Resource Summary Report

Generated by FDI Lab - SciCrunch.org on Apr 17, 2025

PDFX

RRID:SCR_006163
Type: Tool

Proper Citation

PDFX (RRID:SCR_006163)

Resource Information

URL: http://pdfx.cs.man.ac.uk

Proper Citation: PDFX (RRID:SCR_006163)

Description: A fully-automated PDF-to-XML converter service for scientific articles. It takes a full-text PDF article as input and outputs the hierarchy of its distinct logical elements in an XML format. The elements that PDFX can currently extract are: * Front Matter ** title, abstract, author, author footnote * Body Matter ** body text, h1, h2, h3, image, table, figure/table caption, figure/table reference, bibliographic item, bibliographic reference (citation) * Extras ** header, footer, side note, page number, email, URI Note: This system has been designed for processing scientific articles. While virtually any PDF file is acceptable input, quality of the processing output might be degraded e.g. for entire books, slide presentations or spreadsheet/strictly tabular data. There are two ways in which you can use PDFX: * via a web browser * via any other HTTP client, such as the curl command-line tool

Abbreviations: PDFX

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

Keywords: semantic mark up, text extraction, pdf, xml, html

Funding:

Availability: Free

Resource Name: PDFX

Resource ID: SCR_006163

Alternate IDs: nlx_151665

Record Creation Time: 20220129T080234+0000

Record Last Update: 20250417T065239+0000

Ratings and Alerts

No rating or validation information has been found for PDFX.

No alerts have been found for PDFX.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 2 mentions in open access literature.

Listed below are recent publications. The full list is available at FDI Lab - SciCrunch.org.

Jain S, et al. (2016) Weakly supervised learning of biomedical information extraction from curated data. BMC bioinformatics, 17 Suppl 1(Suppl 1), 1.

Mizuno S, et al. (2016) The Pre-Eclampsia Ontology: A Disease Ontology Representing the Domain Knowledge Specific to Pre-Eclampsia. PloS one, 11(10), e0162828.