

OpenRefine als Hilfsmittel für die systematische Recherche in Leitliniendatenbanken und Studienregistern

Helge Knüttel
Fachreferent Medizin
UNIVERSITÄTSBIBLIOTHEK



Universität Regensburg

Problem

Methodische Anforderungen bei Systematischen Recherchen:

- Suche sensitiv => zahlreiche Suchterme
- Leitliniensuche: breite Themen => zahlreiche Suchterme
- Dokumentation der Suchstrategien
- Weiterverarbeitung der Treffer in anderer Software

Leitliniendatenbanken und Studienregister:

- Oft Ressourcenlimitierung beim Anbieter
- Oft **Einschränkungen bei Suchfunktionen**, u.a.
 - Suchsyntax limitiert
 - Länge des Suchstrings limitiert
 - Keine Suchhistorie, keine Wiederverwendung von Suchen
 - => viele Einzelsuchen notwendig
 - Eingeschränkte oder keine Treffer-Exporte

Comparison

Herkömmliche, **manuelle Praxis:**

- Viele Einzelsuchen (Suchschritte) durchführen
- Dabei viel händische Arbeit:
 - Eingabe der Suchstrings
 - Dokumentation der Trefferzahlen
 - Dokumentation/Speichern der Treffer
 - Deduplizierung der Treffer aus einer Datenquelle
- Erfahrungsgemäß fehlerträchtig!
- Unbefriedigendes Erlebnis

Intervention

OpenRefine als Werkzeug für die Teilautomatisierung von Suchen:

- Zahlreiche Einzelsuchen durchführen, d.h. Webseite aufrufen & herunterladen
- Ergebnisse extrahieren

Software [OpenRefine](#):

- Open Source
- Läuft ohne Installation
- Bedienung im Webbrowser
- Vielfältige Funktionen (Auswahl):
 - Arbeiten mit Textdaten, v.a. Tabellen, HTML, XML, JSON
 - Undo/Redo ☺ Automatisierung mit „Makros“
 - Abfrage von Webseiten/APIs
- OpenRefine lernen: U.a. [Lektion bei Library Carpentry](#)

Outcome 1

- **Teilautomatisierte Suchen** bei
 - [AWMF-Leitlinien-Register](#) (API/JSON)
 - [Guidelines International Network library](#) (HTML/web scraping)
 - [ISRCTN-Studienregister](#) (API/XML)
- Kein Erfolg beim
 - [Deutschen Register Klinischer Studien \(DRKS\)](#) 
- Vorbereitete Suchstrategie wird abgearbeitet:
Suchterme werden einzeln gesucht
- Ergebnis gespeichert als HTML/JSON/XML
- Extraktion von Trefferzahlen ☾ **Suchhistorie** exportieren
- Extraktion von Treffern
- **Deduplizierung** von Treffern aus den Einzelsuchen
- Export der **Treffer im RIS-Format**
- **Reproduzierbar dokumentiert**

Outcome 2

- Nachnutzung möglich
Cave: Anpassung nötig, wenn sich Datenquellen verändern
- Doku & OpenRefine-Makros auf Github:

<https://github.com/knh11545/openrefine4searching>



- Feedback und Erweiterungen erwünscht!