

Dies ist eine unbegutachtete Vorversion des Artikels. Das Original finden sie in den Kartographischen Nachrichten.

Wikipedia als Datengrundlage für Sachdaten von thematischen Karten

Tobias Dahinden; Hannover

Die Online-Enzyklopädie Wikipedia beinhaltet eine Vielzahl von Sachdaten. In diesem Artikel werden zwei Wege erläutert, wie diese Daten in der Kartographie genutzt werden können: Man kann die primären Daten der Infoboxen auswerten, dabei besteht natürlich immer die Gefahr von fehlerhaften Daten. Verlässlicher ist es die verschiedenen Seiten miteinander in Beziehung zu setzen und so neue Informationen zu erhalten.

1 Einführung

In den vergangenen 20 Jahren hat das Internet mit dem World Wide Web eine nicht vergleichbare Erfolgsgeschichte erlebt. Während Anfang der Neunzigerjahre vernetzte Computer noch ein Werkzeug für Experten waren, so ist es heute Standard, dass man mit drahtlos-verbundenen mobilen Kleincomputer (Smartphones) jederzeit und an jedem Ort Informationen zur Verfügung stellt und abrufen kann. Fast jede Information ist irgendwo verfügbar.

Früher war es schwierig, an Informationen zu gelangen, weil sie oft gar nicht zur Verfügung standen. Heutzutage besteht die Schwierigkeit im Wiederfinden der Informationen. Allerdings lassen sich für die Informationswiederfindung automatische Prozesse einsetzen. Dies führt zu scheinbar skurrilen Informationsseiten wie *pleaserobme.com*, in denen die Plattformen *Twitter* oder *Foursquare* danach untersucht werden, ob jemand Hinweise darauf gibt, ob er/sie nicht zuhause ist.

Natürlich kann man das Internet auch verwenden, um Informationen für Karten zu gewinnen. Dazu benötigt man einerseits eine Suchmaschine, um Informationen zu finden, andererseits braucht man ein Koordinatenverzeichnis, um bestimmten Namen einen Erdbezug zu geben. Schließlich muss der Erdbezug verifiziert werden (*named entity recognition*), denn mit *Hilton, Paris* ist nicht unbedingt das Hilton in Paris gemeint, sondern möglicherweise die Person Paris Hilton. Und falls ein Ort gemeint ist, ist noch nicht klar, ob es sich um Paris, Frankreich oder Paris, Texas handelt.

Beispiele für diese Art der geographischen Informationsbeschaffung (*geographic information retrieval*) wurden z. B. bei (Jones et al., 2008) demonstriert. Die Probleme liegen neben der Mehrdeutigkeit von Ortsnamen auch in der unstrukturierten Form, in der man die Informationen erhält.

Eine einigermaßen strukturierte Form in der Wissen im Web abgelegt ist, ist die Online-Enzyklopädie Wikipedia. Man findet dort eine Vielzahl von Beschreibungen geographischer Objekte mit Angaben zu Ausdehnung, Wirtschaft, Klima, ... sowie über Personen mit Geburts-, Wirkens- und Sterbeort. Zudem beinhalten viele Artikel zu geographischen Objekte eine Koordinate. Die Koordinate ist eindeutig einem Text zugeordnet, d. h. es gibt keine Mehrdeutigkeiten wie im oben erwähnten Falle von Paris. Im Durchschnitt hat jeder zehnte Wikipediaartikel einen Koordinateneintrag.

Aus diesen beiden Gründen soll die Informationswiederbeschaffung in Wikipedia in Bezug auf die Kartographie im Folgenden etwas genauer dargestellt werden. In Abschnitt zwei werden wir deshalb die Struktur von Wikipedia beleuchten. Im dritten Abschnitt werden Möglichkeiten dargestellt, wie man direkt Daten aus Wikipedia gewinnen kann. Im vierten Abschnitt wird eine Auswahl an Arbeiten vorgestellt, die ihre Daten indirekt aus Wikipedia gewonnen haben. Im letzten Abschnitt ziehen wir ein Fazit.

