

NAME

`djvutxt` - Extract the hidden text from DjVu documents.

SYNOPSIS

djvutxt [*options*] *inputdjvufile* [*outputtxtfile*]

DESCRIPTION

Program **djvutxt** decodes the hidden text layer of a DjVu document *inputdjvufile* and prints it into file *outputtxtfile* or on the standard output. The hidden text layer is usually generated with the help of an optical character recognition software.

Without options **-detail** and **-escape**, this program simply outputs the UTF-8 text. Option **-detail** cause the output of S-expressions describing the text and its location. Option **-escape** uses C-style escape sequences to represent nonprintable non-ASCII characters.

OPTIONS**--page=pagespec**

Specify which pages should be processed. When this option is not specified, the text of all pages of the documents is concatenated into the output file. The page specification *pagespec* contains one or more comma-separated page ranges. A page range is either a page number, or two page numbers separated by a dash. For instance, specification **1-10** outputs pages 1 to 10, and specification **1,3,99999-4** outputs pages 1 and 3, followed by all the document pages in reverse order up to page 4.

--detail=keyword

This options causes **djvutxt** to output S-expressions specifying the position of the text in the page. See the manual page **djvused**(1) for a description of the output format. Argument *keyword* specifies the maximum level of detail for which text location is reported. The recognized values are: **page**, **column**, **region**, **para**, **line**, **word**, and **char**. All other values are interpreted as **char**.

--escape

Output escape sequences of the form **"ooo"** for all non ASCII or non printable UTF-8 characters and for the backslash character.

REMARKS

Use program **djvused**(1) for more control over the text layer.

CREDITS

This program was initially written by Andrei Erofeev <andrew_erofeev@yahoo.com> and was then improved Bill Riemers <docbill@sourceforge.net> and many others. It was then rewritten to use the ddjvuapi by Leon Bottou <leonb@sourceforge.net>.

SEE ALSO

djvu(1), **djvused**(1)