

Aufbau eines crowdsourcingbasierten Verzeichnisses für historische Ortsnamen

Kai Walter und Ralf Bill

Professur für Geodäsie und Geoinformatik, Universität Rostock · kai.walter@uni-rostock.de

Full paper double blind review

Zusammenfassung

Die räumliche Dimension ist – neben der thematischen und der zeitlichen – ein zentrales Konzept der inhaltlichen Beschreibung von Informationsressourcen. In kulturellen und wissenschaftlichen Medien werden räumliche Bezüge in der Regel textuell über Ortsnamen hergestellt. Entsprechend erschlossene Informationsressourcen bieten wichtige Forschungsgrundlage für Wissenschaftsdisziplinen, die räumlich und zeitlich arbeiten. Sie liefern allerdings nur dann einen nachhaltigen Wert, wenn Ortsnamen georeferenziert, sprich mit geographischen Koordinaten eindeutig identifiziert, vorliegen. Insbesondere in community-basierten Online-Kartenforen und -portalen haben sich derartige thematisch/räumliche Recherchemöglichkeiten im Laufe der letzten Jahre hervor getan. Im Rahmen des DFG-Projekts „Virtuelles Kartenforum 2.0“ soll eine umfangreiche Sammlung von Altkarten durch den Aufbau eines crowdsourcingbasierten historischen Ortsnamensverzeichnis via Verknüpfung mit Ortsobjekten räumlich erschlossen werden. Als Grundlage dienen hierfür eine Reihe von teilweise historischen dienstebasierten Ortsnamensverzeichnissen (z. B. GeoNames, Genealogisches Orts-Verzeichnis, OpenStreetMap Nominatim etc.), die sich als einschlägige Grundlage für raumbasierte Recherche in virtuellen Forschungsumgebungen und Kartenportalen etabliert haben. Die Aussagekraft der z. T sehr spezialisierten Verzeichnisdienste soll so kombiniert und gebündelt werden, dass sie Nutzern und Bearbeitern des aufzubauenden historischen Ortsnamensverzeichnis eine optimale und umfangreiche Recherchegrundlage zur Verfügung stellen. Eine entsprechende Zusammenführung geschieht dabei mithilfe von räumlichen Aggregierungsmethoden.

Motivation

Ortsnamen

„Whatever occurs, occurs in space and time“ (M. Wegener, zitiert nach HILL 2006, 1). Die räumliche Dimension ist – neben der thematischen und der zeitlichen – ein zentrales Konzept der inhaltlichen Beschreibung von Informationsressourcen. Um Informationen gezielt recherchieren, verknüpfen und auswerten zu können, ist es notwendig, geographische Bezüge in einer strukturierten Form zu erschließen. In den Medien der kulturellen Überlieferung sowie der Wissenschaft werden räumliche Bezüge in der Regel textuell über Ortsnamen hergestellt. Ortsnamen, spezifischer ausgedrückt Toponyme bzw. Geonamen, bezeichnen Namen topographischer Objekte und können mithilfe geographischer Koordinaten Teilen der realen Welt zugeordnet werden. Entsprechend erfolgt auch die Recherche von

relevanten Dokumenten über eine textbasierte Suche in Titeln, Volltexten, Schlagworten oder Klassifikationssystemen. Der räumliche Bezug ist hierbei relevant und wird auch oft über Orts Schlagworte in Katalogbeständen hergestellt. Derart erschlossene Informationsressourcen bieten wichtige Forschungsgrundlage für Wissenschaftsdisziplinen, die räumlich und zeitlich arbeiten (z. B. Historiker, Geographen, Planer, Ökologen, Statistiker, Demographen usw.). Sie sind jedoch nur dann von nachhaltigem Wert, wenn sie georeferenziert, sprich mit geographischen Koordinaten eindeutig identifiziert, vorliegen