2 Aufbau von Wikipediaartikeln

In Abbildung 1 ist der Kopf und in Abbildung 2 der Fuß eines Wikipediaartikels abgebildet. Aus den Bildern kann man erkennen, dass Wikipediaartikel neben einem Titel (Name) und einem Textblock eine Vielzahl von Strukturen besitzen. Einige dieser Strukturen sind für alle Artikel gleich, einige passen sich dem Artikelthema an.

2.1 Einzellinks und Vorlagen

Die wohl prominenteste Struktur ist die Verlinkung von den Seiten (in der Abbildung mit 1 bezeichnet). Dabei zeigt der Link entweder auf einen anderen Wikipediaartikel (1) oder auf einen Abschnitt desselben Artikels (1*).

Ein Artikel kann explizit eine Koordinate besitzen (2). Zu finden ist diese jeweils oben rechts auf der Seite. Es handelt sich immer um Punktkoordinaten, auch wenn das beschriebene Objekt linien- oder flächenförmig ist. Als Referenzsystem wird in der Regel WGS84 verwendet, es können aber auch Koordinaten in anderen Systemen, z. B. in Schweizer Landeskoordinaten angegeben werden. Um eine Koordinate anzugeben, müssen im Quelltext bestimmte Zeichen (Klammern, Striche) verwendet werden. Wie das genau auszusehen hat, wird in einer sogenannten *Vorlage* definiert. Diese Vorlage verlangt, dass zusätzlich zur Koordinate noch bestimmte Eigenschaften des Objekts (z. B. der Objekttyp) erfasst werden. Diese werden allerdings nicht direkt auf der Seite angezeigt, sondern dienen als Metadaten.

Unter der Koordinate kann es eine *Infobox* (3) geben. Diese Kästen enthalten strukturierte Informationen in der Form von Schlüssel-Wert-Paaren. Die Struktur ist vom Thema abhängig: die Vorlage für die Infobox für Flüsse unterscheidet sich z. B. von jener für Ortschaften. Insgesamt gibt es über 400 verschiedene Vorlagen für geographische Objekte. Dabei handelt es sich teilweise um allgemeine Vorlagen (*Infobox See*, *Infobox Gebirgsgruppe*), teilweise um sehr spezielle (*Infobox Stadtteil in Stuttgart*, *Infobox Whisky-Destillerie*).

Neben der expliziten Koordinate können im Artikel noch weitere Koordinaten erwähnt werden. Etwa 200 Vorlagen zu Infoboxen verlangen Angaben zur geographischen Lage (4). Dabei kann es sich um eine oder mehrere Punktkoordinaten handeln. Zudem können Koordinaten im Fließtext des Artikels erwähnt werden (4*).

2.2 Speziallinks und Linksammlungen

Linkerhand des Haupttextes gibt es eine Linksammlung (5), die auf die korrespondierenden Artikel in anderen Wikipedia-Sprachversionen verweist. Diese Links heißen Interwikilinks. Der Name stammt daher, dass jede Sprachversion als eigene Wikipedia verstanden wird. Die Links stellen Verbindungen zwischen verschiedenen Wikipedien her.

Viele Links, die um den Haupttext angeordnet sind, verweisen auf Spezialseiten (6). Diese Seiten dienen zur strukturierten Haltung von Metadaten. Es handelt sich um Seiten mit der Möglichkeit den Artikel zu bearbeiten, über ihn zu diskutieren oder man kann sich Versionsgeschichte des Artikels anschauen. Es gibt spezielle Seiten, auf denen für einen bestimmten Artikel eine Auszeichnung z. B. als Lesenswert oder die Löschung beantragt werden kann. Weiter gibt es Seiten in denen die Autoren irgendwelche persönlichen Informationen ablegen können oder Seiten, auf denen steht, wie man Wikipedia bearbeitet.

Erwähnenswert erscheint zudem die Möglichkeit Artikel in Portalen (7) oder Kategorien zu ordnen. Um Kategorien zu bilden, werden unten an der Seite Kategorienlinks (8) angefügt. Die Links führen auf eine Seite, in der alle Artikel einer bestimmten Kategorie verlinkt sind. Kategorien können ebenfalls in Kategorien geordnet werden.

