



Search + K

Quantitative Analyse der Medienwellen der Spanischen Grippe (1918/19). Eine Fallstudie

Präambel

Groblerziele dieser Fallstudie

Technische Voraussetzungen

Fragestellung und Operationalisierung. Einführung in die Fallstudie

Einführung

Fragestellung

Operationalisierung

Resümee

Korpusaufbau. Auswählen, sammeln, dokumentieren

Einführung

Korpora als Forschungsobjekte der Digital Humanities

Elemente von Korpora: Texte als digitale Objekte

Metadaten

Aufbau des Forschungskorpus

Resümee

OCR. Von Bild zu Text

Einführung

OCR als Methode, um Text maschinenlesbar zu machen

OCR in Python mit PyTesseract

Messung der OCR-Qualität

Messung der OCR-Qualität in Python

Resümee

OCR-Nachbereitung. Manuell, automatisch, LLMs

Einführung

Einführung in die OCR-Nachbearbeitung

Regelbasierte OCR-Nachbearbeitung

LLM-basierte OCR-Nachbearbeitung

Resümee

Korpusverarbeitung. Von Strings zu Token

Einführung

NLP als Methode zur "Semantisierung" von Text

Korpusverarbeitung – Annotation mit spaCy

Resümee

Korpusanalyse. Von Häufigkeiten zu Diagrammen

Einführung

Häufigkeitenanalyse, Visualierung und KWIC

Analyse 1: Diachrone Frequenzdiagramme

Analyse 2: Keyword in Context (KWIC)

Resümee

Reflexion und Resümee

Reflexion

Ausblick: Die Auswirkung der LLM-Nachkorrektur auf die Suche nach Textspuren der Spanischen Grippe

Ausblick

Resümee

Poster: Henny Sluyter-Gäthje, Daniil Skorinkin, Peer Trilcke, Evgenia Samoilova, Melanie Seltmann, Maria Chlastak



Contents

Fallstudien

Jupyter Book als OER
Struktur
Infrastruktur
Elemente

Scalable OER

QUADRIGA-Fallstudien zur Datenkompetenz in den Humanities: Jupyter Books als “scalable Open Educational Resources”

⚠ Feinlernziel(e) dieses Kapitels

- Sie können aufzählen, welche Vorteile fallstudienbasierte Open Educational Resources haben.
- Sie können erklären, was ein Jupyter Book ist und auf welcher Infrastruktur es basiert.
- Sie können begründen, warum sich Jupyter Books gut als OER eignen.
- Sie können erläutern, was mit dem Begriff “scalable OER” gemeint ist und den Zusammenhang zu unterschiedlichen Levels of Expertise von Lernenden darstellen.

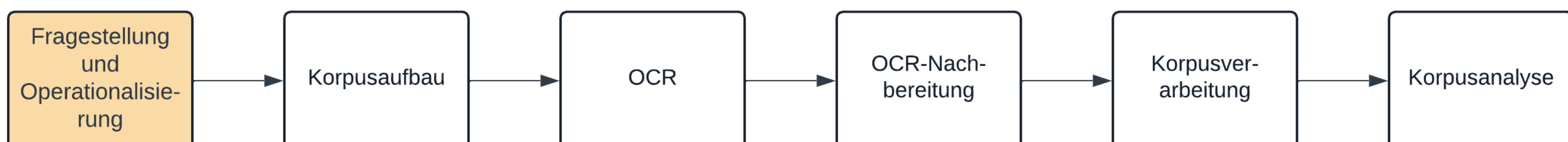
Fallstudien

Wir gehen von einer konkreten Fragestellung aus und setzen so **problemorientiertes Lernen** ein:

💡 Forschungsfrage

Lassen sich für die Spanische Grippe 1918/19 mit Fokus auf den Berliner Raum Muster in der öffentlichen Aufmerksamkeit ausmachen, die eine wellenartige Verlaufsform aufweisen?

