

Die computergestützte Erschließung und Visualisierung historischer Itinerare

Blank, Daniel

daniel.blank@uni-bamberg.de
Universität Bamberg, Deutschland

Henrich, Andreas

andreas.henrich@uni-bamberg.de
Universität Bamberg, Deutschland

Einleitung und Ziele

Die Itinerarforschung beschäftigt sich mit der Erschließung historischer Straßennetze. Sie differenziert zwei Itinerararten. Als Itinerare werden einerseits historische Reisewege hochstehender Personen und Herrscher bezeichnet. Diese Reisewege wurden und werden meist anhand historischer Dokumente und Urkunden rekonstruiert, die Auskunft darüber geben, zu welcher Zeit sich gewisse Personen an bestimmten Orten aufgehalten haben. Andererseits bezeichnen Itinerare auch Reisewegverläufe, einzeln oder in Form von Sammlungen, die unmittelbar als solche zusammengetragen wurden (Szabó 2009: 85).

Die Erforschung historischer Itinerare ist ein wichtiger Arbeitsschwerpunkt in verschiedenen Wissenschaftsdisziplinen. Dies ist in Teilen dadurch bedingt, dass historische Personen, die die Itinerare entweder direkt erstellt haben oder auf Basis deren Vita die Itinerare durch Dritte erstellt wurden, häufig in verschiedenen Rollen unterwegs waren. So tritt etwa Hieronymus Münzer auf seiner Spanien- und Frankreichreise gleichzeitig als Arzt, Historiker, Kaufmann, Pilger und Geograph in Erscheinung (Hurtienne 2009: 268). Nicht zuletzt deshalb ist die „Altwegeforschung“ ein stark interdisziplinäres Forschungsfeld (Veling 2014). Charakteristisch für die Itinerarforschung ist eine manuell geprägte und zeitaufwändige Arbeitsweise. Ein wesentlicher Aspekt bei der Erschließung ist etwa die Identifizierung der in den Itineraren genannten Orte (Hurtienne 2009: 269).

Der Ansatz, der in dieser Arbeit beschrieben wird, versucht Werkzeuge zu entwerfen, die Forscher_innen in der Itinerar- und Altwegeforschung in verschiedenen Wissenschaftsbereichen unterstützen. Ziel ist es, die zeitaufwändige, manuelle Erschließung der Itinerare effizienter zu gestalten und später auch den Vergleich verschiedener Itinerare im großen Stile zu

fördern. Außerdem soll es ermöglicht werden, leichter Fehler und Inkonsistenzen in den Itinerarquellen zu identifizieren. Darüber hinaus soll die Erweiterung von Ortsverzeichnissen, insbesondere um historische Informationen, erleichtert werden. Ortsverzeichnisse, sog. *Gazetteers*, sind häufig unvollständig und lückenhaft, insbesondere wenn es um historische Informationen geht. Ferner beschränken sich historische Gazetteers häufig auf bestimmte geografische Gebiete und/oder Zeitperioden. Für die Anreicherung der Gazetteers stellen Itinerare eine wesentliche Datenbasis dar, aus der sich computerunterstützt mit Hilfe des hier skizzierten Ansatzes wichtige Informationen ableiten lassen.

Während Blank und Henrich (2015) bereits die grundlegende Idee und eine Abgrenzung gegenüber verwandten, technischen Arbeiten im Geografischen Information Retrieval adressieren, beleuchtet die vorliegende Arbeit insbesondere die Anwendbarkeit und Einsatzmöglichkeiten des Ansatzes.

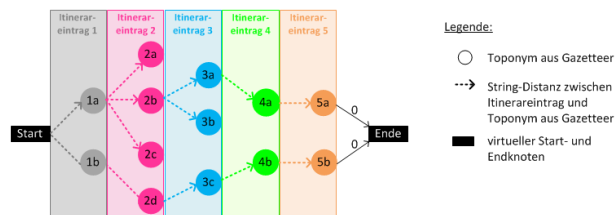
Vorgehen und Systembeschreibung

Die computergestützte Erschließung historischer Itinerare lässt sich in vier Teilschritte zerlegen. Nachdem in Schritt (1) eine optische Zeichenerkennung durchgeführt wird und die Dokumente, die in der Regel als Scans vorliegen, eingelesen werden, muss in Schritt (2) die Struktur der Itinerare erfasst werden. Diese liegen gelegentlich in Tabellenform vor und enthalten beispielsweise wichtige Distanzangaben, die für eine Auflösung von Mehrdeutigkeiten in den Ortsnamen in Schritt (3) ein wesentliches Kriterium sind. Abschließend geht es in Schritt (4) darum, die exakten Wegeverläufe im Gelände zu rekonstruieren.

