

von Andreas Henrich und Tobias Gradl

Erforschung der Weltenwanderer

Über die Nutzung der Wikipedia zur Identifikation grenzüberschreitender Lebensläufe

Für die Geschichtswissenschaften können die Datenbestände von Wikipedia und Wikidata wertvoll sein. Dies betrifft etwa die Erforschung von Biografien: War jemand im 19. Jahrhundert bereits international vernetzt? Die Nationalbiografien geben hierüber mitunter nur unzureichend Auskunft, da die nationale Perspektive überwiegt. Wikipedia und Wikidata liefern wertvolle zusätzliche Informationen. Für die Auswertung der Daten entwickelt die Bamberger Medieninformatik derzeit einen Prototypen, Historiker und Informatiker arbeiten Hand in Hand.

Wanderer zwischen den Welten oder stets zuhause geblieben? Bei der Untersuchung personenbezogener Eigenschaften existieren Indikatoren, die Rückschlüsse auf die erwartete Internationalität von Personen erlauben. So wäre man bei Berufsgruppen wie Missionaren oder Diplomaten überrascht, wenn eine zugeordnete Biografie keine Grenzüberschreitung beinhaltet. Weniger offensichtliche Anzeichen umfassen beispielsweise die Religion, den Ausbildungsstand oder den sozialen Status einer Person.

Gemeinsam mit Historikern des Leibniz-Instituts für Europäische Geschichte (IEG) in Mainz untersuchen Bamberger Medieninformatiker Möglichkeiten zur Unterstützung historischer Forschung durch die Analyse der Datenbestände von Wikipedia und Wikidata. Motiviert wird die interdisziplinäre Zusammenarbeit durch die Frage, inwieweit die qualitative, historische Untersuchung grenzüberschreitender Lebensläufe in europäischen Nationalbiografien durch weitere biografische Quellen und die Anwendung quantitativer Verfahren unterstützt werden kann. Die Mainzer Historiker interessieren sich dabei vor allem für Personen des 19. Jahrhunderts, die über politische Grenzen

hinweg Spuren in verschiedenen Ländern hinterließen. Die Besonderheit bei der Analyse solcher transnationaler Lebensläufe besteht nun darin, dass sich diese oft erst nach der Betrachtung mehrerer biografischer Quellen als solche zu erkennen geben: Biografische Texte – insbesondere in den jeweiligen Nationalbiografien – beschreiben eine Person aus einer nationalen Perspektive und vernachlässigen oder verschweigen oft Einflüsse der Person auf andere Nationen oder Kulturkreise.

Jakob Heinrich Schiff,
jüdischer Bankier



Transnationale Lebensläufe im 19. Jahrhundert: ein Beispiel

Betrachten wir als Beispiel den 1847 geborenen, jüdischen Bankier Jakob Heinrich Schiff. Nach Geburt und Kindheit in Frankfurt migrierte dieser zunächst im Alter von 18 Jahren und – nach drei Jahren in Hamburg und Frankfurt – 1875 ein weiteres Mal in die USA. Der rund 950 Wörter umfassende Eintrag zu Jakob Schiff in der deutschsprachigen Wikipedia gibt Aufschluss über die Transnationalität in seinem Leben und betont insbesondere die beruflichen Stationen als Bankier. Der mit rund 2.350 Wörtern umfassendere englischsprachige Artikel unterscheidet sich vor allem durch die differenzierte Betrachtung des Philanthropen und Geschäftsmanns und seine weitreichende finanzielle Unterstützung Japans im Krieg gegen Russland 1904 bis 1905. Obwohl beide Artikel die wesentlichen Aspekte seines Lebens jeweils umfassen, so enthalten diese auch Informationen, die dem jeweils anderen fehlen: So erwähnt nur der deutsche Ein-

trag Schiffs Brüder und beschreibt seine Rolle als Gründungsmitglied der Goethe-Universität Frankfurt am Main. Im englischsprachigen Beitrag fehlen diese Informationen, wohingegen eine detaillierte Auflistung der von ihm unterstützten, in den Vereinigten Staaten ansässigen Einrichtungen vorgelegt wird. Insgesamt aber bieten die Texte der Wikipedia – besonders in ihrer Kombination – detaillierte Informationen über das Leben von Jakob Schiff in Deutschland und den USA sowie seine Position zu Japan und Russland.

Da Wikipedia-Artikel grundsätzlich nicht an Qualitäts- oder Langfristigkeitsanforderungen gebunden sind, können Historiker für eine fundierte Auseinandersetzung mit dem Leben von Jakob Schiff auf einen Eintrag der Datenbank von *immigrantentrepreneurship.org* zurückgreifen. In dieser führt das 1987 gegründete Deutsche Historische Institut Washington (DHI) wissenschaftlich



fundierte und redaktionell geprüfte Einträge zu Deutsch-Amerikanischen Unternehmern. Schiff ist dort mit einem über 10.000 Worte umfassenden Artikel verzeichnet. Und obwohl der Artikel eine historisch differenzierte Analyse seines Lebens und Wirkens liefert: Einige in der Wikipedia verfügbare Informationen fehlen hier.