Aus der Fragestellung entwickeln wir eine Forschungsstudie, orientiert an einem **DH-Workflow**:



Jupyter Book als OER

Bilddigitalisate von Text

Reiner Text, „Plain Text“

XML/TEI

Beispiel für eine XML-Kodierung nach TEI-Standard. Im Kopfbereich der Datei steht der <?xml>-Element, wo folgt das <text>-Element, in dem der Text mit Strukturmöglichkeiten wie z.B. Kapitelüberschriften, Fußnoten, Zitate und sprachliche Besonderheiten markiert ist. Ein Beispiel für die XML-Kodierung nach TEI-Standards ist im Dokument unter [Quellen](#) dargestellt.

Struktur

Jupyter Book (Python-Bibliothek) generiert HTML-Dateien aus **Markdown- und Jupyter Notebook-Dateien** sowie aus Bildern und bibliographischen Referenzen (BibTeX).

Die Inhalte sind in hierarchisch angelegte Kapitel strukturiert, die chronologisch durchlaufen oder direkt angesteuert werden können.

→ Ermöglichung des **aktiven Lernens**

Infrastruktur

Die Dateien liegen auf GitHub und werden über GitHub Actions als GitHub Pages deployed.

→ Schnelles **Feedback** (über Issues) und **Update** der Inhalte

Elemente

Die MyST-Markdown-Erweiterung erlaubt die multimodale Gestaltung von Inhalten: Abbildungen, GIFs, Tabellen, Querverweise, Informationsboxen etc.

In Markdown und Jupyter Notebooks können interaktive Elemente gestaltet werden und Jupyter Notebooks können komplett ausgeführt werden.

→ **Scannable** und **skimmable** Inhalte

Scalable OER

call the function to plot the frequencies plot_with_js(corpora_annotations_merged, search_terms=word_list)

Frequency of words [Influenza', 'Grippe', 'Grippeepidemie', 'Grippewelle', 'Grippeskranker', 'Grippepandemie']

Worteingabe für die Suche (für Cloud Mode und Local Mode)

```

text_input = input("Geben Sie die zu suchenden Wörter ein und trennen Sie sie durch Komma, # Convert the input to a list by splitting the input by comma, text_input = [x.strip() for x in text_input.split(',')]")
plot_with_js(corpora_annotations_merged, search_terms=text_input)
  
```

Es gibt drei **Modi der Ausführung**, angepasst an verschiedene Softwarekompetenz-Niveaus:

1. **Book-Only Mode** : Lesen der Fallstudie im Browser mit eingeschränkter Möglichkeit der Interaktion (z.B. bei Graphiken)

2. **Cloud Mode** : Ausführen der Jupyter Notebooks in Cloud-basierten Diensten, z.B. Google Colab. Eigene Daten müssen hochgeladen werden.

3. **Local Mode** : Herunterladen der Jupyter Notebooks und Ausführen in einer lokalen Umgebung mit lokal gespeicherten Daten.

< Previous

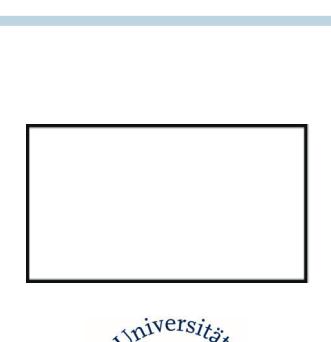
Next >



GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung



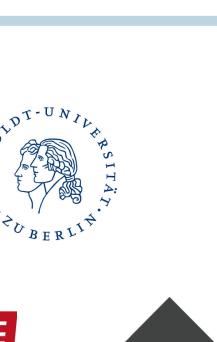
Finanziert von der
Europäischen Union
NextGenerationEU



FILMUNIVERSITÄT
BABELSBERG
KONRAD WERNER



FH;P
Fachhochschule Potsdam
University of Applied Sciences



FRAUNHOFER
FOKUS



FREIE
UNIVERSITÄT
BERLIN



TECHNISCHE
UNIVERSITÄT
BERLIN



GESELLSCHAFT
FÜR INFORMATIK