Diese Arbeit fokussiert auf Schritt (3). Grundlage des Verfahrens sind Itinerare, die neben potentiell mehrdeutigen bzw. in historischer Schreibweise enthaltenen Toponymen auch geografische Distanzen zwischen den einzelnen Wegpunkten erfassen. Distanzen können explizit (s. Spalte *Eingabe* in der folgenden Tabelle; Zahlen entsprechen Meilen) oder implizit zum Beispiel durch eine Angabe von Tagesetappen vorhanden sein.

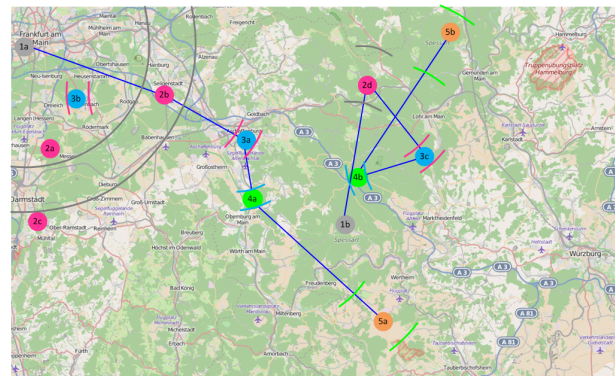
Eingabe	Ergebnis	Korrekte Lösung	Distanz
Frankfurt, -	Frankel	Frankfurt am Main	66,2 km
Aschenburg, 5	Scheringen	Aschaffenburg	57,3 km
Mildtenburg, 4	Kaltenbrunn	Miltenberg	9,1 km
Bischoffshaim, 4	Tauberbischofsheim	<i>Tauberbischofsheim</i>	0 km
Würzburg, 4	Würzburg	<i>Würzburg</i>	0 km
Detelbach, 4	Dettelbach	<i>Dettelbach</i>	0 km
Haßfurt, 5	Westheim bei Haßfurt	Haßfurt	6,5 km
Bamberg, 4	Bischberg	Bamberg	5,1 km
Lichtenfelß, 4	Steinfeld	Lichtenfels	19,9 km
Kulmbach, 4	Kulmbach	<i>Kulmbach</i>	0 km

Das Verfahren generiert einen Entscheidungsgraph mit einem virtuellen Start- und Zielknoten. Ein solcher Graph ist exemplarisch in der folgenden Abbildung für ein fiktives Itinerar mit fünf Wegpunkten skizziert.



Für jedes Toponym des Itinerars wird zunächst die minimale Distanz zu den Toponymen des deutschen Teils des Geonames-Gazetteers (Wick 2005-2016) berechnet. Dabei können Distanzen verwendet werden, die rein syntaktisch Zeichenketten vergleichen; auch phonetische oder semantische Distanzen sind denkbar (alle im Folgenden vereinfachend als String-Distanzen bezeichnet). Auf den so ermittelten, minimalen String-Distanzwert pro Toponym des Itinerars wird ein Delta addiert, um pro Itinerareintrag einen Schwellwert zu erhalten, mit dessen Hilfe eine Kandidatenmenge aus der Menge der Toponyme des Gazetteers identifiziert werden kann. Die String-Distanz aller Toponyme der Kandidatenmenge muss kleiner oder gleich dem Schwellwert sein. Alle Toponyme der Kandidatenmenge werden anschließend nach Anwendung diverser Filter als Knoten in den Entscheidungsgraph aufgenommen. Als Kantengewicht wird die String-Distanz zwischen dem jeweiligen Toponym des Itinerars und dem des Gazetteers erfasst. Dieser Verarbeitungsschritt wird für alle Toponyme des Itinerars wiederholt, sodass ein Graph wie in obiger Abbildung entstehen kann. Abschließend können die gemäß String-Distanz kürzesten Wege vom virtuellen Start- zum Endknoten und damit die Toponyme des Gazetteers mit den in Summe geringsten String-Distanzen zu den Itinerareinträgen ermittelt werden.