Eine zusätzliche Möglichkeit, um Artikel zu ordnen, bieten die sogenannten Naviblocks (9). Bei diesen Kästen handelt es sich ebenfalls um Vorlagen. Die Blöcke erscheinen jeweils unter allen Artikeln, die in einem Naviblock erwähnt werden. In den Naviblocks werden Artikel mit einem verwandten Thema zusammengefasst, beispielsweise alle Stadtteile einer Stadt oder alle Artikel zu den Prinzessinnen eines bestimmten Herrschaftshauses.

Wikipediaartikel können Medien (Fotos, Videos, Bilder, Tonaufnahmen, Karten, ...) beinhalten. Diese Medien sind in der Regel nur mit dem Artikel verlinkt und werden für alle Sprachversionen

gemeinsam auf dem *Commons-Server* gespeichert. Die Medien können in ihren Metadaten Koordinaten enthalten, dies kann etwa die Aufnahmeposition eines Fotos sein.

3 Primärdaten

3.1 Datengewinnung

Für Menschen ist die natürlichste Art Daten aus Wikipedia zu gewinnen, einzelne Artikel durchzulesen. Wenn es darum geht, größere Informationsbestände zu analysieren, sollte das allerdings automatisch ablaufen. Die Verwaltungssoftware von Wikipedia (MediaWiki) unterstützt automatische Zugriffe mit einem Programmierinterface (API). Zudem kann man den Inhalt von Wikipedia als Textdatei oder Datenbank-Speicherauszug (Dump) beziehen. Weiter gibt es spezielle Datenbanken in denen bestimmte Daten bereits vorverarbeitet zur Verfügung stehen: Besonders erwähnenswert sind die Datenbank *Wikipedia-World* (Kühn und Alder, 2009) und die Datenbank *Vorlagenauswertung* (Alder, 2009).

Die Verwendung der API erfolgt über <http://de.wikipedia.org/w/api.php>. Beim Aufruf dieses Interfaces im Webbrowser erhält man auch Hinweise darauf, wie man das Interface verwenden kann. Über die API kann man einzelne Eigenschaften von einer bestimmten Seite abfragen. Zum Beispiel ist es möglich, sämtliche Links von oder zu einer Seite in Listenform zu erhalten. Man kann abfragen, ob eine bestimmte Seite ein Artikel, eine Kategorie, eine Benutzer- oder eine Diskussionsseite ist. Man kann abfragen, welche Bilder auf einer Seite verwendet werden oder welche Benutzer die Seite bearbeitet haben. Die Seite kann zudem über die API bearbeitet werden.

Von Wikipedia werden regelmäßig Sicherungen erstellt. Diese kann man frei beziehen (Wikimedia_Foundation, 2010). Es werden regelmäßig die gesamte Wikipedia mit oder ohne Versionsgeschichte, sowie verschiedene Teil (Kategorien, Zusammenfassungen, Linkstruktur, Statistiken, ...) gesichert. Zur Auswertung dieser Sicherungsdateien programmiert man vorteilhaft einen speziellen Textparser.

Die Datenbank *Wikipedia-World* umfasst eine Liste mit sämtlichen Koordinaten sämtlicher Wikipedien. Zu jeder Koordinate gibt es den Hinweis, wo sie gefunden wurde. Zudem beinhaltet sie die Information, in welchen Wikipedien es eine Seite gibt, die dasselbe Objekt behandelt und wie diese Seiten heißen. Diese Datenbank eignet sich hervorragend zur Auswertung von geographischen Eigenschaften.

In der Datenbank *Vorlagenauswertung* sind die Informationen aus den Infoboxen und weiteren Elementen, die auf Vorlagen beruhen, gespeichert. Dabei wurde der Typ der Vorlage, der Name des Artikels, in dem die Vorlage gefunden wurde, sowie sämtliche Schlüssel-Wert-Paare dieser Vorlage erfasst. In dieser Datenbank kann man beispielsweise nach sämtlichen Artikeln suchen, welche die Vorlage *Infobox See* verwenden und sich den Namen des Artikels und Wert für einen Schlüssel, z. B. Volumen ausgeben lassen. Mit der *Wikipedia-World*-Datenbank zusammen, lässt sich dann dieser Wert geographisch verorten.

