



QUADRIGA

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

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

Zielgruppe
Die Fallstudie richtet sich an Geisteswissenschaftler:innen auf fortgeschrittener Qualifikationsstufe. Kenntnisse der Digital Humanities sind nicht erforderlich, wohl aber eine prinzipiell Neugier und Offenheit gegenüber digitalen Arbeitsweisen und quantitativen Forschungsansätzen.

Struktur der Fallstudie
Die Gliederung der Fallstudie nachfolgend. Insgesamt 6 Kapitel.

Struktur der Fallstudie

- Im 1. Schritt entwickeln wir die quantitative Analyse auf die Forschungsfrage.
- Im 2. Schritt bauen wir den Korpus auf (siehe Kapitel [Korpusaufbau](#)).
- Im 3. Schritt machen wir die Daten für die Analyse zugänglich (siehe Kapitel [Korpusaufbau](#)).
- Im 4. Schritt evaluieren wir die Qualität des Korpus (siehe Kapitel [Korpusaufbau](#)).
- Im 5. Schritt reichern wir den Korpus mit Metadaten an (siehe Kapitel [Korpusaufbau](#)).
- Im 6. Schritt führen wir die quantitative Analyse durch (siehe Kapitel [Korpusanalyse](#)).

Die Fallstudie schließt mit einer Reflexion und einem Ausblick (siehe Kapitel [Reflexion und Ausblick](#)).

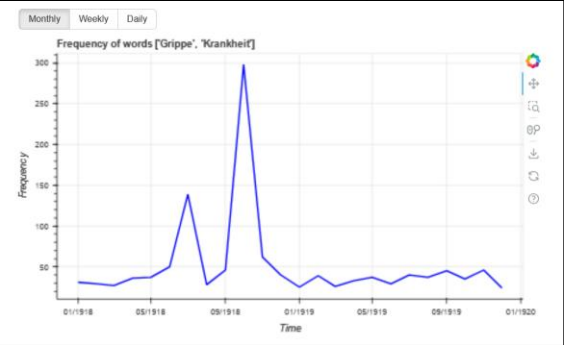
Next
[Lernziele dieser Fallstudie](#)



Die vorliegende Fallstudie bereitet – in Form eines [„Learner Book“](#) – den Prozess und die Ergebnisse eines Forschungsprojekts aus dem Digital Humanities didaktisch auf. Schritt für Schritt wird nachvollziehbar:

- wie eine **Forschungsfrage** entwickelt und operationalisiert wird,
- ein entsprechendes **Korpus** aufgebaut, bereinigt und angereichert wird,
- um schließlich **quantitative Analysen** auf diesem Korpus durchzuführen.

Anhand von historischen Tageszeitungen wird dabei eine Frage aus dem Feld der Digital History nachgegangen: **Welchen quantitativen Mustern folgte die Berichterstattung über die Spanische Grippe in den Jahren 1918/1919?**



Forschungsgeleitete Vermittlung von Datenkompetenz: Mediendidaktische Aufbereitung von Fallstudien zu Bildungsangeboten

Evgenia Samoilova, Henny Sluyter-Gäthje, Daniil Skorinkin, Hannes Schnatter, Peer Trilcke, Ulrike Lucke

