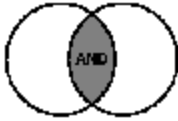


November 21, 2002

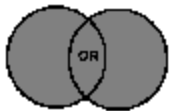
Helpful Hints for Searching Text with ADAMS Find

- ! Every search engine processes search requests differently. ADAMS uses a commercial search engine provided by FileNET.
- ! The retrieval of externally generated documents and older NRC documents, which are stored as images with text generated by using optical character recognition (ocr), is limited by the quality of image scanned.
- ! You **can** now search for **Stop Words**. These are commonly occurring words, such as; the, and, or, of, etc. Many text search engines exclude these words for searching.
- ! **When in doubt, use the wildcard** * after words that might have relevant ending, such as *es, ed, edly, ing*, etc.
- ! The full text search feature in ADAMS handles some variant forms of words for you.
 - , **Hyphenated Words** -- When in doubt, use the hyphen between words that might be hyphenated, for example, ***high-level*** or ***risk-informed***.
 - , **Plurals** -- For words in the text that end in s or 's, a search will retrieve those variants when you enter only the root.
 - , Searches on words or phrases are ***not case sensitive***
- ! There are a small number of files from which searchable text could not be extracted. These are files from non-standard applications or files where the file extension was not recognizable. These files represent only about 0.2 percent of the official agency records collection.
- ! In Advanced Find, **proximity** is measured by a range in the number of **characters, not words**. **Accounting for punctuation** -- In some cases, depending on the source file, the quality of the ocr'ed text, and/or the type of punctuation, there are non-alpha and hidden characters that must be accounted for.
- ! In Advanced Find, if you specify **more than two words** per condition, a **Words In Order** search will retrieve **more** documents than a **Words Any Order** search, even though you think that using **Any Order** would be less restrictive than **In order**. Until you become an experienced user, use only two-word phrases
- ! **Find Count** can not be used when text criteria is part of the search.
- ! The Legacy Library, which will be available soon, has only a limited number of text files.



For higher precision or relevancy, or when retrieving too many documents, to narrow your search:



- ! In Simple Find, use quotes around phrases.
 - , But remember, given variants in punctuation and the limitations of OCR software, if you are not finding what you need using Simple Find, try using the proximity capabilities of Advanced Find.
- , In Advanced Find,
 - , For **two-word** phrases, use **Words In Order** and low range count
 - , For **three or more word phrases**, use **Words Any Order** with range value equal to the number of characters between the first and last word in the phrase, plus 3 more to account for punctuation.



For higher recall, or when not retrieving what expected, to broaden your search:

- ! Use **Wildcards** in Simple Find or Advanced Find.
- ! Try synonyms. Use Advanced Find; add additional rows of word or phrase criteria; and **OR** your text criteria.
- ! **Accounting for punctuation** -- In some cases, depending on the source file, the quality of the ocr'ed text, and/or the type of punctuation, there are non-alpha and hidden characters that must be accounted for. Therefore, if you are not finding what required using Simple Find using quotes, we recommend that you use Advanced Find and set your Range to at least 3 in order to assure higher recall.
- ! In Advanced Find,
 - , For **two-word** phrases, use **Words Any Order** and use a higher range count
 - , For **three or more word phrases**, use **Words In Order** with low range value or **Words Any Order** with a high range value.

Finding Text Within a Document -- You may want to see how the words and phrases are used within the document to ensure that your search is retrieving relevant documents. The technique for searching for text within a document depends on the document's file type.

Document Formats in ADAMS	Steps to search for text within a document
 <p>A <u>PDF</u> file is an electronic document created by Adobe Acrobat.</p>	<p>Double-click the document icon to display the document in the ADAMS Viewer.</p> <p>Use the <i>Find</i> button (binoculars) on the Adobe Acrobat toolbar (the bottom toolbars.)</p> <p>You may leave the ADAMS Viewer window open if you expect to view other documents.</p>
 <p>A <u>TIFF</u> file is a paper document that has been scanned to create an electronic document. A red "T" appears if the document's scanned image has been processed by OCR, Optical Character Recognition, to create a searchable, reusable text document.</p>	<p><u>Right</u>-click the document icon and select <u>Open Text</u> to display the OCR'ed text in the Notepad program.</p> <p>Use the Notepad's Search menu.</p> <p>Close the Notepad when you are finished.</p>