

Resource Summary Report

Generated by [SPARC SAWG](#) on Aug 7, 2026

[lapdftext](#)

RRID:SCR_006167

Type: Tool

Proper Citation

lapdftext (RRID:SCR_006167)

Resource Information

URL: <http://code.google.com/p/lapdftext/>

Proper Citation: lapdftext (RRID:SCR_006167)

Description: Software that facilitates accurate extraction of text from PDF files of research articles for use in text mining applications. It is intended for both scientists and natural language processing (NLP) engineers interested in getting access to text within specific sections of research articles. The system extracts text blocks from PDF-formatted full-text research articles and classifies them into logical units based on rules that characterize specific sections. The LA-PDFText system focuses only on the textual content of the research articles. The current version of LA-PDFText is a baseline system that extracts text using a three-stage process: * identification of blocks of contiguous text * classification of these blocks into rhetorical categories * extraction of the text from blocks grouped section-wise.

Abbreviations: lapdftext, LA-PDFText,

Synonyms: Layout-Aware PDF Text Extraction, Layout-Aware Text Extraction from Full-text PDF of Scientific Articles, lapdftext: Layout-Aware Text Extraction from Full-text PDF of Scientific Articles

Resource Type: software application, text extraction software, software resource

Defining Citation: [PMID:22640904](#)

Keywords: text mining, pdf, text extraction, natural language processing

Funding: NSF 0849977;
NIGMS RO1-GM083871;
NIMH 1R01MH079068-01A2;
NCRR U24 RR025736-01

Availability: Acknowledgement requested, GNU General Public License, v3

Resource Name: lapdftext

Resource ID: SCR_006167

Alternate IDs: nlx_151668

Record Creation Time: 20220129T080234+0000

Record Last Update: 20260806T054427+0000

Ratings and Alerts

No rating or validation information has been found for lapdftext.

No alerts have been found for lapdftext.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We have not found any literature mentions for this resource.