Hochschule 2034

24.09.24 | Wiesbaden





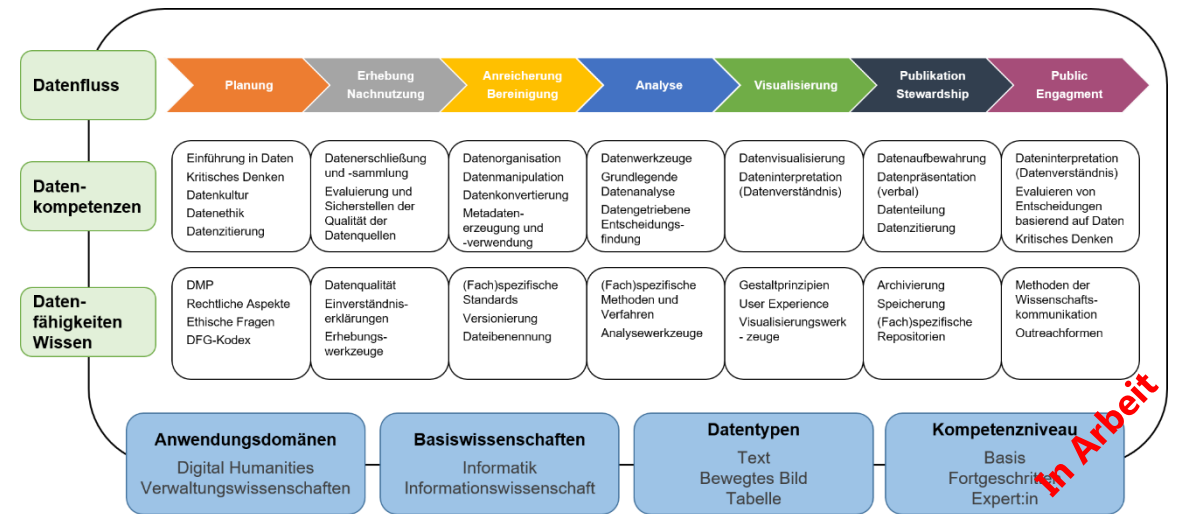
- Berlin-Brandenburgisches **Datenkompetenzzentrum** für Digital Humanities, Verwaltungswissenschaft, Informatik und Informationswissenschaft
- **QUADRIGA-Zielgruppen:**
 - **Promovierende** und **promovierte Wissenschaftler:innen** aller Karrierestufen
 - Lehrende
 - mit dem Fokus auf **Digital Humanities** und **Verwaltungswissenschaft**
- **Informatik und Informationswissenschaft** unterstützen die Anwendungsdomänen (z.B. in den Bereichen Data Science und Data Stewardship).
- **Datentypen:** Text | Bewegtes Bild | Tabelle

QUADRIGA (11.23-11.26)



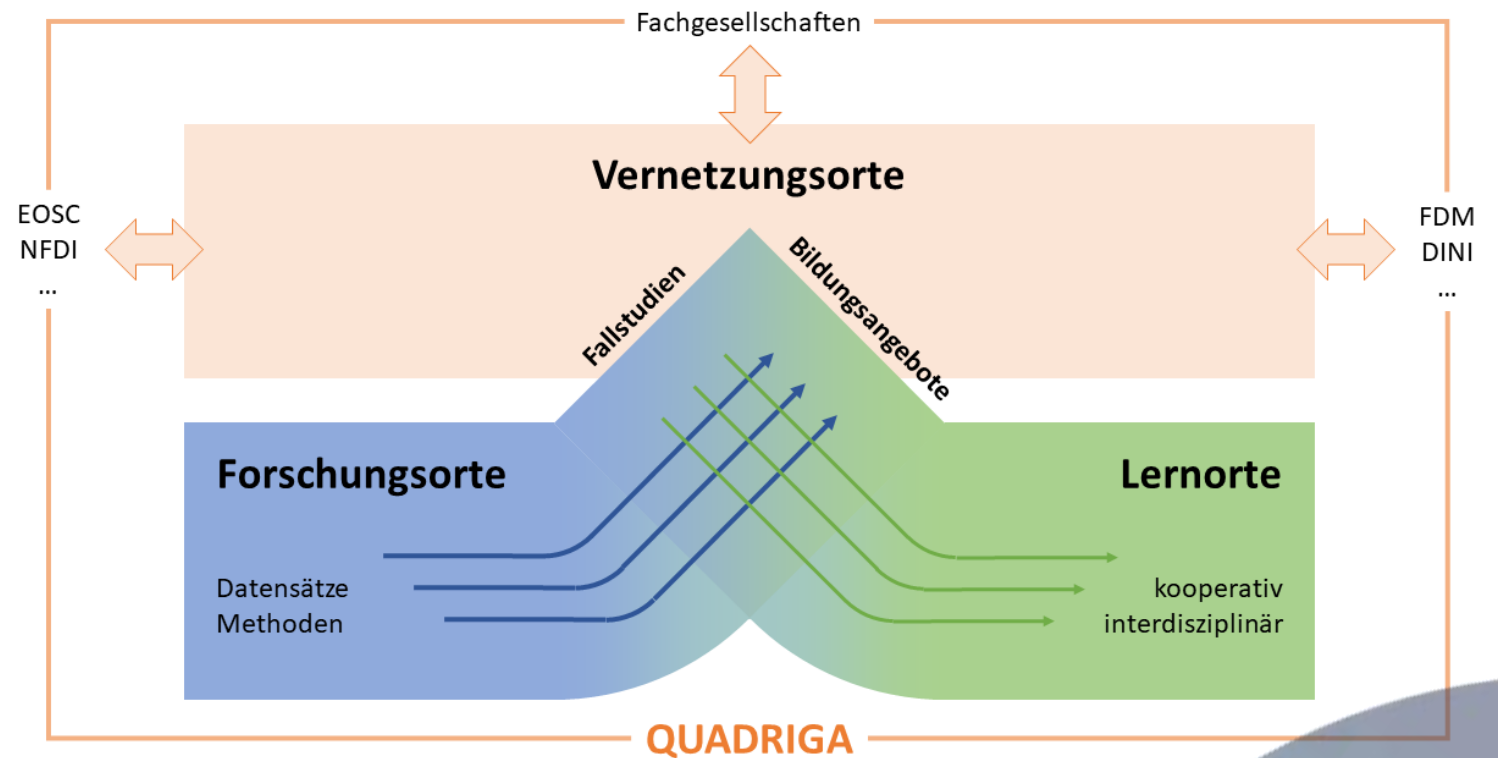
- Datenkompetenz beinhaltet kritische und wohlüberlegte Sammlung, Verwaltung, Bewertung und Anwendung von Daten, die über kognitive Aspekte hinaus auch affektive Einstellungen umfasst [Ri15, SBH19]
- Die Vermittlung von Datenkompetenzen erfolgt derzeit häufig über einen Top-down Ansatz: Datenkompetenzframeworks [Bu24]