Als Filter sind verschiedene Kriterien denkbar. Ein erstes Kriterium ist der bereits genannte Schwellwert der String-Distanz, mit dessen Hilfe Toponyme des Gazetteers gefiltert werden. Ein zweites Kriterium ist die geografische Distanz. Hierzu wird die Entfernungsangabe im Itinerar in einen Kilometerkorridor transformiert. Die Nutzer_innen der Anwendung können die Breite des Korridors vorgeben bzw. diese je nach Kontext festlegen. Der Korridor ist in der folgenden, fiktiven Abbildung durch Sphärenschalen angedeutet.



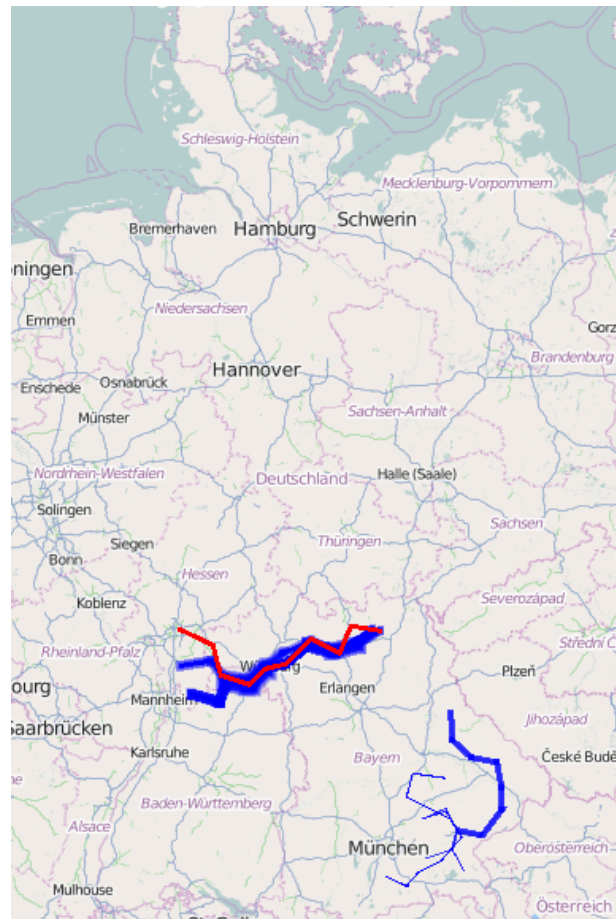
Die Orte 2a und 2c scheiden als nachfolgende Etappenziele von Ort 1a aus, da sie nicht innerhalb der grauen Sphärenschale um Ort 1a liegen. Nur Toponym 2b qualifiziert sich für die weitere Analyse, da es innerhalb liegt. Ein drittes Filter-Kriterium ist die Gerichtetheit der Wegverbindungen. Indem an jedem Zwischenstopp ein Peilungswinkel ermittelt wird, können Orte und Wegverläufe ausgeschlossen werden, die der Gerichtetheit eines Itinerars widersprechen. Somit würde unter Umständen von Ort 2b ausgehend das Toponym 3b verworfen und nur 3a weiter betrachtet, wenn nur hier die geforderte Gerichtetheit gegeben ist. Weitere Kriterien ermöglichen es, ganze Pfade herauszufiltern. Wenn beispielsweise die Vertauschung zweier aufeinanderfolgender oder beliebiger Wegpunkte dazu führt, dass sich die Gesamtdistanz vom Start- zum Zielpunkt des Itinerars reduziert, kann dies ein Indiz für eine inkorrekte Lösung sein, die der Algorithmus verwerfen kann. Ein solches Beispiel ist in der obigen Abbildung durch den Wegverlauf, der bei Toponym 1b beginnt, angedeutet.

