Portable Document Format (PDF) is a file format used to present documents in a manner independent of application software, hardware, and operating systems.[2] Each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, graphics, and other information needed to display it. In 1991, Adobe Systems co-founder John Warnock outlined a system called "Camelot"[3] that evolved into PDF.

While Adobe Systems made the PDF specification available free of charge in 1993, PDF was a proprietary format, controlled by Adobe, until it was officially released as an open standard on July 1, 2008, and published by theInternational Organization for Standardization as ISO 32000-1:2008,[4][5] at which time control of the specification passed to an ISO Committee of volunteer industry experts. In 2008, Adobe published a Public Patent License to ISO 32000-1 granting royalty-free rights for all patents owned by Adobe that are necessary to make, use, sell and distribute PDF compliant implementations.[6] However, there are still some proprietary technologies with published specification defined only by Adobe, such as Adobe XML Forms Architecture, and JavaScript for Acrobat, which are referenced by ISO 32000-1 as normative and indispensable for the application of ISO 32000-1 specification.[7][8][9][10][11] The ISO committee is actively standardizing many of these as part of ISO 32000-2.

Portable Document Format (pdf) is a file format used to present documents in a manner independent of application software, hardware, and operating systems. Each pdf file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, graphics, and other information needed to display it. In 1991, Adobe Systems co-founder John Warnock outlined a system called "Camelot"[3] that evolved into pdf.

While Adobe Systems made the pdf specification available free of charge in 1993, pdf was a proprietary format, controlled by Adobe, until it was officially released as an open standard on July 1, 2008, and published by theInternational Organization for Standardization as ISO 32000-1:2008, at which time control of the specification passed to an ISO Committee of volunteer industry experts. In 2008, Adobe published a Public Patent License to ISO 32000-1 granting royalty-free rights for all patents owned by Adobe that are necessary to make, use, sell and distribute PDF compliant implementations. However, there are still some proprietary technologies with published specification defined only by Adobe, such as Adobe XML Forms Architecture, and JavaScript for Acrobat, which are referenced by ISO 32000-1 as normative and indispensable for the application of ISO 32000-1 specification. The ISO committee is actively standardizing many of these as part of ISO 32000-2.

File structure

A PDF file is a subset of a COS ("Carousel" Object Structure) format. COS is also used with FDF files.[53] A COS tree file consists primarily of objects, of which there are eight types:[54]

Boolean values, representing true or false

Numbers

Strings

Names

Arrays, ordered collections of objects

Dictionaries, collections of objects indexed by Names

Streams, usually containing large amounts of data

The null object

Objects may be either direct (embedded in another object) or indirect. Indirect objects are numbered with an object number and a generation number. An index table called the xref table gives the byte offset of each indirect object from the start of the file.[55] This design allows for efficient random access to the objects in the file, and also allows for small changes to be made without rewriting the entire file (incremental update). Beginning with PDF version 1.5, indirect objects may also be located in special streams known as object streams. This technique reduces the size of files that have large numbers of small indirect objects and is especially useful for Tagged PDF.

There are two layouts to the PDF files: non-linear (not "optimized") and linear ("optimized"). Non-linear PDF files consume less disk space than their linear counterparts, though they are slower to access because portions of the data required to assemble pages of the document are scattered throughout the PDF file. Linear PDF files (also called "optimized" or "web optimized" PDF files) are constructed in a manner that enables them to be read in a Web browser plugin without waiting for the entire file to download, since they are written to disk in a linear (as in page order) fashion.[26] PDF files may be optimized using Adobe Acrobat software or QPDF.