3.2 Verlässlichkeit

Wikipediaartikel können von jeder Person mit Internetzugang bearbeitet werden. Die Bearbeitung erfolgt oft, indem sich ein Benutzer anmeldet und unter einem Spitznamen Änderungen vornimmt. Allerdings ist diese Anmeldung nicht zwingend, die Kennung der Autorin erfolgt dann lediglich über die IP-Adresse. Daher stellt sich die Frage nach der Verlässlichkeit der Daten für Wikipedia in besonderem Maße.

Die Administratoren von Wikipedia sind natürlich bemüht, möglichst genaue Informationen in den Artikeln zur Verfügung zu stellen. Wo es möglich ist, wird daher auf offizielle Daten zurückgegriffen. Ein Beispiel dazu sind etwa die Einwohnerzahlen von den Gemeinden in Deutschland: diese werden direkt von den statistischen Ämtern bezogen. In den Artikel bindet man lediglich einen Verweis auf eine Metadatenvorlage ein. Die Zahlen werden dann automatisch in den Artikel eingefügt. Solche statistischen Daten sind allerdings für den Kartenhersteller von untergeordnetem Interesse, da diese bereits bisher von den Ämtern zur Verfügung gestellt wurden. Einzig die Vereinheitlichung des Quellenzugriffs könnte von Interesse sein.

Von Interesse sind allerdings jene Daten, die sonst nirgends zentral zur Verfügung gestellt werden. Dies können Unternehmensinformationen oder Transportinformationen sein. Bei diesen Informationen muss man bedacht sein, dass sie falsch erfasst oder falsch geschrieben sein könnten und dass ihre Genauigkeit und ihr Erfassungszeitpunkt unterschiedlich sein kann. Ein großes Problem liegt darin, dass in Wikipedia sämtliche Daten in Textform erfasst werden. Für Infoboxen gibt es ein Werkzeug namens *Vorlagen-Meister* (Wikipedia, 2011), das hilft, die Vorlage richtig auszufüllen. Dieses sieht zwar vor, dass es Werte geben kann, die aus einer Liste ausgewählt werden müssen. Dass ein Wert ein bestimmtes Zahlenformat haben muss, ist jedoch nicht vorgesehen.

Eine Möglichkeit um abzuschätzen, ob ein Artikel zuverlässige Informationen enthält, kann über die Analyse seiner Metadaten (Versionsgeschichte, Umfang, Länge der Diskussion) erfolgen. Diese Analysestrategie verfolgt etwa das Projekt *Wikibu* (Hielscher und Stöcklin, 2010). Insbesondere werden die Zahl der Autoren, der Änderungen, der Quellennachweise und der Besucher berücksichtigt und die zu erwartende Qualität als Wert von 0 bis 10 ausgegeben.

3.3 Anwendungsbeispiel

(Bittner, 2010) untersuchte die Möglichkeit, anhand der Primärdaten von Wikipedia eine thematische Karte herzustellen. Sie erstellte eine Karte mit der Anzahl Flugpassagiere und der Kapazitäten von den zivilen, europäischen Flughäfen. Abbildung 3 zeigt eine Karte, die auf der Grundlage dieser Arbeit hergestellt wurde.

Betreffend Vollständigkeit der Daten konnte folgendes festgestellt werden werden:

- Insgesamt werden 575 Verkehrsflughäfen beschrieben. Zum Vergleich: der kommerzielle Flugreiseanbieter *www.fairliners.com* zählt „nur“ 310 Flughäfen auf, wobei alle bei Wikipedia beschrieben werden.
- In der Kategorie Flughafen werden 1530 Flughäfen aufgeführt, bei neun dieser Flughäfen (0,58%) fehlt die Infobox Flughafen.
- Passagierzahlen wurden für 573 Flughäfen ausgewiesen, die Kapazität lediglich für 160 (27,8%).

