

pdftotext

Overview

pdftotext is a utility that comes as part of the Foolabs Xpdf package. It is used by the [PDF Solution Pack](#) to extract text from text-based PDFs so that it can be appended to the object as a FULL_TEXT datastream.

Provisions

- [PDF Solution Pack](#)

Downloads

pdftotext is installed as part of Xpdf, which can be found at Foolabs' official site, <http://www.foolabs.com/xpdf/download.html>. For Windows and Mac installations, a binary installer exists there; for Linux installations, however, you may compile it from source, use the binaries from the site, or much more simply use your distribution's package manager to install it automatically; on Debian- and Ubuntu-based systems, this can be accomplished by running:

```
sudo apt-get install xpdf-utils  
or  
sudo apt-get install poppler-utils
```

Usage

More information on how to integrate pdftotext with Islandora can be found on the [PDF Solution Pack](#) page.