Chapter 4. QDF Mode

In QDF mode, qpdf creates PDF files in what we call *QDF form*. A PDF file in QDF form, sometimes called a QDF file, is a completely valid PDF file that has `%QDF-1.0` as its third line (after the pdf header and binary characters) and has certain other characteristics. The purpose of QDF form is to make it possible to edit PDF files, with some restrictions, in an ordinary text editor. This can be very useful for experimenting with different PDF constructs or for making one-off edits to PDF files (though there are other reasons why this may not always work).

It is ordinarily very difficult to edit PDF files in a text editor for two reasons: most meaningful data in PDF files is compressed, and PDF files are full of offset and length information that makes it hard to add or remove data. A QDF file is organized in a manner such that, if edits are kept within certain constraints, the **fix-qdf** program, distributed with qpdf, is able to restore edited files to a correct state. The **fix-qdf** program takes no command-line arguments. It reads a possibly edited QDF file from standard input and writes a repaired file to standard output.

The following attributes characterize a QDF file:

- All objects appear in numerical order in the PDF file, including when objects appear in object streams.
- Objects are printed in an easy-to-read format, and all line endings are normalized to UNIX line endings.
- Unless specifically overridden, streams appear uncompressed (when qpdf supports the filters and they are compressed with a non-lossy compression scheme), and most content streams are normalized (line endings are converted to just a UNIX-style linefeeds).
- All streams lengths are represented as indirect objects, and the stream length object is always the next object after the stream. If the stream data does not end with a newline, an extra newline is inserted, and a special comment appears after the stream indicating that this has been done.
- If the PDF file contains object streams, if object stream *n* contains *k* objects, those objects are numbered from *n +1* through *n+k*, and the object number/offset pairs appear on a separate line for each object. Additionally, each object in the object stream is preceded by a comment indicating its object number and index. This makes it very easy to find objects in object streams.
- All beginnings of objects, `stream` tokens, `endstream` tokens, and `endobj` tokens appear on lines by themselves. A blank line follows every `endobj` token.
- If there is a cross-reference stream, it is unfiltered.
- Page dictionaries and page content streams are marked with special comments that make them easy to find.
- Comments precede each object indicating the object number of the corresponding object in the original file.

When editing a QDF file, any edits can be made as long as the above constraints are maintained. This means that you can freely edit a page's content without worrying about messing up the QDF file. It is also possible to add new objects so long as those objects are added after the last object in the file or subsequent objects are renumbered. If a QDF file has object streams in it, you can always add the new objects before the xref stream and then change the number of the xref stream, since nothing generally ever references it by number.

It is not generally practical to remove objects from QDF files without messing up object numbering, but if you remove all references to an object, you can run qpdf on the file (after running **fix-qdf**), and qpdf will omit the now-orphaned object.

When **fix-qdf** is run, it goes through the file and recomputes the following parts of the file:

- the `/N`, `/W`, and `/First` keys of all object stream dictionaries