

NOTEWORTHY, SOFTWARE

# ACADEMIC TEXT PARSING

19.09.2024

I used to [parse PDFs](#) using the [Allenai](#) method and the [layoutparser](#).

This worked in many instances but is no longer maintained.

I still have [Nougat](#) on my to do list while a [new paper](#) now points to [AceParse](#)

AceParse includes various types of structured text, such as formulas, tables, algorithms, lists, and sentences embedded with mathematical expressions, among others. We provide examples of several dataset samples to give you a better understanding of our dataset.

CC-BY-NC Science Surf accessed 11.03.2026

---