

# LexHelper: die Digitalisierung der 6 Millionen Belege des Bayerischen Wörterbuchs

Manuel Raaf

Referat für IT & Digital Humanities, Bayerische Akademie der Wissenschaften

## Handhabung hunderttausender digitalisierter Fragebögen und Textzeilen

**LexHelper** bietet einige Grundfunktionen, um digitalisierte Fragebögen oder abgetippte Belege zu durchsuchen und zu editieren:

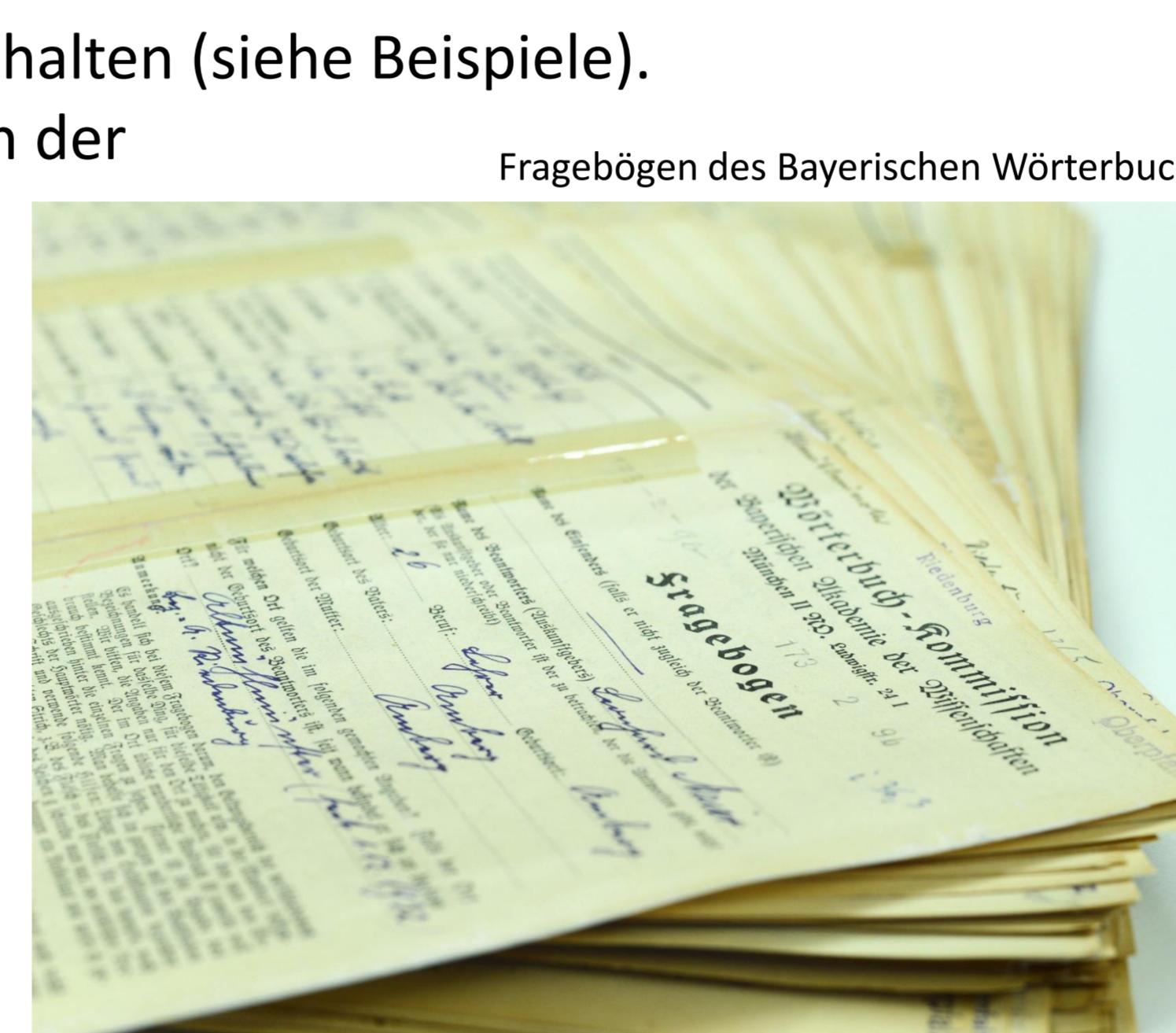
- Import von Scans; Import von Excel-Listen, die Transkriptionen und weitere text-basierte Informationen zu jedem Einzelbeleg enthalten
- halbautomatische Generierung der Schnipsel sowie Schwärzung personenbezogener Daten im Scan
- schneller Zugriff auf alle Daten; schnelles Durchblättern der Belege
- eingeschränkter Lese-Zugriff für die Allgemeinheit
- Vernetzung einzelner Projekte innerhalb **LexHelpers** und somit übergreifende Suche in diesen Projekten

Neben diesen allgemeinen Funktionen sind weitere **projekt-spezifische Funktionen** enthalten (siehe Beispiele).