Vor allem im Bereich der Kartenbibliotheken ist ein entsprechender Bedarf offensichtlich, weil die konventionelle Erschließung mit Autor, Titel, Schlagwort wesentliche Einschränkungen aufweist. Insbesondere in communitybasierten Online-Kartenforen und -portalen haben sich hier thematisch/räumliche Recherchemöglichkeiten im Laufe der letzten Jahre hervorgetan (DR 2014, OMO 2014). Insgesamt fehlt es jedoch noch an nutzerfreundlichen integrativen Lösungen, welche die Suche, die Visualisierung und anschließende Einbindung und Nutzung von raumbezogenen Informationsressourcen in beliebige Arbeitsumgebungen von Praktikern und Forschern ermöglichen.

Ortsnamensverzeichnisse

Ortsnamensverzeichnisse bzw. Gazetteers führen im klassischen Sinn Auflistungen von Toponymen bzw. Geonamen in einer bestimmten Region, um diese anhand zusätzlicher raumbezogener Angaben genauer identifizieren bzw. räumlich zuordnen zu können. Aktuelle webbasierte Ortsnamensdienste existieren mit verschiedenen Funktionsumfängen (z. B. Geocoding/Reverse Geocoding via Toponym, Adresse, PLZ etc.) und in z. T. spezifischen Anwendungsbereichen (z. B. Genealogie, Crowdsourcing, amtliche Nachweise etc.). Für den Einsatz im Rahmen von z. B. virtuellen Forschungsumgebungen und Kartenportalen haben sich eine Reihe von dienstebasierten Ortsnamensverzeichnissen (z. B. GeoNames, OpenStreetMap Nominatim etc.) als Grundlage für eine raumbasierte Recherche in den angebotenen Datenbeständen etabliert. Als Basis stehen hierbei kollaborativ gesammelte aktuelle topographische Daten im Vordergrund. Die Arbeit in einem räumlich-historischen Kontext, wie z. B. Arbeit mit Altkarten, stellt die Verwendung solcher Verzeichnisse jedoch vor besondere Herausforderungen. Insbesondere in den östlichen Regionen des ehemaligen Deutschen Reichs ist die Namensgenese von Ortschaften über die letzten 150 Jahre sehr wechselhaft. Eine beispielhafte Suche nach dem vor 1945 gebräuchlichen Namen *Horst* für das heutige polnische *Niechorze* würde keine brauchbaren Treffer ergeben. Für diesen Verwendungszweck existieren zwar spezialisierte Verzeichnisse wie z. B. das sehr umfangreiche Genealogische Ortsverzeichnis (SCHNADT 2011), die jedoch aufgrund ihrer sehr eng abgesteckten Verwendungskontexte, in diesem Fall vor allem die Sammlung von Siedlungsnamen für die Ahnenforschung, nicht unmittelbar für andere Anwendungszwecke verwendbar sind.

DFG-Projekt Virtuelles Kartenforum 2.0

Im DFG-Projekt „Virtuelles Kartenforum 2.0 – Eine Service-orientierte virtuelle Forschungsumgebung in der Deutschen Fotothek“ wird ein integrativer und nutzerfreundlicher Zugang zu der dann georeferenziert vorliegenden Kartensammlung der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) konzipiert und entwickelt. Ziel ist es, die gescannt vorliegenden topographischen Karten in der Gebietsaus-

dehnung des ehemaligen Deutschen Reiches georeferenziert für Raum-Zeit-Forscher über einen standardisierten Dienst, dem vom Open Geospatial Consortium (OGC) spezifizierten Web Map Service (WMS), bereitzustellen¹ (vgl. BILL et al. 2014).

