

Gender-Aspekte der Data Science

Sitzung 1b: Einführung Wikipedia & Wikidata

19.03.2026

Dozentin: Dr. Daria Tisch

Roadmap

	Einführung: Gender & Data Science
19.03.2026	Einführung in Wikipedia und Wikidata; Arbeiten mit R; Bildung von Projektteams
16.04.2026	Der männliche Prototyp Themensuche; Daten extrahieren; Forschungsdatenmanagement
30.04.2026	Datenfeminismus Versionskontrolle, Datenanalyse
28.05.2026	Geschlechterbezogene Verzerrungen im gesamten Datenprozess Projektvorstellungen: Zwischenstand; Datenanalyse; Peer Feedback
11.06.2026	Thematische Vertiefung: Gender Bias im Gesundheitsbereich Projektvorstellungen: Zwischenstand; Datenanalyse; Peer Feedback
18.06.2026	Thematische Vertiefung: Algorithmische Diskriminierung Projektvorstellungen: Zwischenstand; Datenanalyse; Peer Feedback
25.06.2026	Thematische Vertiefung: wird gemeinsam bestimmt Gemeinsame Erarbeitung eines Posters zu Gender & Data Science
09.07.2026	Projektpräsentationen & Reflexion

Wer kennt...

- Wikipedia?
- Wikidata?

Wikidata, ein Wissensgraph

The screenshot shows the Wikidata page for Bill Gates (Q5284). The page is in English and displays the following information:

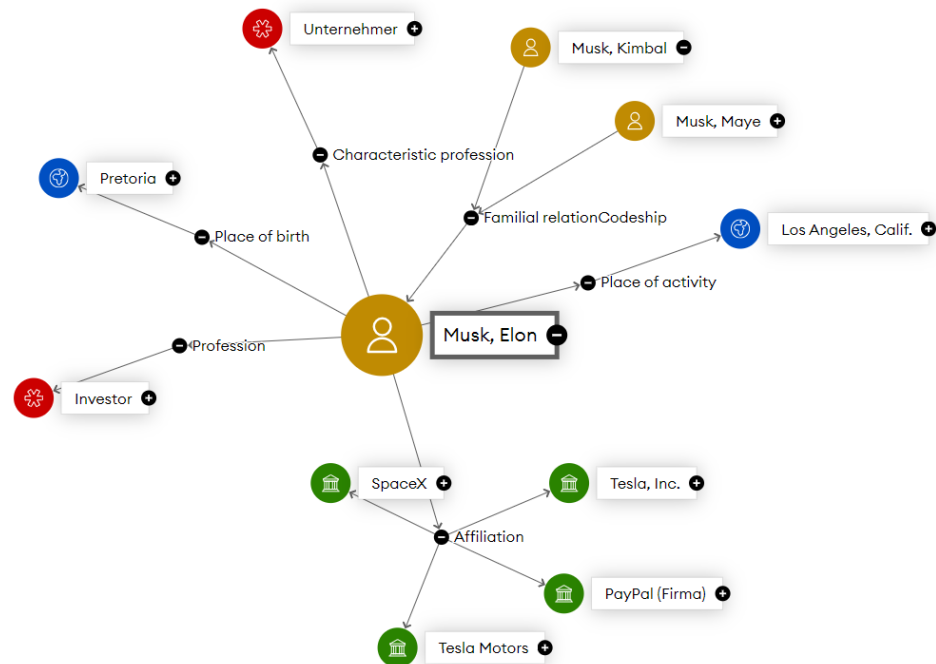
- Item:** Bill Gates (Q5284)
- Description:** American businessman, investor, and philanthropist (born 1955)
- Statements:**
 - instance of:** human (1 reference)
 - image:** Bill Gates - 2023 - F002021-007902 (cropped).jpg (1,393 × 2,091; 869 KB) (point in time: 2023) (0 references)
 - signature:** Bill Gates signature.svg (594 × 141; 10 KB) (0 references) and Bill Gates signature (short).svg (200 × 80; 10 KB) (point in time: 1976) (0 references)
 - sex or gender:** male (4 references)
 - country of citizenship:** United States of America (2 references)
 - name in native language:** Bill Gates (English) (1 reference)
 - birth name:** William Henry Gates III (Footish) (1 reference)

The page also includes a sidebar with navigation options and a list of languages.

- Wikidata ist eine freie, gemeinschaftliche und mehrsprachige Datenbank.
- Wikidata sammelt strukturierte Daten zur Unterstützung von Wikipedia und anderen Projekten.
- Sowohl von Menschen als auch von Maschinen lesbar und bearbeitbar.
- Jedes Element ist mit Attributen verknüpft, was komplexe Abfragen und Datenextraktionen ermöglicht.

Was sind Knowledge Graphs (Wissensgraphen)?