Außerdem ist die Möglichkeit gegeben, neue zu entwickeln, um künftige Anforderungen der Redaktion erfüllen zu können.

### Technische Details:

- Web-Anwendung (Client-Server)
- OpenSource-Standards in der Entwicklung:
  - Server: LAMP
  - Client: CSS, JavaScript, HTML5
- kostengünstig, da keine Lizenzgebühren anfallen
- anpassbar, wartbar, archivierbar



## Beispiel: Bayerisches Wörterbuch

The screenshot shows the LexHelper interface for the Bavarian Dictionary. It includes a search bar, filters for location (Niederbayern), and a list of results. One result is highlighted with handwritten notes: "Nur nur als 'Tiefat' (Diafat, Dufat)? Was verstehen Sie darunter? Genaue Aussprache? Bitte Satzbeispiel. *Dein Mō schteckt ztuifat'st em Dreck drin.*" Another result shows: "60/ 4. Kennen Sie das Wort Tiefat (Diafat, Dufat)? Was verstehen Sie darunter? Genaue Aussprache? Bitte Satzbeispiel. *Gib acht, daß die nicht ins Tiefat einkommen (in das tiefe Wasser).*" A third result shows: "60/ 4. Kennen Sie das Wort Tiefat (Diafat, Dufat)? Was verstehen Sie darunter? Genaue Aussprache? Bitte Satzbeispiel. *Asier fän Tiefat „Du bist a ganzar Doizte“*"

- > 400.000 digitalisierte, handschriftlich ausgefüllte Fragebögen
  - > 6.000.000 halbautomatisch erstellte Schnipsel
  - Filtermöglichkeiten für Belege:
    - fertig bearbeitete Antworten
    - auskommentierte, unbrauchbare Antworten (z. B. leere oder solche, die lediglich „unbekannt“ oder „nein“ etc. beinhalten)
  - Suchoptionen zur geographischen Filterung (z. B. Regierungsbezirk, Stadt)
  - einfache Zuweisung eines Lemmas zum Beleg
  - PDF-Export der Suchergebnisse oder ganzer Lemmalisten samt Schnipsel
  - Erfassung neuer Fragebögen in nativ digitaler Form
- ✓ schnelles Auffinden von Belegen  
✓ Vorbereitung der Artikelerstellung durch schnelle Lemmatisierung der Belege  
✓ direkte Veröffentlichung der bearbeiteten Belegmaterialien online und gedruckt (via XML-Export und Weiterbearbeitung in Oxygen).

⇒ <http://www.bwb.badw.de/bwb-digital/datenbank/>

## Beispiel: Fränkisches Wörterbuch

The screenshot shows the LexHelper interface for the Franconian Dictionary. It includes a search bar, filters for location (Erste), and a table of results. One result is highlighted: "Erste Getredene Getredene SF NonSg die Getredene Getredene Gauerschein a2b,8". The table has columns for Lemma, Grammatik, Bedeutung, Ortskennung, Umschrift, Ort, Flächenbezeichn., Kommentar GF, Kommentar B, GF, Bogen, Frage, Akzessor.

- basierend auf Excel-Listen
  - Zuweisung jeder Zeile zu Bogen und Frage der entsprechenden Gewährsperson
  - Such- und Anzeigefilter
  - Export nach **REDE**<sup>1</sup> zur halbautomatischen Kartierung (künftig: vollautomatisch)
  - automatische Kompilierung und Onlinestellung des zitierfähigen Lemma-Artikels
  - Anbindung der Belegmaterialien an das Wörterbuchnetz (künftig)
  - CSV-Export der Suchergebnisse für individuelle Weiterbearbeitung / Forschung
  - Sachgruppenzuordnung (nach Post) und Suche über diese
- ✓ schnelles Auffinden von Belegen, die in einem Ort bereits erfasst wurden und somit nicht erneut erfasst werden müssen (→ Zeitersparnis)  
✓ Zugriff auf komplettes Belegmaterial; Suchfilter für entsprechend exakte Suchanfragen  
✓ automatische Artikelerstellung

⇒ <http://www.wbf.badw.de/wbf-digital/wörterbuch/>

<sup>1</sup> SprachGIS (<https://www.regionalsprache.de/SprachGIS/Map.aspx>) zur individuellen Kartenerstellung