Ein wesentlicher Punkt für eine verbesserte Nutzbar- und Recherchierbarkeit ist neben der Georeferenzierung (BILL & WALTER 2015) vor allem die Verknüpfung der Kartenblätter mit Ortsobjekten. Neben der reinen räumlichen Recherche durch Verschiebung des Kartenbilds nach Länge und Breite, ist die Verwendung von Ortsnamensverzeichnissen bzw. Gazetteer-Diensten in (Web)GIS-Anwendungen ein praktizierter Standard, um Nutzern einen intuitiven Einstiegspunkt in teils überwältigend große Datenbestände via Verortung zu ermöglichen. Eine neue Dimension der Recherche sowohl in Karten- als auch in Medien- und Bibliotheksbeständen soll durch ein flächendeckendes Ortsnamensverzeichnis entstehen.

Ziel des Projekts ist es deshalb, die in den Messtischblättern enthaltenen Ortsnamen (sowohl Toponyme generell als auch speziell Siedlungsnamen) und -lagen über alle vorliegenden Zeitschnitte (1868-1945) zu extrahieren, sie mit anderen bestehenden Verzeichnissen zu verknüpfen und in Form eines historischen dienstebasierten Ortsnamensverzeichnisses Forschern für die zukünftige Arbeit bereitzustellen. Das im Rahmen verschiedener Projekte etablierte Prinzip des Crowdsourcing (siehe z. B. <http://www.bl.uk/maps/>) wurde auch bereits in diesem Projekt erfolgreich für die Georeferenzierung der Messtischblätter angewendet (siehe BILL & WALTER 2015) und kommt auch hier wieder zum Einsatz. Um die freiwilligen und mitunter fachfremden Nutzer sowie zukünftige Forschungsarbeiten bei der Erfassung von historischen Ortsnamen bestmöglich zu unterstützen, liegt eine Bündelung aller verfügbaren und bisher teilweise angesprochenen Datenquellen nahe.

Datengrundlage

Ausgangslage und Anforderungen

Neben einer reinen Suchfunktion über aktuelle und historische Ortsnamen zur ad hoc-Referenzierung und Sprung auf entsprechende Messtischblätter, soll das aufzubauende Ortsnamensverzeichnis vor allem auch die Annotation der im Kartenbild vorhandenen Ortsobjekte durch den Nutzer unterstützen. Um hier den crowdsourcingbasierten Ansatz größtmöglich zu vereinfachen und zu automatisieren, soll noch vor Beginn für die Ausdehnungen des ehemaligen Deutschen Reichs eine möglichst flächendeckende Sammlung von Ortsobjekten innerhalb der Messtischblätter für den Nutzer zur Auswahl zur Verfügung stehen. Das auch immer mit einer gewissen Unsicherheit verbundene händische Setzen von komplett neuen Ortsobjekten soll somit auf ein gewisses Maß reduziert werden. Von der Namensvarianz abgesehen, sind die aus den einzelnen vor allem zeitgenössischen Ortsnamensverzeichnissen gespeisten Lagen der Ortsobjekte mit geringen Abweichungen größtenteils auf die in den Messtischblättern verzeichneten übertragbar, da die meist an historischen Ortskernen, Hauptstraßen, Marktplätzen etc. orientierten Mittelpunkte über die knapp 150-jährige Zeitspanne bis auf geringe Abweichungen einheitlich sind.

¹ <http://www.deutschefotothek.de/cms/kartenforum.xml> (14.01.2015)

Während die Nutzung verschiedener z. T. historischer Ortsnamensverzeichnisse (GOV, HONV MV, siehe LABAHN et al. 2102) für den aufzubauenden Datensatz auf der Hand liegen, sollen auch unterschiedliche aktuelle Datensätze zum Einsatz kommen. Auf den ersten Blick liefern die umfangreichen Datensätze von GeoNames und OpenStreetmap Nominatim, neben vielen Übereinstimmungen, auch eine nicht unerhebliche Menge von Ortslagen, die im jeweils anderen nicht zu finden sind. Insbesondere bei kleineren Ortslagen sind diese Unterschiede deutlich zu beobachten. Eine Kombination erscheint somit auch hier sinnvoll. Auf der anderen Seite ist eine Vereinheitlichung bzw. Zusammenführung der verschiedenen Datensätze mithilfe eines einzigen harmonisierten Datenmodells im Rahmen der Projektarbeiten als nicht praktikabel anzusehen und wird nicht angestrebt. Jedes der Ortsnamensverzeichnisse stellt, für sich im jeweiligen Anwendungskontext gesehen, ein optimiertes Werkzeug dar. Vielmehr soll die geplante Bündelung der einzelnen Datenquellen via Referenzierung mithilfe räumlicher Aggregation umgesetzt werden, wodurch Modularität der externen Datensätze und Autorität der jeweiligen Datenbetreiber gewahrt wird.

