

# Resource Summary Report

Generated by RRID on Aug 11, 2026

## CERMINE

RRID:SCR\_008993

Type: Tool

---

### Proper Citation

CERMINE (RRID:SCR\_008993)

---

### Resource Information

**URL:** <http://services.ceon.pl/cermine/index.html>

**Proper Citation:** CERMINE (RRID:SCR\_008993)

**Description:** Java library and a web service for extracting metadata and content from scientific articles in born-digital form. The system analyses the entire content of a PDF file containing a publication and attempts to extract information such as: the title of the article, journal information (title, etc.), bibliographic information (volume, issue, page numbers, etc.), authors and affiliations, keywords, abstract, bibliographic references and structured sections hierarchy.

**Abbreviations:** CERMINE

**Synonyms:** CERMINE - Content ExtRactor and MINEr, Content ExtRactor and MINEr

**Resource Type:** analysis service resource, production service resource, software application, software toolkit, software library, service resource, text extraction software, software resource, data analysis service

**Keywords:** java library, java, metadata extraction, page segmentation, content classification, bibliographic reference parsing, parse, extraction system, pdf

**Funding:** National Centre for Research and Development Poland SP//I/1/77065/10

**Availability:** GNU Affero General Public License, v3

**Resource Name:** CERMINE

**Resource ID:** SCR\_008993

**Alternate IDs:** nlx\_152512

**Record Creation Time:** 20220129T080250+0000

**Record Last Update:** 20260811T043336+0000

---

## Ratings and Alerts

No rating or validation information has been found for CERMINE.

No alerts have been found for CERMINE.

---

## Data and Source Information

**Source:** [SciCrunch Registry](#)

---

## Usage and Citation Metrics

We found 5 mentions in open access literature.

**Listed below are recent publications.** The full list is available at [RRID](#) .

Tong Q, et al. (2026) Decoding enzymatic landscapes: a knowledge graph-enhanced large language model framework for microbial enzyme production and catalysis systems. aBIOTECH, 7(3), 100059.

Guo Z, et al. (2025) Sm-Nd Isotope Data Compilation from Geoscientific Literature Using an Automated Tabular Extraction Method. Scientific data, 12(1), 203.

Acheson E, et al. (2021) Extracting and modeling geographic information from scientific articles. PloS one, 16(1), e0244918.

Carrio-Cordo P, et al. (2020) Geographic assessment of cancer genome profiling studies. Database : the journal of biological databases and curation, 2020.

Leal MC, et al. (2020) Fifty years of capacity building in the search for new marine natural products. Proceedings of the National Academy of Sciences of the United States of America, 117(39), 24165.