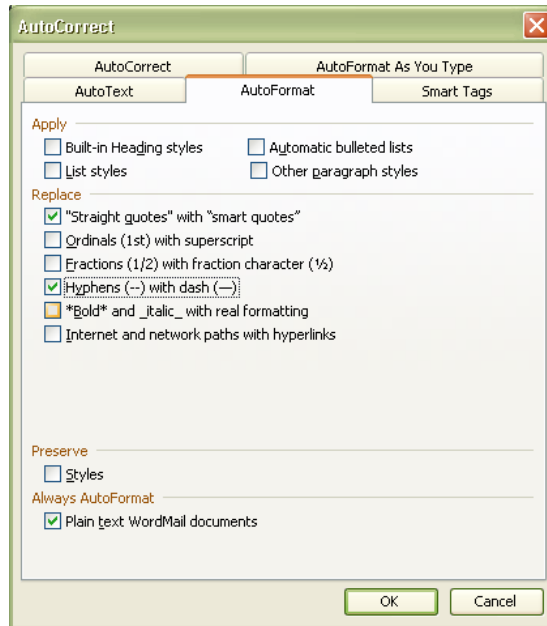


Figure 10 — Confirm Word’s AutoFormat Settings

Dismiss this dialogue, then “AutoFormat” now by clicking on OK. (Note: If your AutoFormat options are already established, you can press **Ctrl-Alt-K** to invoke the AutoFormat process). When done, paragraphs will reflow, and most manually-terminated line endings will be removed.

4. **Cool trick:** “Shake off” all leading spaces turning them into first-line indents with this multi-part tip:
 - Edit|Replace, and in the “Find what:” box enter ^p^w
 - In the “Replace with:” box assign first-line indent formatting (More|Format|Paragraph|First Line|.5). Replace All.



Figure 11 – Replace with First-Line Indents

- Next, select all paragraphs with this applied formatting: Edit|Find, “Find what:” ^p^w, adding the First Line Indent to the further qualify your candidates. Check the “Highlight all items found box”.



Figure 12 – Select All Applied First-Line Indents

- Dismiss the dialogue, but *don't* select anything in the document as this dismisses your document-wide selections. Press **Ctrl-E** to center everything, then left align it with **Ctrl-L**. All leading spaces are now gone.



5. Using DocXtools “Apply” features, apply:

- Outlines
- Titles
- Body Text
- Single Lists
- Bullets

NOTE: If you don't have DocXtools, follow these guidelines:

- Restyle body text elements in document. Quickly do so using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁷. Otherwise a series of ‘replace’ operations will also accomplish this.
- Restyle title text elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise, a series of ‘replace’ operations will also accomplish this.
- Restyle bulleted elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise a series of ‘replace’ operations will also accomplish this.
- Restyle any paragraphs which indicate that outline or single-list numbering should be applied, if necessary (e.g., if a Table of Contents, navigation using Word’s Document Map, or Adobe Reader’s Bookmarks pane when producing a PDF are desired). Additionally, automatic numbering may be required to facilitate any further editing.

⁷ This feature only enabled when Tools|Options|Edit| “Keep Track of Formatting” and “Mark formatting inconsistencies” are checked. Also, Tools|Options|Save| “Disable features introduced after:” must remain *unchecked*.

6. Navigate to the first “Table” tag present in the document — Edit|Find|, and enter <TABLE> in the “Find What:” box. Click OK.
7. Locate the line where column indicators (<C>) demark the tables' structure:

<S>	<C>	<C>	<C>	<C>	<C>	<C>	<C>
258508	MENLO PARK	CA	94025	SFD	5.875	5.483	\$2,721.08
259118	THOUSAND OAKS	CA	91362	SFD	5.375	4.983	\$1,813.01

Figure 13 – Sample ASCII Text Table Excerpt

Copy text from this line to the end of the “table”.