Systemverwendung

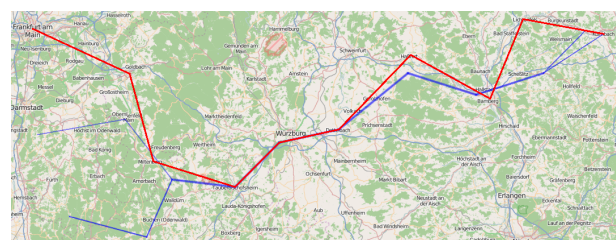
Die Evaluation basiert zunächst auf dem Itinerar Jörg Gails aus dem Jahr 1563 (Krüger 1974). Dies ist „der erste selbständig gedruckte Reiseführer des deutschen Schrifttums“ (Krüger 1974: 1). Erste Analysen basieren dabei auf einem Wegverlauf mit zehn Zwischenstopps aus Route 1 des Itinerars (s. obige Tabelle). Dabei findet die Jaro String-Distanz Anwendung (Winkler 1990). Geografische Distanzfilter werden definiert, indem von einem typischen Umwegfaktor von 1,2 ausgegangen wird. Der Radius der Sphären beträgt 6000 bzw. 8000 Meter, sodass eine mittelalterliche deutsche Meile zwischen 7200 und 9600 Metern modelliert wird (inspiriert durch den Wikipedia-Eintrag 'Meile' , 25.11.2015). Der Winkel für die Gerichtetheit beträgt 90 Grad. Den String-Distanzschwellwert setzten wir auf 0,24. Dieser ist grob per Hand abgestimmt, alle anderen Parameter wurden initial festgelegt und verblieben unverändert.

Die oben dargestellte Tabelle zeigt die durch den Algorithmus ermittelte beste Route. Kursiv dargestellte Toponyme (4/10 = 40%) werden korrekt identifiziert. Im Gazetteer gefundene Toponyme werden aufgetrennt, weil z. B. Frankfurt, der Ausgangspunkt der Route, als *Frankfurt am Main* im Gazetteer enthalten ist. Partielle Übereinstimmungen erhalten keinen Malus. *Frankfurt am Main* soll gleichberechtigt zu *Frankfurt* behandelt werden, da etwa auch *Frankfurt* allein als Toponym im Gazetteer enthalten ist, in Form eines kleinen Dorfs weit von der Route entfernt. Eine Konsequenz dieser Entscheidung ist, dass eine fünfte Übereinstimmung durch *Westheim bei Haßfurt* verhindert wird, das den Vorzug vor *Haßfurt* erhält. Es kann außerdem festgestellt werden, dass einige der nicht korrekt identifizierten Toponyme geografisch sehr nah an den tatsächlichen Orten liegen. Dies ist anhand der Entfernungsangaben in der letzten Spalte der obigen Tabelle ersichtlich. Es zeigt sich auch, dass es schwierig ist, den korrekten Ausgangspunkt der Route zu finden. Nur zwei der zehn Ortsnennungen des Itinerars sind als eindeutiger Eintrag im Gazetteer vorhanden (*Haßfurt* und *Bamberg*). Die besondere Behandlung solch eindeutiger Orte wird in zukünftigen Arbeiten in das Verfahren integriert, um die Qualität weiter zu verbessern.

Die folgende Abbildung zeigt eine Visualisierung der 30 besten Routen, die der Algorithmus findet. Dabei führen Mehrfachnennungen zu dickeren blauen Linien. Rot dargestellt ist die korrekte Lösung. Wegpunkte sind stets durch gerade Linien verbunden. Es lassen sich grob zwei geografische Regionen identifizieren, in denen der Algorithmus die korrekte Lösung vermutet.



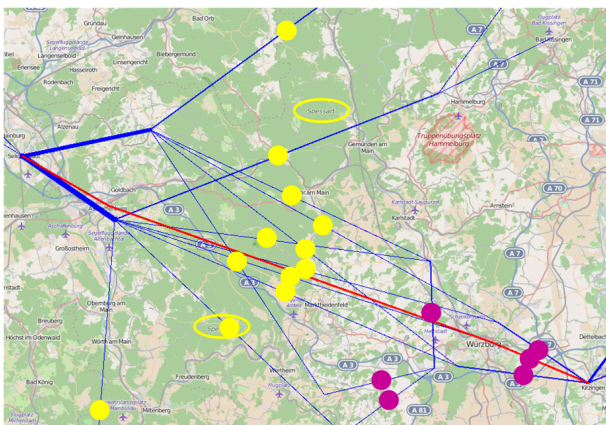
Eine Überlagerung der fünf besten Vorschläge zeigt die nachfolgende Abbildung. Der Fokus liegt nun auf nur einer geografischen Region. Hierbei ist zu erkennen, dass die Vorschläge (blau) nahe an der tatsächlichen Route (rot) liegen.



