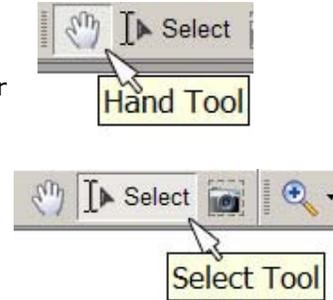


21. QUESTION: How can I extract text and other information from a PDF file?

Much of the content on the MIT OCW Web site is published in Portable Document Format (PDF) format. [Adobe® PDF](#) is the published specification used by standards bodies around the world for more secure, reliable electronic document distribution and exchange. PDF files look exactly like original documents and preserve the fonts, images, graphics, and layout of any source file — regardless of the application and platform used to create it. They can be shared, viewed and printed by anyone, on any system, using [free Adobe Reader® software](#) — regardless of the operating system, original application, or fonts. PDFs have full text search features for locating words, bookmarks, and data fields in documents, and MIT OCW content within a PDF file can be extracted and manipulated for use in other document formats.

There are 3 ways to extract content from an MIT OCW PDF:

- A. To copy text from a PDF document, you must first change the pointer from the **Hand Tool** to the **Select Tool**. In the Acrobat Reader toolbar, the Hand Tool is selected by default. Click the Text Select Tool button on the tool bar. When the Text Select Tool is selected, the Hand-pointer is replaced by a standard text-editing I-beam pointer. To select text in the PDF document, hold down the left mouse button and drag until the desired text is highlighted. Copy the selected text by:
- Clicking the **right** mouse button and selecting Copy
 - By pressing Ctrl+C on the keyboard
 - By clicking the Copy button on the Acrobat toolbar

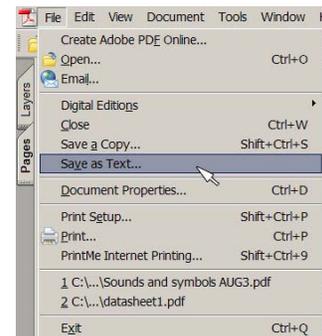


The text of an entire PDF document may be selected by placing the I-beam anywhere within the PDF document and clicking the **right** mouse button and choosing the "Select All" option. It is not necessary to select text with the mouse prior to clicking Select All. When the desired text has been copied, open or return to the word processing application and paste the selected text by:

- Clicking the **right** mouse button and selecting Paste
- Pressing Ctrl+V
- Clicking the Paste toolbar button in the word processing application

Click the Hand Tool button on the Acrobat Reader toolbar to resume scrolling through the PDF document.

- B. In Adobe Reader® 7.0, there is an option in the "File" menu to "Save As Text", which will create a file containing all the text from the PDF, formatted only in rough paragraphs.



- C. On Adobe's ® Web site, you can enter the URL to an MIT OCW PDF [into a form](#) and it will convert the PDF to a rough HTML version of the document.

Adobe PDF Conversion by Simple Form
To convert an Adobe® Portable Document Format (PDF) file to HTML or text, simply type a URL for an Adobe PDF document into this electronic form and select "Convert". The Adobe PDF document will be converted and returned to you in your browser application. Depending on the complexity of the document and the number of requests being made on the server at the time, the conversion process may take a while.

Enter a URL to an Adobe PDF document below:

URL:

Format: HTML 3.2 Text