DATENKOMPETENZ



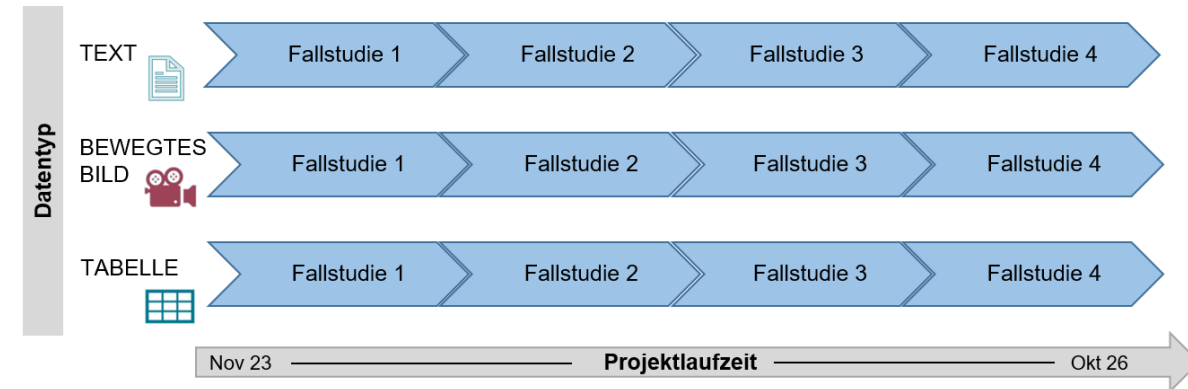
Datenkompetenzframeworks im Projekt QUADRIGA (Bu24)

- In **Forschungsorten** entstehen Fallstudien aus Datenbasierter Forschungstätigkeit.
- Diese werden in **Lernorten** zu mediengestützten Bildungsangeboten aufbereitet.
- Die **Vernetzungsorte** sorgen für die Integration von neuen Fallstudien und Bildungsangeboten in relevante Fachgesellschaften und Initiativen.