8. Create a new, Landscape document. Paste the content inside. If lines wrap, adjust page margins and/or paper size until all lines maintain their single-line presence.
9. Edit|Select All (**Ctrl-A**). While maintaining the active selection, click in the ruler to insert a left-aligned tab marker above each left angle bracket (<).
10. Collapse only columnar spaces into Tabs using this ingenious Replace string:
 - Edit|Replace, in the “Find What:” box type a space, followed by ^w. (The space is crucial, as it preserves the spaces between words; without it, even words will become separated with tab markers.)
 - In the “Replace with:” box, enter ^t., and Replace All
11. Convert remaining text into a table using Tables|Convert|Text to Table.
12. Save the document, then clean up, touch up and edit the table structure and its properties as desired. When finished, move the table back into the governing document.
13. Repeat steps for each remaining table.

Special note: Editing tables in a separate document greatly accelerates Word's performance during this demanding editing phase: there is less document content to paginate, thus minimizing on-screen refresh events.

14. Remove all leftover tags using this wildcard search string: Edit|Find, enable both the “Use wildcards” and “Highlight all items found in” and enter the following string in the “Find what:” box: \<*\> as illustrated in Figure 14 below.

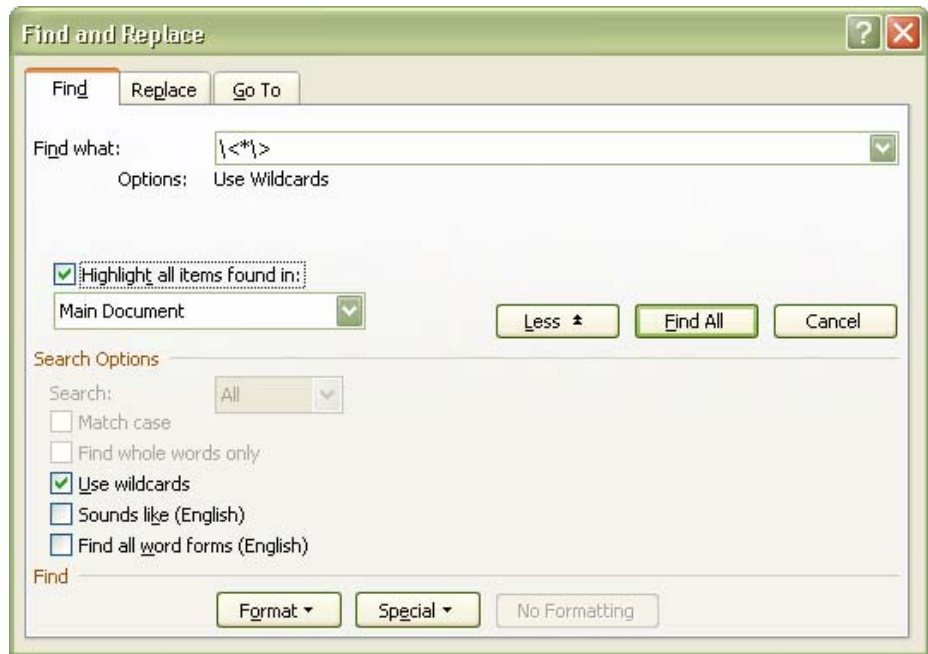


Figure 14 – “Use wildcards” to Eliminate Leftover Tags

15. Touch up the remaining content, using the Document Map to quickly navigate the document.
16. Regenerate the table of contents, and resolve any page numbering or section formatting.
17. Verify your formatting work using DocXtools “Discover” which assesses the document for document quality, and ensures best practices.
18. If you have Adobe Acrobat, you can verify your result matches that of your source using the technique described in the “Verifying Original PDF vs. Resulting OCR’ed & Cleaned Copy” segment on Page 21.
19. Your result is now ready for content edits.

Manual Cleanup Process from EDGAR (Supplemental PDFs)

The cleanup trail from a posted supplemental PDF takes on one of two forks:

- Some form of paper capture or OCR scanning (*most* tedious)
- Methods which leverage a combination of the tips and tricks noted in the “Manual Cleanup from Modernized EDGAR (HTML)” and “Manual Cleanup Process from EDGAR (ASCII)” sections above.