Exemplarisch soll nun gezeigt werden, wie die Techniken auch im Bereich Visual Analytics und im Speziellen beim Auffinden von Auffälligkeiten und Fehlern verwendet werden können. Hierzu wird ein Ausschnitt einer Route des Brügger Itinerars mit vier Wegpunkten und einer textuellen Anmerkung *Per nemora* analysiert (vgl. Hamy 1908: 170):

Eingabe	korrektes Ergebnis
Zelghenstat, -	Seligenstadt
Asscaffengerhe, III	Aschaffenburg
Witsbuerch, IX	Würzburg
<i>Per nemora</i>	<i>durch den Wald</i>
Litsinghen, III	Kitzingen

Die Anmerkung *Per nemora* ist zwischen Würzburg und Kitzingen angesiedelt. Die hier dargestellte Karte, unter Verwendung der Levenshtein-Distanz (Levenshtein 1966) mit einem Schwellwert von 0,2, gibt Hinweise, dass in direkter Umgebung Würzburgs kein größeres bewaldetes Gebiet zu finden ist (magenta farbige Kreise: Kreismittelpunkt ist jeweils der Mittelpunkt einer Strecke zwischen möglichen Interpretationen von Witsbuerch und Litsinghen).



Eine Positionierung der Anmerkung eine Etappe früher zwischen Asscaffengerhe und Witsbuerch (gelbe Kreise) deckt auf, dass mit der Bezeichnung *Per nemora* wohl die Wegstrecke durch den Spessart gemeint ist. Die Tatsache, dass diese Anmerkung im Itinerar verzeichnet ist, in dem Anmerkungen eher selten zu finden sind, deutet auf eine Besonderheit hin. Müller (1907: 175) schreibt über den entsprechenden Wegabschnitt, dass der Spessart im 15. Jahrhundert vom Verkehr gemieden wurde, obwohl die direkte Verbindung zwischen Nürnberg und Frankfurt durch den Spessart verlief.

Bibliographie

Blank, Daniel / Henrich, Andreas (2015): "Geocoding place names from historic route descriptions", in: *Proceedings of the 9th ACM Workshop on Geographic Information Retrieval, Paris, France*.

Hamy, Ernest-Théodore (1908): *Le Livre de la description des pays de Gilles le Bouvier, dit Berry*. Paris: Ernest Leroux.

Hurtienne, René (2009): "Ein Gelehrter und sein Text: Zur Gesamtedition des Reiseberichts von Dr. Hieronymus Münzer, 1494/95 (Clm 431)", in: Neuhaus, Helmut (ed.): *Erlanger Editionen. Grundlagenforschung durch Quelleneditionen (= Erlanger Studien zur Geschichte 8)*. Erlangen / Jena: Palm & Enke 255-272.

Krüger, Herbert (1974): *Das älteste deutsche Routenhandbuch*. Jörg Gails Raibbüchlin. Graz: Akademische Druck- und Verlagsanstalt.

Levenshtein, Vladimir I. (1966): "Binary codes capable of correcting deletions, insertions, and reversals", in: *Soviet Physics Doklady* 10, 8: 707-710.

Müller, Johannes (1907): "Geleitswesen und Güterverkehr zwischen Nürnberg und Frankfurt im 15. Jahrhundert", in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 5: 173-196, 361-409.

Szabó, Thomas (2009): "Die Itinerarforschung als Methode zur Erschließung des mittelalterlichen Straßennetzes", in: Szabó, Thomas (ed.): *Die Welt der europäischen Straßen. Von der Antike bis in die Frühe Neuzeit*. Köln / Weimar / Wien: Böhlau Verlag 85-96.

Veling, Alexander (2014): "Altwegeforschung: Forschungsstand und Methoden", in: *aventinus - Geschichtswissenschaften im Internet* 44: http://www.aventinus-online.de/no_cache/persistent/artikel/9847/ [letzter Zugriff 13. Oktober 15].

Winkler, William E. (1990): "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage", in: *Proceedings of the Section on Survey Research Methods, American Statistical Association*: 354-359.

Wick, Marc (ed.) (2005-2016): *GeoNames*. Unxos GmbH, Switzerland <http://www.geonames.org/> [letzter Zugriff 30. Januar 2016].

Wikipedia (25.11.2015) „Meile“, in: *Wikipedia*. Die freie Enzyklopädie <https://de.wikipedia.org/wiki/Meile> [letzter Zugriff 30. Januar 2016].