

# Show merge dialog when importing a single PDF

## Context and Problem Statement

PDF files are one of the main format for transferring various documents, especially scientific papers. However, by itself, PDF is like a picture, it contains commands solely for displaying the human-readable text, but it might not contain computer-readable metadata.

To overcome these problems various heuristics and AI models are used to “convert” a PDF into a BibTeX entry. However, it also introduces a level of problems, as heuristics are not ideal: sometimes it works perfectly, but on others it generates random output.

PDF importing in JabRef is done via `PdfImporter` abstract class and its descendants, and via `PdfMergeMetadataImporter`. `PdfImporter` is typically a single heuristics or method of extracting a `BibEntry` from PDF. `PdfMergeMetadataImporter` collects `BibEntry` candidates from all `PdfImporter`s and merges them automatically into a single `BibEntry`.

The specific problem JabRef has: should JabRef automate all heuristics (automatically merge all `BibEntry`ies from several `PdfImporter`s) when importing PDF files or should every file be analysed thoroughly by users?

## Decision Drivers

- Option should provide a good-enough quality.
- It is desired to have a fine-grained controls of PDF importing for power-users.

## Considered Options

- Automatically merge all `BibEntry` candidates from `PdfImporters`.
- Open a merge dialog with all candidates.
- Open a merge dialog with all candidates if a single PDF is imported.

## Decision Outcome

Chosen option: “Open a merge dialog with all candidates if a single PDF is imported”, because comes out best (see below).

# Pros and Cons of the Options

Automatically merge all `BibEntry` candidates from `PdfImporters`

- Good, because minimal user interaction and disruption of flow. It also allows batch-processing.
- Bad, because heuristics are not ideal, and it is even harder to develop a “smarter” merging algorithm.

Open a merge dialog with all candidates

- Good, because allows for fine-grained import. Some correct field may be overridden by a wrong field from other importer, which is undesirable for power-users.
- Bad, because it is a dialog. If lots of PDFs are imported, then there will be lots of dialogs, which might be too daunting to process manually.

Open a merge dialog with all candidates if a single PDF is imported

Explanation:

- If a single PDF is imported, then open a merge dialog.
- If several PDFs are imported, merge candidates for each PDF automatically.

Outcomes:

- Good, because it combines the best of the other two options: Allow both for PDF batch-processing and for fine-grained control.
-