An image-based PDF

You’ll know it’s an image-based PDF when Edit|Select All in Adobe Reader selects nothing, or when you Search in the PDF and strings are not found. When this happens you’ll need to OCR the content using any of the following methods:

- Adobe Acrobat’s free plug-in, Paper Capture
- OmniPage Pro’s ability to OCR a PDF
- If you have the source pages available in TIFF images, you can use Microsoft’s Office Document Imaging Solution, a little-known but valuable freebie available in Office XP and 2003

Once you have the document as editable text, perform cleanup as described in the sections above.

A Text-Based PDF

If your PDF is text-enabled, the repurposing workflow resembles this:

1. Open the PDF file in Adobe Reader
2. Edit | Copy file to Clipboard
3. In Word, File | New against a “Normal” document, and Edit | Paste, preserving formatting



4. If you have DocXtools, select the following options on the Manual Cleanup dialogue:
 - Reflow end of line markers into paragraphs
 - Convert straight quotes to curly quotes

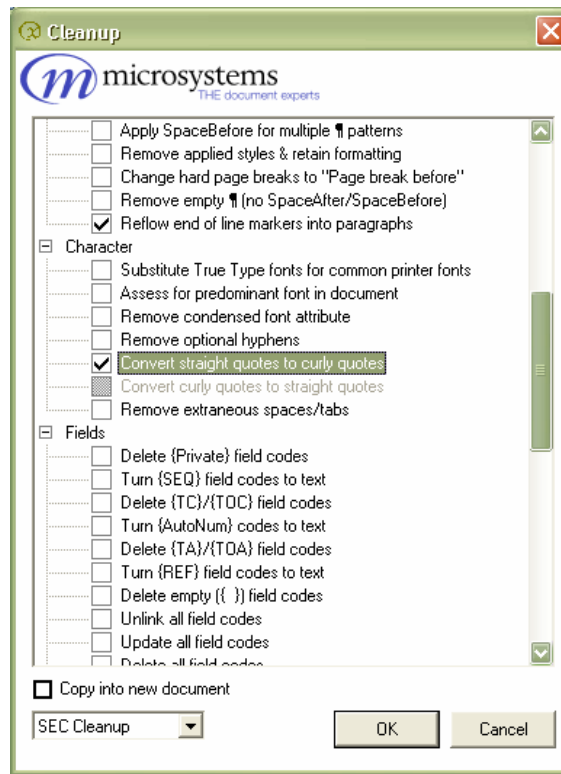


Figure 15 – Manual Cleanup Options for PDF Cleanup

If you don't have DocXtools, you'll need first to confirm your settings for Word's AutoFormat by visiting `Format|AutoFormat|Options` ensuring the following options are selected:

- Straight quotes with “smart quotes”
- Plain text WordMail documents

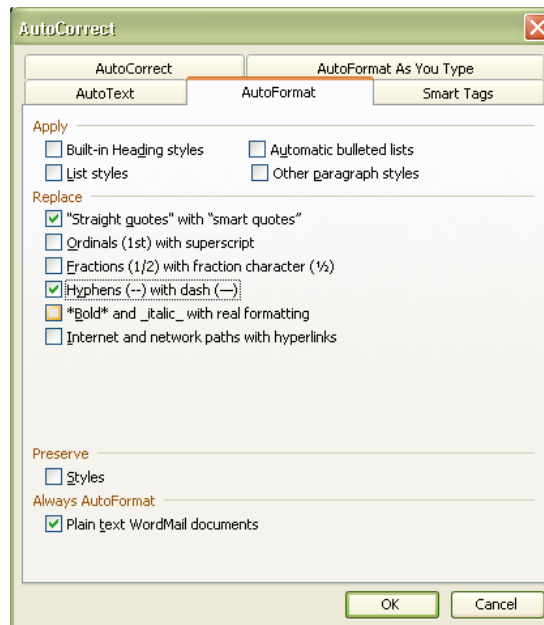


Figure 16 — Confirm Word's AutoFormat Settings

Dismiss this dialogue, then “AutoFormat” now by clicking on OK. (Note: If your AutoFormat options are already established, you can press **Ctrl-Alt-K** to invoke the AutoFormat process). When done, paragraphs will reflow, and most manually-terminated line endings will be removed.