- Fallstudien stellen reale (oder realistische) Ereignisse oder Probleme dar, damit Lernende die Komplexität, Mehrdeutigkeit und Unsicherheiten nachvollziehen können, mit denen die ursprünglichen Beteiligten in der Fallstudie (Forschende) konfrontiert waren [Fo01].
- Fallstudien in QUADRIGA: Untersuchung einer konkreten Forschungsfrage oder eines spezifischen Problems innerhalb einer Studie.
- **Fallstudien** beschreiben spezifische **Aktivitäten** oder **Schritte** aus der realen Forschungspraxis und bieten **narrative Einblicke**.

OFFENE BILDUNGSRESSOURCEN (OER): FALLSTUDIEN





Lassen sich für die Spanische Grippe 1918/1919 mit Fokus auf den Berliner Raum Muster in der öffentlichen Aufmerksamkeit ausmachen, die eine wellenartige Verlaufsform aufweisen?

Forschungsprozess der Studie:

- wie eine Forschungsfrage entwickelt und operationalisiert wird,
- **ein entsprechendes Korpus aufgebaut, bereinigt und angereichert wird,**
- um schließlich quantitative Analysen auf diesem Korpus durchzuführen.

Fragestellung

Dass Pandemien ihre je eigenen Verlaufsprofile entwickeln, konnte Zeitgenoss:innen vor wenigen Jahren am Verlauf der COVID19-Pandemie miterleben. Schnell wurde die Rede von den "Wellen" der Corona-Pandemie sprichwörtlich. Auch während der Spanischen Grippe wurde bereits von "Wellen" bzw. von "Waves" gesprochen und diese spezifische Verlaufsform graphisch dokumentiert (siehe Fig. 1).

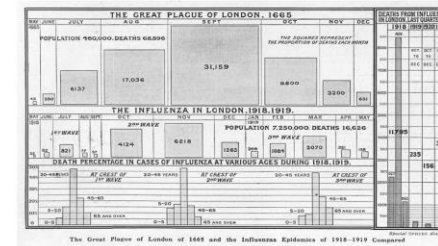


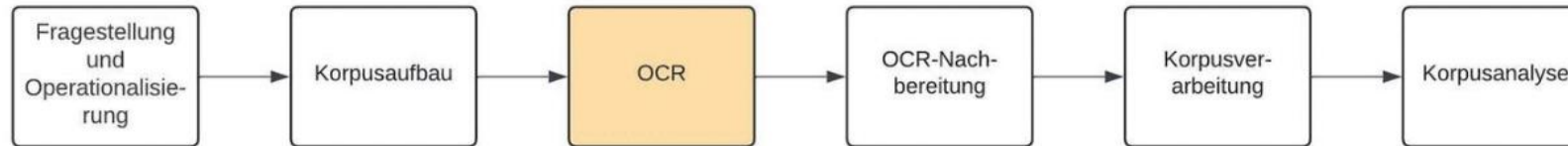
Fig. 3 Vergleichende Grafik zum Verlauf der Pest (1665) und der Spanischen Grippe (1918/19) in London.
Quelle: [Staveley-Wadham, n.d.]

Aus Perspektive einer medienwissenschaftlich informierten historischen Epidemiologie überlagern sich dabei unterschiedliche Wellen: Während 'Fallzahlen' (z.B. die Anzahl der Infektionen oder die Anzahl der Todesfälle), wie sie in Fig. 1 dargestellt werden, ein in erster Linie medizinisch zu erhebendes Maß sind, sind die unterschiedliche Intensität und Extensität der öffentlichen Wahrnehmung des pandemischen Geschehens ein Untersuchungsgegenstand der (historischen) Medienwissenschaft. Auch hier ist davon auszugehen, dass sich charakteristische "Medienzeiten" oder eben: "Medienwellen" zeigen: Mal dominiert die Pandemie den öffentlichen Diskurs, mal tritt sie in der öffentlichen Wahrnehmung zurück.