Betreffend Genauigkeit zeigt sich:

- Lage: Für die zwanzig größten europäischen Flughäfen wurde die Lagekoordinate überprüft. Es wurde keine Abweichung festgestellt.
- Zeit: Bei 35 Flughäfen wurde die Kapazität aufs Jahr 2005, 215 auf 2006, 134 auf 2007, 167 auf 2008 und 2 auf 2009 bezogen angegeben. Bei den restlichen 56 fehlte eine Angabe zum Jahr.
- Numerisch: Bei 68 Flughäfen wurden die Angaben auf Millionen gerundet, bei 16 auf 100.000 bei 16 auf 10.000, bei 30 auf 1000. Die restlichen 423 sind scheinbar Passagiergenau.

Betreffend automatische Auswertung gab es folgende Probleme:

- Für 2,8% Flughäfen konnten die Koordinaten nur über einen Umweg automatisch ermittelt werden. Dies ist in der Verwendung von Sonderzeichen begründet.
- Für den Kartenausschnitt waren 155 Flughäfen relevant. In drei Fällen (1,9%) wurde in den Infoboxen ein Punkt statt einem Komma zur Dezimalzahlendarstellung verwendet.
- Bei einem Flughafen (Franz-Josef-Strauß München) wurden die einzelnen Terminals separat ausgewiesen.
- Bei einem Flughafen wurde die Kapazität in Anzahl pro Stunde anstatt Anzahl pro Jahr angegeben. Bei einem haben die Angaben Wochenbezug.
- Für die Angabe von Millionen wurde 32-mal „Mio.“ und 36-mal „Millionen“ verwendet.
- In 13 Fällen gibt es noch Angaben wie „ungefähr“, „etwa“ oder „ca.“
- Als 1000-er Trennzeichen werden Punkte, Hochkommas oder Leerzeichen verwendet.

Darauf aufbauend kann man folgern, dass eine Darstellung der Passagierzahlen von Flughäfen nahezu vollständig sein dürfte. Besondere Probleme bereitet aber die zeitliche Genauigkeit, die

sich auf 5 Jahre erstreckt. Zudem dürfte es immer notwendig sein, die Daten zu sichten – von einer reinen automatischen Auswertung ist eher abzuraten.

4 Abgeleitete Daten

4.1 Graphstrukturen

Die Strukturen von Wikipedia kann man als Graphen interpretieren (Hecht und Raubal, 2008). Diese lassen sich dann analysieren und daraus Sachinformationen ableiten.

Ein Graph ist eine Menge $G(v,e)$ mit Knoten v und Kanten e . Dabei stellen die Kanten Verbindungen zwischen Knoten dar. Wenn die Kanten in eine bestimmte Richtung zeigen (Einbahnstraßen), spricht man von einem gerichteten Graphen. Eine Kante kann zudem ein Gewicht haben, z. B. die Distanz zwischen zwei Knoten.

Die Artikel von Wikipedia kann man als Knoten, die Links zwischen den Artikeln als Kanten interpretieren. In diesem Fall spricht man vom Wiki-Artikel-Graph (WAG). Ebenso kann man die Kategorien Knoten und die Kategorielinks als Kanten auffassen. Daraus ergibt sich der Wiki-Kategorien-Graph. Man kann sich leicht vorstellen, dass dieses Konzept auf alle anderen Strukturen (Diskussionsseiten, Benutzersseiten, Medien, etc.) erweiterbar ist. Von besonderem Interesse in diesem Artikel sind der Interwikilink-Graph, der die Artikel als Knoten und die Interwikilinks als Kanten verwendet, sowie der Wiki-Text-Graph, der einzelne Texte als Knoten und Seitenlinks als Kanten verwendet.

4.2 Beispiele und Verlässlichkeit

Im Folgenden werden ein paar Untersuchungen aus dem Bereich der Geographie und Kartographie zur Datengewinnung aus Wikipedia erläutert. Wie verlässlich die Resultate solcher Analysen sind, lässt sich oft nur anhand von Beispielen zeigen.

