

Is there a simple way to identify a file as being a PDF?

The first line of every PDF document is a file header with the characters "%PDF-" followed by a version number, shown as "1.n," where "n" would be a digit from 0 to 7. So the first line of a PDF document might look like this:

```
%PDF-1.7
```

See Section 7.5.2, "File Header," in the ISO 32000 Reference, 1.7, page 39. Find this document on the web store of the International Standards Organization (ISO).

Note that the "%PDF-" string is not required to be at the beginning of the file. Acrobat and the Library only require that the header appear somewhere within the first 1024 bytes. If you look at random PDF documents, however, you should find that they're all consistent in the first line. After that, you won't be able to predict the file structure from one file to another. You can open them up for inspection in a text editor such as Microsoft WordPad. The body of the PDF coding may be largely gibberish, especially for compressed documents, but you should find a %PDF declaration within the first 1024 bytes of the file.

See Annex C, "Implementation Limits," in the ISO 32000 Reference, page 649, and Annex I, "PDF Versions and Compatibility," page 727.

If a Version entry in the PDF document's catalog dictionary exists, it is used in place of the version shown in the header. The right version value is found on the catalog.