Diese Fallstudie führt durch eine Digital Humanities-Forschungsprojekt, das aus medienhistorischer Perspektive den Verlauf der "Medienwellen" während der Spanischen Grippe 1918/19 – bekannt auch als "the Mother of All Pandemics" [Taubenberger and Morens, 2006] – in Berlin und Brandenburg untersucht hat. Das Forschungsprojekt wurde von der folgenden Forschungsfrage geleitet:

Fallstudie, Datentyp Text (noch in Bearbeitung):
<https://dh-network.github.io/quadriga/markdown/intro.html>

- Die Entwicklung der Lernziele wurde durch die sequenzielle Struktur der empirischen Schritte/Aufgaben oder ausgewählten Aspekten der Fallstudie geleitet (MTS21, Me97).



Beschreibung der Aufgabe in der Fallstudie	Groblernziel	Feinlernziele
<p>Homogenisierung der Daten:</p> <p>Bild/PDF zu TXT: Mittels einer OCR-Pipeline werden aus den im Imageformat (JPG o.Ä.) oder als PDF vorliegenden Dateien Textdaten erstellt und im TXT Format abgespeichert.</p>	<p>1. Lernende können Textdaten mit Jupyter Notebooks homogenisieren und die Qualität der Datensammlung anhand eines Samples bewerten</p>	<p>1. Die Lernenden können die Grundprinzipien und den Workflow einer OCR-Pipeline erläutern. Sie können erklären, wie Bilder und PDF-Dokumente in bearbeitbare Textformate umgewandelt werden.</p> <p>2. ...</p> <p>n. ...</p>

Grob- und Feinlernziele aus einem Ausschnitt für die Fallstudie "Spanische Grippe"



- Kontrolle über Lernerfahrung durch die buchähnliche Struktur
- Modularisierung: Modul=Kapitel=Aufgabe
- Interaktive Lehrbücher fokussieren auf **Text** und **interaktive Elemente**; andere Medien (z.B. Screencasts) bei Bedarf

OERS: INTERAKTIVE LEHRBÜCHER

The screenshot shows the QUADRIGA interface for a case study titled "Quantitative Analyse der Medienwellen der Spanischen Grippe (1918/19). Eine Fallstudie". The left sidebar contains a search bar and a table of contents with sections like "Präambel", "Fragstellung und Operationalisierung", "Korpusaufbau", "OCR", "Korpusverarbeitung", and "NLP". The main content area includes a title, a historical newspaper clipping titled "Die Grippe wüthet weiter", a list of bullet points, a "Zielgruppe" section, a "Struktur der Fallstudie" section with a flowchart, and a list of six steps for the case study.

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

Die vorliegende Fallstudie bereitet – in Form eines „Jupyter Books“ – den Prozess und die Ergebnisse eines Forschungsprojekts aus den Digital Humanities didaktisch auf. Schritt für Schritt wird nachvollziehbar,

- wie eine **Forschungsfrage** entwickelt und operationalisiert wird,
- ein entsprechendes **Korpus** aufgebaut, bereinigt und angereichert wird,
- um schließlich quantitative **Analysen** auf diesem Korpus durchzuführen.

Anhand von historischen Tageszeitungen wird dabei eine Frage aus dem Feld der Digital History nachgegangen: **Welchen quantitativen Mustern folgte die Berichterstattung über die Spanische Grippe in den Jahren 1918/1919?**

Zielgruppe

Die Fallstudie richtet sich an Geisteswissenschaftlerinnen auf fortgeschrittener Qualifikationsstufe. Kenntnisse der Digital Humanities sind nicht erforderlich, wohl aber eine prinzipiell Neugier und Offenheit gegenüber digitalen Arbeitsweisen und quantifizierten Forschungsansätzen.

Struktur der Fallstudie

Die Gliederung der Fallstudie lässt sich jederzeit durch die Menüleiste links im Browser nachvollziehen. Insgesamt vollzieht die Fallstudie 6 Schritte:

```
graph LR; A[Fragstellung und Operationalisierung] --> B[Korpusaufbau]; B --> C[OCR]; C --> D[OCR-Nachbearbeitung]; D --> E[Korpusverarbeitung]; E --> F[Korpusanalyse];
```

Fig. 1 Flussdiagramm der Fallstudie, die sich aus sechs Arbeitspaketen zusammensetzt.