Verwendete Ortsnamensverzeichnisse

Die Auswahl der zu verwendenden Ortsnamensverzeichnisse fiel aufgrund von Umfang sowie historischer als auch amtlicher Relevanz auf die folgenden Quellen:

- GeoNames (GN)
<http://www.geonames.org/>
- Genealogisches Ortsverzeichnis (GOV)
<http://gov.genealogy.net/search/index>
- OpenstreetMap (OSM)
<http://www.openstreetmap.org/>
- Bundesamt für Kartographie und Geodäsie – Geographische Namen (BKG)
<http://www.geodatenzentrum.de/geodaten/>
- GeoBeacon-Datensatz inkl. HONV-MV (siehe WALTER et al. 2014)
<http://139.30.132.26/beacon/search.html>

Technische Durchführung

Datenaufbereitung

Die Datenbeschaffung erfolgt meistens über die auf den Projektseiten zur Verfügung stehenden und regelmäßig aktualisierten Austauschdateien in CSV- oder Shape-Format. Da die Daten in unterschiedlicher räumlicher Granularität vorliegen, muss zunächst eine rein räumliche Einschränkung auf die Grenzen des ehemaligen Deutschen Reichs erfolgen. Nach der zuvor noch teilweise notwendigen Extraktion der Punktobjekte (im OSM-Austauschformat der Typ ‚node‘), liegen in einigen Fällen (GeoNames und OSM) noch deutlich kleindetailliertere Objekte vor als eigentlich benötigt werden (Denkmäler, Geschäfte, sonstige POI). Entsprechend werden die genannten Datensätze über entsprechend enthaltene Objektklassen gefiltert. Für OSM gilt beispielsweise eine Einschränkung der Objektklassen ‚city‘ (Stadt) bis zur Größe ‚hamlet‘ (Weiler). Für den GeoNames-Datensatz ist zudem aufgrund der breiteren Klassenbildung eine Ausdünnung der Klasse S (spot, building, farm)

um Objekte, z. B. Hotel, notwendig. Das Vorliegen von gewissen Lokalitäten (z. B. Forsthaus) bleibt, soweit möglich, erwünscht, da sich diese bei der Erschließung von Objekten auf historischen Karten als durchaus hilfreich erweisen können.

Nach der Durchführung der beschriebenen Filterung stellt sich die Anzahl der vorliegenden Ortsobjekte in den einzelnen Datensätzen wie folgt dar:

- GeoNames: 345.331
- Genealogisches Ortsverzeichnis: 283.496
- OpenStreetMap: 218.913
- GeoBeacon: 91.000 (davon 7.135 HONV-MV)
- BKG Geographische Namen: 45.061