Wikipedia als biografische Quelle

Das Beispiel Jakob Schiffs erlaubt zwei direkte Rückschlüsse: Erstens, dass oft erst durch die Kombination nationaler Perspektiven ein objektiver Gesamteindruck über eine transnationale Biografie entsteht. Zweitens kann die Wikipedia zwar aufgrund ihrer Intention und Ideologie nicht als direkte Quelle einer historischen Forschung dienen; für die Identifikation und initiale Analyse der Transnationalität von Biografien bietet die Wikipedia jedoch den Vorteil einer – insbesondere gegenüber

den Nationalbiografien – oft weitaus geringeren nationalen Prägung. Vor allem jedoch stehen Wikipedia-Artikel in den verschiedensten Sprachen frei und ohne Zugriffshürden zur Verfügung: Allein die deutschsprachige Wikipedia beinhaltet etwa 560.000 Einträge zu Personen und bietet so eine für die Anwendung quantitativer Methoden mehr als ausreichende Datenbasis. Im Vergleich hierzu stellen die ebenfalls beachtlichen Bestände der *Allgemeinen Deutschen Biographie* (ADB) rund 26.500 Einträge zu Personen bis einschließlich des 19. Jahrhunderts; die *Neue Deutsche Biographie* (NDB) derzeit knapp 22.000 Einträge.

Auf Basis der Wikipedia-Daten könnten die Historiker nun beispielsweise bei ihren Fragen nach den Hintergründen und Motiven für Transnationalität unterstützt werden. Neben Berufen eignet sich eine Vielzahl von Eigenschaften wie Geschlecht, Religion, sozialer Status, Ort und Art der Ausbildung oder die politische Orientierung zur Bildung von Personengruppen, welche dann miteinander verglichen werden können. Durch die große Menge biografischer Texte kann hierbei erreicht werden, dass selbst zu den fein differenzierten Gruppen zumeist eine ausreichende Zahl korrespondierender Personen identifiziert werden kann. So können erste fachliche Rückschlüsse gezogen werden, um gegebenenfalls Impulse für die eigentliche, historische Forschung zu setzen.

Investigating Travelers Between Worlds

Using Wikipedia to identify world-spanning life paths

The wealth of data available on Wikipedia and Wikidata can be highly valuable to research in the humanities, especially in the field of biographical research. Who, for example, was already cultivating international networks in the 19th century? Because biographies tend to be dominated by a particular national perspective, questions like this are hard to answer adequately without taking the larger picture into account. Wikipedia and Wikidata can provide valuable supplementary information, and in Bamberg's media informatics programme, historians and computer scientists are working hand in hand to develop a prototype for the analysis of such data.

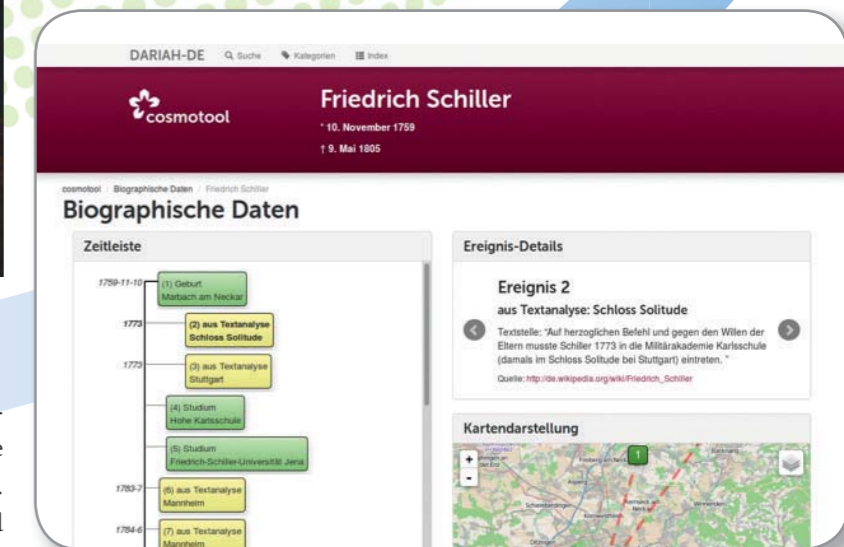


Ableitung biografischer Profile

Durch die Kombination von Daten aus unterschiedlichen Quellen kann oft nicht nur eine größere Anzahl biografischer Ereignisse erkannt werden. Auch die Qualität der individuellen Profile wird gesteigert. Für die Identifizierung und Analyse von Zusammenhängen in natursprachlichen Texten können wir Menschen auf unsere kognitiven Fähigkeiten zurückgreifen. Computerlinguistischen Verfahren fehlen diese Fähigkeiten, wodurch Bezeichnungen und Zusammenhänge oft nicht zweifelsfrei aufgelöst werden können.

Die Kombination und gegenseitige Validierung von Daten aus unterschiedlichen Quellen kann hier helfen. Ein einfaches Beispiel: Die Abbildung zeigt einen Überblick über erkannte Ort-/Zeit-Korrelationen im Lebenslauf Friedrich Schillers. Mit Hilfe der mehrfach erkannten und dadurch relativ sicheren, in der Abbildung grün markierten Information zu seiner Geburt 1759 können weitere Ergebnisse, die im Widerspruch zu diesem Geburtsjahr stehen, von der Aufnahme in das Profil Schillers ausgeschlossen werden. So tauchen aus der Volltextanalyse nur diejenigen Ergebnisse auf, die sich in einer plausiblen Lebensspanne von Schiller befinden, in der Abbildung gelb markiert.

Der von der Bamberger Informatik entwickelte Prototyp erlaubt es, biografische Daten aus unterschiedlichen Quellen zu validieren.



Als Ergebnis bleibt festzuhalten: Aufgrund ihrer uneingeschränkt verfügbaren, oft mehrsprachigen Einträge ist die Wikipedia eine hervorragende Datenbasis für die Erprobung von Analyse- und Visualisierungstechniken, die eine Unterstützung der historischen Forschung bieten könnten.

