

Annotate Ease: PDF Metadata Extraction Application Specializing in Research Publications

Nivetha K. K.^{1*}, Samanthi Priyasadini²

¹Kingston University, UK,²ESOFT Metro Campus, Colombo 4, Sri Lanka
*nive2012@gmail.com

Annotate Ease is an innovative software application designed to address the challenges of extracting metadata from research publications, predominantly in Portable Document Format (PDF). Utilizing advanced artificial intelligence (AI) models, Annotate Ease automates the extraction of critical information such as titles, authors, abstracts, affiliations, and bibliographic data, organizing it into structured JavaScript Object Notation (JSON) files for seamless integration with other research tools and databases. By incorporating Open Researcher and Contributor Identifiers (ORCID), the tool enhances the accuracy of author identification and affiliation validation, achieving a tested precision rate of over 95%. Annotate Ease specializes in processing structured PDFs, providing reliable outputs with consistent formatting and significantly reducing the effort required for manual extraction. Although bulk processing is not yet supported, the tool can process up to 10 structured PDFs in approximately 20 minutes, making it well-suited for smaller-scale academic workflows. Future versions aim to enhance throughput, enabling the handling of larger document batches. However, Annotate Ease faces challenges with corrupted or poorly formatted PDFs, which may require preprocessing to ensure optimal performance. Addressing these issues is a priority for ongoing development, along with cost optimization strategies to reduce API-related expenses. Comparative analysis demonstrates that Annotate Ease excels in accuracy, user-friendliness, and efficiency when compared to other metadata extraction tools. Many competing solutions struggle with academic workflows, lack intuitive interfaces, or deliver inconsistent outputs. Annotate Ease's structured JSON outputs seamlessly integrate with downstream tools such as citation managers and analytics platforms, positioning it as a transformative solution for academic research. Beyond its technical advantages, Annotate Ease saves researchers time by automating repetitive tasks and enabling them to focus on critical analysis and discovery. Future enhancements will aim to expand the tool's applicability to a broader range of academic documents, such as white papers, theses, and conference proceedings while introducing preprocessing steps to handle variability in document formatting. These improvements will ensure Annotate Ease remains an indispensable resource for researchers, publishers, and database administrators, driving efficiency and innovation in academic workflows.

Keywords: *Artificial Intelligence, Academic Tools, Metadata Extraction, ORCID Validation, PDF Processing, Research Publications.*