Räumliche Aggregation

Um die, durch die Kombination der Datenquellen bedingten, nun vorliegenden Doppelungen bzw. Ähnlichkeiten von Ortsobjekten bereinigen zu können, bedarf es einer räumlichen Zusammenfassung solcher Punkte, ohne dabei den Informationsgehalt der individuellen Objekte zu verlieren (siehe Abb. 1). Hierzu werden die einzelnen Datensätze mithilfe der räumlichen Funktionen eines PostGIS-Datenbanksystems, z. B. via Funktion ‚ST_DWithin‘, anhand ihrer Koordinaten verknüpft. Beginnend mit dem umfangreichsten Datensatz (GeoNames) werden alle Informationen der Objekte der nächstkleineren Datensätze (GOV, OSM, GeoBeacon, BKG) an die Objekte des Größeren angehängt, wenn sie sich innerhalb eines gewissen Umkreises (z. B. 1.000 Meter) um diese herum befinden. Mithilfe von Mengenoperationen kann in einem nächsten Schritt nun überprüft werden, welche Objekte des kleineren Datensatzes (GOV) nicht unter diese Bedingung fallen. Man erhält somit die Objekte, die im größeren Datensatz (GeoNames) nicht enthalten sind (Differenzmenge GOV\GeoNames). Die so ermittelten Differenzobjekte können wiederum wie im ersten Schritt genutzt werden, um die benachbarten Objekte aller übrigen nächstkleineren Datensätze (OSM, GeoBeacon, BKG) ausfindig zu machen, um deren Informationen entsprechend an sie anzuhängen. Werden die beschriebenen Schritte bis hin zum kleinsten Datensatz (BKG) wiederholt, werden so automatisch alle benachbarten Objekte aller Datensätze verknüpft bzw. die allein gestellten Objekte jedes Datensatzes ausfindig gemacht. Insgesamt ergibt sich somit eine Summe von knapp 378.000 individuellen Ortsobjekten in den Ausdehnungen des ehemaligen Deutschen Reiches.



Abb. 1: Zusammenführung von Ortsobjekten verschiedener Quellen (links) durch räumliche Aggregation (rechts) (Screenshot, Kartenmaterial via [openstreetmap.org](https://www.openstreetmap.org))

Im Folgenden wird aufgelistet, wieviel Ortsobjekte jede Datenquelle entsprechend der oben genannten Methodik beiträgt:

- GeoNames: 345.331
- Genealogisches Ortsverzeichnis: 19.587
- OpenStreetMap: 11.695
- GeoBeacon: 1200
- BKG Geographische Namen: 47
- Gesamt → 377.946

Umsetzung als Webclient



Abb. 2: HONV Web-Client (Screenshot, Kartenmaterial via openstreetmap.org)

Der nach Abschnitt 3.2 erzeugte Datensatz resultiert in einer Sammlung von Ortsobjekten, die durch ihre Lage identifiziert sind. Durch die Beibehaltung der Fremd-Identifikatoren aller externer Datenquellen ist es zudem möglich, alle in einem Objekt zusammengeführten Informationen auf ihre jeweilige Quelle zurückzuführen. Das Resultat ist eine Sammlung von URLs bzw. Weblinks mit Verweis auf den jeweiligen Datensatz in der Webpräsenz des Datenanbieters. In einer ersten prototypischen Anwendung im Virtuellen Kartenforum 2.0 wird der Datensatz über einen Web Map Service unter zusätzlicher Nutzung der Schnittstelle ‚GetFeatureInfo‘ bereitgestellt und in einem, auf der Javascript-Bibliothek OpenLayers 3 basierenden, WebGIS visualisiert (siehe <http://139.30.132.26/vk2/honvWMS.html> sowie Abb. 2). Die Erfassung von ergänzenden Informationen durch den Nutzer auf Grundlage der Messtischblätter wird wie geplant unterstützt. So entstehende Annotationen werden im Datenmodell lediglich an die Koordinaten des jeweiligen Ortsobjekts gebunden – nicht an das Objekt selbst – und können so zu einem späteren Zeitpunkt beliebig flexibel mit allen externen Datenquellen über einen räumlichen Abgleich rekombiniert werden.