5. Delete any unnecessary textual elements, such as those which reflect page numbers, or in-line header or footer content.
6. Because tables are delivered into the document as typed text, you will need to reconstruct them. The easiest method would be to use Adobe Acrobat’s Select Table tools to select, then copy each segment as a structure, then paste it into the appropriate location in the target document you’re building — a very laborious task. If you don’t have Acrobat — or scanning software which recognizes the tables as tables — plan on spending a significant amount of time on this step.
7. Use DocXtools Expert Toolbar, Cleanup | Tools | Apply Normal font to document content to eliminate inappropriate fonts from the document.

If you don’t have DocXtools, visit Tools | Options | Compatibility | Font Substitution to make sure all fonts are substituted. If not, “Convert Permanently” until all substitutions are made.

8. Next, Ctrl-A to Select All, then Ctrl-C to copy the entire document’s content. File | New to the appropriate SEC template, and paste your modified results inside.



9. Using DocXtools “Apply” features, apply:

- Outlines
- Titles
- Body Text
- Single Lists
- Bullets

NOTE: If you don’t have DocXtools, follow these guidelines:

- Restyle body text elements in document. Quickly do so using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁸. Otherwise a series of ‘replace’ operations will also accomplish this.
- Restyle title text elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise, a series of ‘replace’ operations will also accomplish this.

⁸ This feature only enabled when Tools|Options|Edit| “Keep Track of Formatting” and “Mark formatting inconsistencies” are checked. Also, Tools|Options|Save| “Disable features introduced after:” must remain *unchecked*.

- Restyle bulleted elements in document. Quickly do so by placing cursor in an appropriately-formatted paragraph, using right-click, “Select Text with Similar Formatting” found in Word 2002 or Word 2003, only⁶. Otherwise a series of ‘replace’ operations will also accomplish this.
 - Restyle any paragraphs which indicate that outline or single-list numbering should be applied, if necessary (e.g., if a Table of Contents, navigation using Word’s Document Map, or Adobe Reader’s Bookmarks pane when producing a PDF are desired). Additionally, automatic numbering may be required to facilitate any further editing.
10. Take care of any necessary page setup, header/footer or page numbering issues.
 11. Verify your content rebuild using the PDF compare technique noted below.

Verifying Original PDF vs. Resulting OCR'ed & Cleaned Copy

A little known method exists to aid in verifying your resulting OCR’ed and cleaned content does not deviate from the originally-supplied PDF. This technique requires Adobe Acrobat. Here’s how:

1. Create a PDF result of your OCR’ed and cleaned document
2. Launch Adobe Acrobat 6, and open both the original PDF you worked from, and the one you’ve just created.
3. Choose Document|Compare Documents, and identify the ‘newer’ and ‘older’ documents.
4. Check “Textual Differences” and select either a side-by-side or consolidated report. Click OK.

The resulting PDF will visually display where differences appear. You can now return to Microsoft Word to make any necessary corrections.

To EDGAR — Creating Modernized HTML from Word

Manual Process to Create Modernized EDGAR (HTML) from Word

The preparation challenge when taking documents into modernized EDGAR HTML largely involves ‘unfolding’ formatting disallowed by the on-line EDGAR system, yet required during the:

- editing
- content development
- collaboration
- comparison and

- printing or PDF-ing

stages in Word. While there are *many* individual tasks, general document-wide guidelines are:

- Remove all unused styles (keeps the resulting HTML file very tidy and compact)
- Remove all table borders
- Center all table structures to appear in the center of the viewable HTML page.
- Eliminate all pagination control (e.g., “Keep with Next”, “Keep Lines Together”, “Widow/Orphan Control”)
- Remove content housed in text boxes, inserting it in-line
- Translate paragraph indents (e.g., first line, hanging and left indents) into hard-coded tab markers
- Use empty paragraph markers, in lieu of paragraph spacing
- Convert curly quotes into straight quotes

Manual verification is necessary for:

- Assessment that all special symbols, such as trademark symbols and fractions, comply with EDGAR guidelines.
- Assessment of all inserted graphics — EDGAR does not allow OLE-linked objects or graphics, requiring graphics in .gif or .jpg form.