- Im **1. Schritt** entwickeln wir eine Forschungsfrage und operationalisieren diese Forschungsfrage für die quantitative Analyse, entwickeln also ein Konzept, wie wir mittels Meßoperationen zu einer Antwort auf die Forschungsfrage kommen.
- Im **2. Schritt** bauen wir ein Korpus aus Textobjekten für die Analyse auf, das zunächst aus PDF-Dateien besteht (siehe Kapitel ["Korpusaufbau"](#)).
- Im **3. Schritt** machen wir die Textobjekte im Korpus, die zunächst nur als Bilddateien vorliegen, mittels Optical Character Recognition (OCR) maschinenlesbar (siehe Kapitel ["OCR – Vom Bild zum Text"](#)).
- Im **4. Schritt** evaluieren wir die OCR-Ergebnisse und testen Optionen zur Nachkorrektur (siehe Kapitel ["Nachkorrektur der OCR-Ergebnisse"](#)).
- Im **5. Schritt** reichern wir mithilfe von Verfahren des Natural Language Processing (NLP) die Textobjekte im Korpus mit linguistischen Informationen an. (siehe Kapitel ["Korpusverarbeitung – Von Strings zu Token"](#)).
- Im **6. Schritt** führen wir die quantitativen Analysen auf dem Korpus durch und visualisieren die Ergebnisse (siehe Kapitel ["Korpusanalyse"](#)).

Die Fallstudie schließt mit einer Reflexion und einem Ausblick (siehe Kapitel ["Reflexion und Resümee"](#)).

Fallstudie, Datentyp Text (noch in Bearbeitung):
<https://dh-network.github.io/quadriga/markdown/intro.html>

Die Struktur und die einzelnen Komponenten der Bildungsangebote orientieren sich am 4C/ID-Modell/Cognitive Apprenticeship [Me97, CBH91, WJ07]:

- Präsentation der **Aufgaben/Probleme**
- Unterstützende Informationen (Theorie + Beispiele)
- **Interaktive Übungen** der Lernaufgaben/Probleme
- Zusammenfassung und **Reflexion**
- Prozedurale Information (für Routineaspekte der Lernaufgaben)
- Assessment

LEHRBÜCHER: STRUKTUR



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

Nach dem vorherigen [Kapitel](#) haben wir also ein Korpus als Sammlung gescannter Bilder. Ein Korpus in dieser Form ist jedoch noch **nicht maschinenlesbar** und kann nicht direkt verarbeitet werden. In diesem Kapitel lernen wir, wie man mit OCR **Bilder in Text umwandelt**.

```

graph LR
    A[Fragestellung und Operationalisierung] --> B[Korpusaufbau]
    B --> C[OCR]
    C --> D[OCR-Nachbereitung]
    D --> E[Korpusverarbeitung]
    E --> F[Korpusanalyse]
  
```

Fig. 6 Flussdiagramm der Fallstudie. Wir befinden uns im dritten Arbeitspaket.

Zunächst werden wir lernen, [was OCR ist](#), warum wir es brauchen und wie es funktioniert. Außerdem werden wir einen Überblick über einige OCR-Tools geben.

Anschließend werden wir [OCR in Python mit PyTesseract](#) durchführen, einem kostenlosen und quelloffenen OCR-Tool.

Schließlich werden wir die Metriken kennenlernen, die zur Messung der [OCR-Qualität](#) verwendet werden, und [Qualitätsmessungen](#) durchführen.