Stand und Ausblick

Die beschriebenen Methoden und Komponenten bilden die Grundlage für ein Recherche- und Annotationswerkzeug, das den Aufbau eines umfangreichen historischen Ortsnamensverzeichnis unter Verwendung von Altkarten effektiv unterstützt. Hierbei werden vor allem die fachspezifischen Stärken aller verwendeten externen Verzeichnisse erhalten und ermöglichen die Verknüpfungen in ihre jeweiligen spezifischen Informationsangebote und Anwendungsbereiche. Das Wechseln und Hinzufügen weiterer Datenquellen ist unter minimaler Anpassung der Aggregierungsmethodik problemlos umsetzbar. Nutzer des Kartenportals erhalten im Rahmen ihrer Arbeiten eine kompakte und einfach zu verwendende Übersicht von verdichteten Informationsressourcen, die nach Belieben weiterverfolgt werden können. Je nach anwendungsspezifischem Bedarf kann in Zukunft die Verzahnung in spezielle Teile der externen Informationsangebote intensiviert werden. So ist beispielsweise eine detaillierte Verarbeitung und Visualisierung von semantischen Objektabhängigkeiten (Semantic Web, z. B. via GeoNames) oder Siedlungsentwicklungen (via GOV) denkbar.

Danksagung

Die Autoren danken der Deutschen Forschungsgemeinschaft (DFG) für die Förderung des Projektes im Rahmen des Förderprogramms Virtuelle Forschungslaboratorien (Förderkennzeichen BU 2228/14-1 und Bi 467/22-1).

Literatur

- BILL, R. (Hrsg.) (2012), Kulturlandschaftsforschung in einer virtuellen Forschungsumgebung auf Basis von Internet-GIS-Technologien. Wichmann Verlag, Berlin/Offenbach, 186 S.
- BILL, R. & WALTER, K. (2015), Crowdsourcing zur Georeferenzierung alter topographischer Karten – Ansatz, Erfahrungen und Qualitätsanalyse. Zeitschrift für Geodäsie, Geoinformation und Landmanagement (eingereicht, DOI 10.12902/zfv-0060-2015).
- BILL, R., WALTER, K. & MENDT, J. (2014), Virtuelles Kartenforum 2.0 – Verfügbarmachung von Altkarten über eine räumliche Portalanwendung. In: STROBL, J., BLASCHKE, T., GRIESEBNER, G. & ZAGEL, B. (Hrsg.), Angewandte Geoinformatik 2014. Beiträge zum 24. AGIT-Symposium Salzburg. Wichmann Verlag, Berlin/Heidelberg, 684-693.
- DR – David Rumsey Map Collection (2014), <http://www.davidrumsey.com/> (14.01.2015).
- HILL, L. L. (2006), Georeferencing. The Geographic Associations of Information. The MIT Press, Cambridge/London.
- LABAHN, K., STEPHAN, R. & ZEPF, R. (2012), Digitale Bibliothek und Virtuelle Forschungsumgebung – raumbezogene Verknüpfung und Recherche von bibliothekarischen Datenbeständen. In: BILL, R. (Hrsg.), Virtuelle Forschungsumgebung für die Kulturlandschaftsforschung auf Basis von Internet-GIS-Technologien. Wichmann Verlag, Berlin/Offenbach, 59-82.
- OMO – Old Maps Online (2014), <http://www.oldmapsonline.org> (14.01.2015).

- SCHNADT, J. (2011), Das genealogische Ortsverzeichnis (GOV). Vermessung Brandenburg 1/2011, 35-42.
- WALTER, K., LABAHN, K., BILL, R. & ZEPF, R. (2014), Kulturhistorische Forschung und Geodatenportale – raumbasierte Recherche in kulturhistorischen Datenbeständen mithilfe einer generisch erweiterbaren Dienstinfrastruktur. In: STROBL, J., BLASCHKE, T. GRIESEBNER, G. & ZAGEL, B. (Hrsg.), Angewandte Geoinformatik 2014. Beiträge zum 26. AGIT-Symposium Salzburg. Wichmann Verlag, Berlin/Offenbach, 719-728.