

MORPEP META CMP

Full-Text-Download-Anleitung

Franz Prante

ToDo:

Ordner in Nextcloud: HIWI_MP-> Full text download-> packages_for_full_text_download.
Darin befinden sich 26 Excel-Dateien (`study_set_XX.xlsx`) mit jeweils ≤ 100 Literatur-Einträgen.

- Matthias bearbeitet `study_set_1` bis `study_set_2`
- Viktoria bearbeitet `study_set_3` bis `study_set_10`
- danach sehen wir weiter

Ziel: Volltexte als PDF-Datei downloaden

- PDFs alle in den ordner HIWI_MP -> Full text download -> PDFs speichern. Keine Unterordner.
- PDFs sollen nach ihrem **BibtexKey** benannt sein, welcher in der jeweiligen Excel-Datei `study_set_XX.xlsx` zu finden ist.
 - Beispiel:
 - * **BibtexKey** = Nguyen_2022_the
 - * Dateiname des PDF's: Nguyen_2022_the.pdf

Umgang mit BibtexKey-Duplikaten:

Wenn beim Speichern eines PDFs festgestellt wird, dass ein Dokument mit diesem Dateinamen bereits im Ordner PDFs existiert, dann die beiden PDFs vergleichen. Entweder es handelt sich um ein echtes Duplikat (gleicher Titel, Autor, Jahr) oder der **BibtexKey** war nur durch Zufall identisch.

- Wenn echtes Duplikat: 1 in die Spalte **duplicate**. Bitte prüfen, welche Version des PDFs aktueller ist und ggf. die ältere Version durch die neuere ersetzen.
- Wenn Zufall: den **BibtexKey** des neuen PDFs in der Datei `study_set_XX.xlsx` ändern (Unterstrich (`_`) + nächstes Wort des Titels hinten anfügen) und das neue PDF unter diesem Namen abspeichern. Das bereits existierende PDF behält den alten **BibtexKey**.

Allgemeines Vorgehen:

Ganz wichtig: keine Änderungen an den bereits ausgefüllten Zellen der Excel-Tabelle vornehmen. Ausnahme: Zufälliges Bibtex-Key-Duplikat, siehe oben.

- Excel-Datei (`study_set_XX.xlsx`) öffnen.
- `url` aufrufen oder `doi`¹ aufrufen. Falls diese ins Leere führen oder die aufgerufene Studie nicht mit den Angaben in `study_set_XX.xlsx` übereinstimmen, dann den Titel aus `study_set_XX.xlsx` (Spalte `title`) über GoogleScholar suchen.
- Abgleichen des PDFs mit den Daten aus der Excel-Tabelle `study_set_XX.xlsx`. Titel und Autor(en) sollten gleich sein. Wenn andere Variablen abweichen, könnte es sich um eine neuere/ältere Version handeln. Dann prüfen und ggf. vermerken (`most_recent_version_available` 1 = ja, 0 = nein).
- PDF als Volltext downloaden und im Ordner HIWI_MP -> Full text download -> PDFs wie oben benannt abspeichern.

Für jedes PDF die folgenden 5 Spalten in `study_set_XX.xlsx` ausfüllen:

- `most_recent_version_available`: 0 oder 1 vergeben [1 = ja, 0 = nein]
 - kontrollieren, ob das PDF die neueste Version des Artikels ist (wenn es ein publizierter JournalArticle ist, dann ist das klar; hier ggf. die Daten zu PublicationTitle, Issue, Volume abgleichen /ergänzen/nachtragen)
 - bei Working-Papern nach der neuesten Version suchen (Google Scholar -> “alle ... Versionen”)
- `not_available`: 0 oder 1 vergeben [1 = ja, 0 = nein]
 - Für den Fall, dass der Volltext nicht als PDF verfügbar ist, bzw. darauf von der TU Chemnitz nicht zugegriffen werden kann. (WICHTIG: Buchkapitel aus Sammelbänden nehmen wir auf, wenn verfügbar! Auch Dissertationen, die als “Buch” publiziert sind. Dies wurde bei der anderen Metastudie eventuell anders gehandhabt.)
- `retracted`: 0 oder 1 vergeben [1 = ja, 0 = nein]
 - nach Hinweisen suchen, ob ein Papier zurückgezogen wurde
 - das kann bspw. so aussehen: <https://academic.oup.com/restud/article/81/2/787/1520611?login=true>
- `notes_PDF_download`:
 - Anmerkungen, Hinweise, Notizen jeder Art können in diese Spalte (Englisch).

¹DOIs (“digital object identifiers”) sind ähnlich wie URLs, können in der Regel also auch einfach in die Adresszeile des Browsers kopiert und geöffnet werden.