Wissensgraphen sammeln und vermitteln Wissen. Die Knoten in Wissensgraphen stellen die relevanten Entitäten dar, während die Kanten die Beziehungen zwischen diesen Entitäten darstellen.



<https://explore.gnd.network/en/gnd/1062481119/rerelations>

Wikidata für die Lehre/Übung

- **Inhalt:** Liefert strukturierte Daten zu verschiedensten Themen
- **Abfragesprache:** SPARQL
- **Linked Open Data:** Verknüpft Daten innerhalb von Wikidata sowie mit externen Datenbanken (z. B. X, ORCID, Film-Datenbanken).
- **Wachsend:** Immer mehr Daten.
- **Mehrsprachigkeit:** Nutzung ohne Sprachbarrieren.
- **Kostenlos:** Frei verfügbar für Nutzung und Weiterverbreitung.
- **EditorInnen:** Die Daten werden meist von anonymen Freiwilligen gepflegt
- **Achtung:** Genauigkeit und Richtigkeit ist nicht immer garantiert.

So kann Wikidata genutzt werden

- **SPARQL:** Die Sprache zur Formulierung von Fragen (Queries) an die Wissensdatenbank
- **Wikidata Query Service:** Online-Tool zum Ausführen von Abfragen (<https://query.wikidata.org/>)
- **WikidataR:** Eine Schnittstelle (API) für die Statistiksoftware R, um Abfragen direkt zu senden.*
- **Lernressourcen:** Wikidata bietet Tutorials und Beispiel-Abfragen an. (z.B. https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial)

*Shafee, Thomas, Oliver Keyes, Serena Signorelli, Alex Lum, Christian Graul, and Mikhail Popov. 2022. "WikidataR: API Client Library for 'wikidata'." <https://github.com/TS404/WikidataR>.

Teamprojekt = Studienarbeit = Prüfungsleistung



- Ziel bis Ende des Semesters: Gemeinsames Projekt zu Gender-Bias/Gender-Data-Gap auf Basis von Wikipedia oder Wikidata Daten
- Alle Mitglieder eines Teams sind für die Arbeit des Teams verantwortlich. Jedes Mitglied des Teams erhält dieselbe Note. Bei der mündlichen Präsentation muss jeder Studierende in der Lage sein, das gesamte Teamprojekt vorzustellen und Fragen dazu zu beantworten.
- Abgabe als Link zu einem github repository
 - R-Code
 - Manuskript (z.B. Quarto Dokument oder Markdown): Text und Code kombinieren
- Teams zu max. 3 Personen zusammen

Ideen für Projekte, die Gender-Bias auf Wikipedia untersuchen:

- Gender-Bias in der Darstellung von Berufen auf Wikipedia: Wie stark spiegeln Wikipedia-Artikel typische Geschlechterrollen in Berufsbildern wider?
- Gender-Bias in medizinischen Wikipedia-Artikeln: Spiegelt Wikipedia gender-spezifische Unterschiede in Gesundheitsinformationen korrekt wider?
- Gender-Bias in der politischen Repräsentation auf Wikipedia: Sind Politikerinnen auf Wikipedia unterrepräsentiert oder anders dargestellt als ihre männlichen Kollegen?
- Gender-Bias in der Sichtbarkeit von Wissenschaftlerinnen und Wissenschaftlern auf Wikipedia:
- Gender-Bias in historischen Artikeln: Werden Frauen und Männer in historischen Kontexten auf Wikipedia unterschiedlich dargestellt?
- Gender-Bias in Artikeln zu Filmen, Serien und Musiker*innen: Welche Geschlechterrollen werden in Wikipedia-Artikeln über Medien und Popkultur reproduziert?
- Gendergerechte Sprache in Wikipedia-Artikeln
- Gender-Bias in der Verlinkung von Personen auf Wikipedia: Werden männliche CEOs stärker verlinkt und sind damit sichtbarer auf Wikipedia?

Ideen für Projekte, in denen Wikipedia nur als Datenbasis genutzt wird:

- **Bias in Datenerhebung:** alle relevanten Artikel oder nur eine Teilmenge (z. B. „Top 100“). Welche Gruppen werden durch die Auswahl über- oder unterrepräsentiert? Welche Verzerrung entsteht?
- **Bias in Datenselektion:** filtern nach Kriterien (z. B. Artikel mit Bildern, Artikel >500 Wörter, nur bestimmte Länder). Wie erzeugen Filter, Kategorien oder Auswahlkriterien Gender-Bias?
- **Bias in der Datenauswertung:** Wie können Analyse-Methoden unbewusst Bias erzeugen?

Bildung von Teams

- 5 min Einzelarbeit: Welche Themen interessieren mich?
- Jede Person stellt erste Ideen vor
- Bildung von Teams