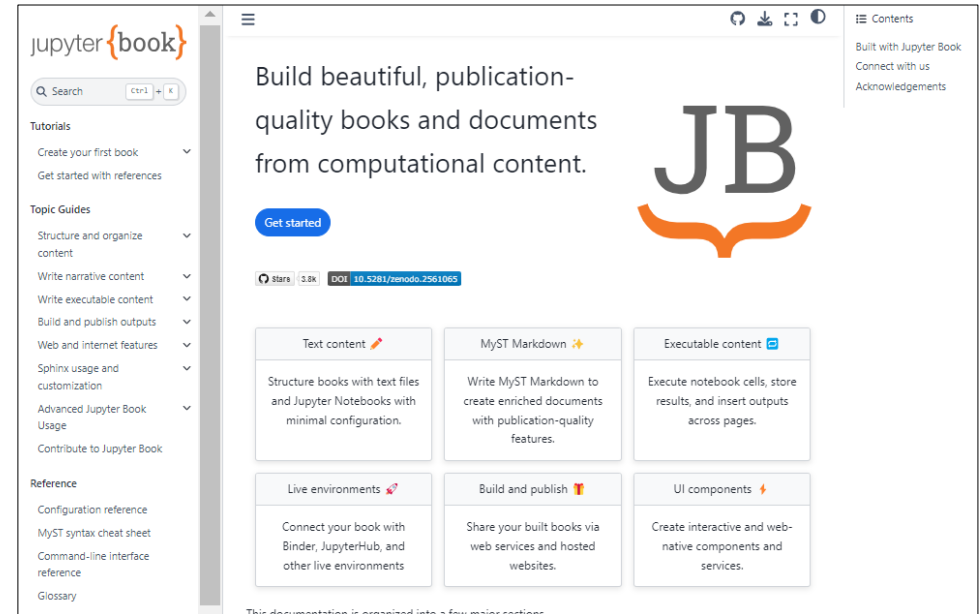
Previous [Resümee](#) Next [OCR als Methode, um Text maschinenlesbar zu machen](#)

Fallstudie, Datentyp Text (noch in Bearbeitung): <https://dh-network.github.io/quadriga/markdown/intro.html>

Sammlung von Markdown-Dateien und Jupyter Notebooks:

- Unterstützt die Ausführung und Anpassung von Quizen, Code, Gleichungen und Visualisierungen
- Unterstützt Julia, Python und R (für zusätzliche Sprachen sind Plugins erforderlich)
- Dynamische Inhaltsaktualisierung (Setzt technische Kenntnisse für die Administration voraus)
- Setzt technische Kenntnisse (Python, Markdown, Git, GitHub) voraus

IMPLEMENTIERUNG: JUPYTER BOOKS

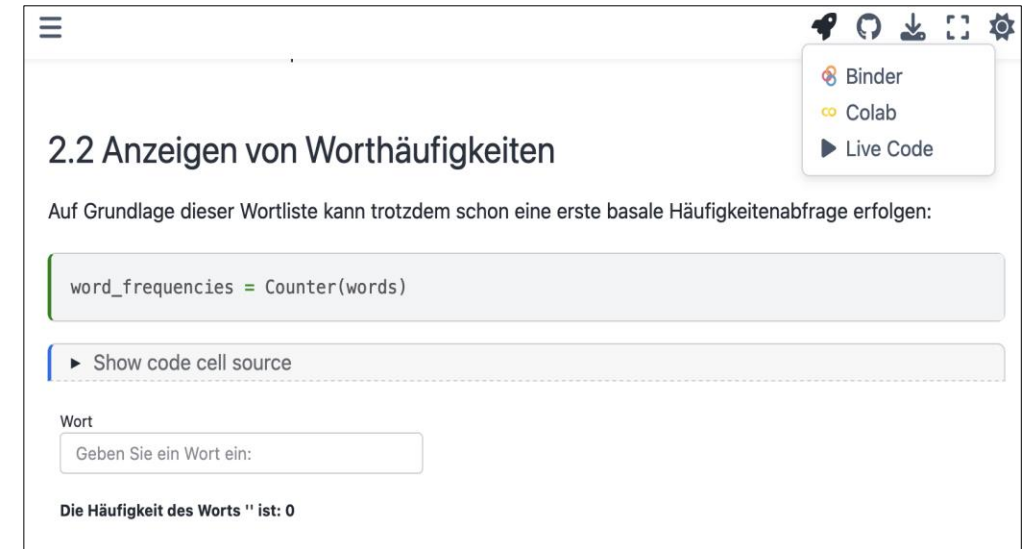


jupyterbook.org

Interaktive Nutzung der Fallstudie für unterschiedliche Bedarfe/Kompetenzniveaus:

- Traditionelles Buch (nicht interaktiv)
- Webseite mit interaktiven Elementen (Live Coding)
- Jupyter Notebooks via Binder/Google Colab/Jupyter Hub
- Download von Büchern oder Notebooks (ermöglicht lokale Anwendung)

IMPLEMENTIERUNG: JUPYTER BOOKS



The screenshot shows a Jupyter Notebook interface. At the top right, there are navigation icons and a dropdown menu with options: Binder, Colab, and Live Code. The main content area has the heading "2.2 Anzeigen von Worthäufigkeiten". Below the heading, there is a text block: "Auf Grundlage dieser Wortliste kann trotzdem schon eine erste basale Häufigkeitenabfrage erfolgen:". This is followed by a code cell containing the Python code: `word_frequencies = Counter(words)`. Below the code cell is a button labeled "Show code cell source". Underneath, there is a section titled "Wort" with an input field containing the text "Geben Sie ein Wort ein:". Below the input field, it says "Die Häufigkeit des Worts '' ist: 0".

Fallstudie, Datentyp Text (noch in Bearbeitung):
<https://dh-network.github.io/quadriga/markdown/intro.html>

- Die Verzahnung von Forschungs-, Lern- und Vernetzungsorten entlang von Fallstudien liefert eine strukturelle und prozessuale Basis für die Bereitstellung von Bildungsangeboten.
- Durch den Einsatz des Vier-Komponenten- Instructional-Design-Modells und des Cognitive Apprenticeship-Ansatzes werden bei der didaktischen Aufbereitung die Forschungsfragen und Aufgaben aus realen Forschungsszenarien in den Mittelpunkt der Lernerfahrung gerückt.
- Interaktive Lehrbücher wurden als Format gewählt und in Jupyter Books implementiert, um eine hohe Kontrolle über die Interaktion mit den Materialien sowie die Möglichkeit, Inhalte dynamisch zu aktualisieren, zu bieten.
- Die Implementierung als interaktive Lehrbücher wurde exemplarisch als Jupyter Book "Spanische Grippe" erläutert, ist jedoch unmittelbar **auf andere Themen oder Werkzeuge übertragbar**.



HERAUSFORDERUNGEN NÄCHSTE SCHRITTE

- Erprobung
- Vervollständigung: Fokus auf Assessment und interaktive Übungen
- Modularisierung
- Verknüpfung verschiedener Fallstudien
- Navigator: Metadaten-basierte Suche

- [Bu24] Buchholz, B. et al.: Umsetzungskonzept QUADRIGA: Berlin-Brandenburgisches Datenkompetenzzentrum für Digital Humanities, Verwaltungswissenschaft, Informatik und Informationswissenschaft. Zenodo, 2024.
<https://doi.org/10.5281/zenodo.10805016>
- [CBH91] Collins, A.; Brown, J. S.; Holum, A.: Cognitive Apprenticeship: Making Thinking Visible. American Educator: The Professional Journal of the American Federation of Teachers 15, 1991
- [Fo01] Foran, J.: The Case Method and the Interactive Classroom. Thought & Action 1/17, S. 41–50, 2001.
- [Me97] van Merriënboer, J. J. G.: Training complex cognitive skills. A four-component instructional design model for technical training. Educational Technology Publ, Englewood Cliffs, N.J, 1997.
- [Ri15] Ridsdale, C. et al.: Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report.
https://www.researchgate.net/publication/284029915_Strategies_and_Best_Practices_for_Data_Literacy_Education_Knowledge_Synthesis_Report, Stand: 16.07.2023.
- [SBH19] Schüller, K.; Busch, P.; Hindinger, C.: Future Skills: Ein Framework für Data Literacy. Zenodo, 2019.
- [WJ07] Woolley, N. N.; Jarvis, Y.: Situated cognition and cognitive apprenticeship: a model for teaching and learning clinical skills in a technologically rich and authentic learning environment. Nurse education today 1/27, S. 73–79, 2007.



Vielen Dank für die Aufmerksamkeit.

evgenia.samoilova@uni-potsdam.de

www.quadrigea-dk.de

quadrigea-management@listserv.dfn.de