Creating the HTML file

While native Word provides a File|Save As Web Page, this result pollutes the delivered HTML codes, generating several errors as it is posted onto the EDGAR system. Use of the “filtered” HTML converter, however, produces results more palatable to EDGAR.

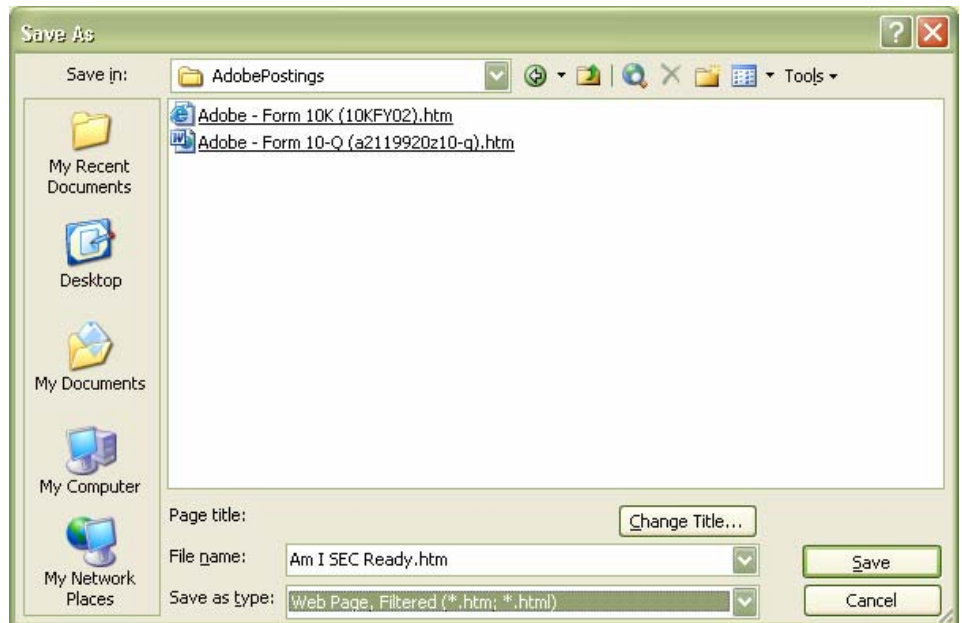


Figure 17 – File|Save As|Web Page, Filtered HTML

The DocXtools SEC Accelerator



The SEC Accelerator, an enriched collection of cleanup tools and Discovery rules to complement those already available in DocXtools, accelerates many of the manual tasks discussed in this document, most notably:

When Repurposing from EDGAR:

- Specialized handling based on source type:
 - ASCII/Plain Text
 - Modernized EDGAR
 - Existing Word work product
- Preparatory cleanup of:
 - Spurious Hard Spaces
 - Paragraph reflow
 - Hard Spaces into Paragraph Indents
 - One-click application of Styles
 - eliminating graphic lines
 - producing tables from text
 - ‘cleansing’ table structures

When Inserting New Content:

- Manipulating columns to remove outdated financial content
- Shifting columnar data to the left or right

When filing to EDGAR:

- Remove all table borders
- Centering table structures
- Eliminating all pagination controls
- Relocating content housed in text boxes to in-line text
- Converting:
 - paragraph indents into hard-coded tab markers
 - paragraph spacing into empty paragraph markers

About Microsystems

Microsystems is the acknowledged leader in software-based document formatting and clean-up solutions for the legal and life sciences markets and other document-intensive industries. Its Knowledge Partnership program provides more than 165 law firms and pharmaceutical companies with an innovative combination of software products, consulting services, and document support services designed to improve the total document structuring and production processes. Microsystems also supplies document conversion and migration services; the company estimates that its products have migrated over 300 million documents from WordPerfect or legacy versions of Microsoft Word to current versions of Word for more than 400 organizations worldwide. Microsystems is headquartered near Chicago in Lombard, Illinois, with satellite offices in California and Washington, DC.