Eine Anwendung des WAG ist es, denjenigen Artikel zu finden, von dem man mit durchschnittlich am wenigsten Klicks auf jede beliebige Wikipediaseite kommt. (Dolan, 2008) bestimmte dies für die englischsprachige Wikipedia: Es zeigte sich, dass der Artikel zu *United Kingdom* der zentralste Artikel ist. Von dort aus gelangt man mit durchschnittlich 3,6 Klicks zu jeder anderen Seite.

(Hecht und Raubal, 2008) verwendeten den WAG, um beliebige Objekte (z. B. *Fidel Castro*) zu verorten. Dabei gewichteten sie die Links anhand der Anzahl Links in dem Artikel, von dem sie ausgehen und der Anzahl Links des Artikels, auf den sie zeigen. Wenn ein Artikel keine Koordinate enthält, werden alle verlinkten Artikel untersucht, ob sie Koordinaten enthalten. Werden mehrere Koordinaten gefunden, werden diese anhand des Gewichtes der Links gemittelt.

Basierend auf dieser Idee werden in (Dahinden, 2009) geographische Objekte lokalisiert. Dabei wurde allerdings auf eine Gewichtung der Links verzichtet, dafür eine Glättung mittels Dichteberechnung durchgeführt. Es zeigt sich, dass die Verteilung der verlinkten Koordinaten gut mit der Ausdehnung der Objekte übereinstimmt. Beachtet man zusätzlich die Nachbarschaftsstrukturen (Cliques) des Graphen, kann man feststellen, welche Objekte horizontal benachbart sind. Dadurch kann man die Ausdehnung von mehreren flächenhaften Objekten gleichzeitig bestimmen. Als Beispiel wurde die Ausdehnung der Kantone der Schweiz anhand dieser Methode bestimmt. In [Abbildung 4](#) ist ein Resultat zu sehen. Die Fläche ist zu 78% korrekt. Allerdings erscheinen die Grenzen deutlich abgerundet.

(Hecht und Moxley, 2009) überprüften das erste Gesetz der Geographie von Tobler („Nahe Objekte sind stärker miteinander in Bezug als entfernte“) anhand verschiedener Wikipedien. Das besagte Gesetz ließ sich nicht falsifizieren. Es zeigte sich aber, dass entfernte Objekte sogar deutlich *nicht* miteinander in Bezug stehen. Damit ist gemeint, dass man gemäß der Wahrscheinlichkeitsrechnung mehr Bezüge erwarten würde, als es tatsächliche gibt.

(Paelke et al., 2010) untersuchten die Texte von Wikipediaartikeln, deren Koordinaten in ein bestimmtes Gebiet verweisen. Es wurde eine Statistik der Wörter erstellt und die häufigsten Stichworte in einer Schlagwortwolke dargestellt. Stichprobenweise wurden die Schlagworte mit den Namen der auffälligsten Landmarken dieses Gebietes verglichen. Ein Beispiel dazu ist in

Abbildung 5 zu sehen. Man kann tatsächlich eine Korrelation der Begriffe erkennen. Ziel dieser Untersuchung war es, übergeordnete räumliche Informationen für ein Gebiet zu finden.

In (Dahinden, 2011) werden Sprachräume anhand der Verwendung von Interwikilinks gezeigt. Dazu wird für jede Sprache die jeweilige Wikipedia analysiert. Wenn ein Artikel eine Koordinate besitzt, wird dies als Indiz gewertet, dass dort die Sprache gesprochen wird. Es gibt allerdings berühmte Objekte, wie etwa das Brandenburger Tor, die in vielen Wikipedien erwähnt werden. Der Einfluss dieser Objekte wurde dadurch reduziert, indem jede Koordinate Invers zur Anzahl der Interwikilinks gewichtet wird. Danach wird eine Dichteschätzung durchgeführt. In Abbildung 6 ist das Resultat dieser Schätzungen für Limburgisch (blau) und Alemannisch (gelb) zu sehen. Je intensiver die Farbe, desto wahrscheinlicher wird die Sprache gesprochen. Zum Vergleich ist eine Sprachkarte hinterlegt.

5 Fazit

Wikipedia ist ein leistungsfähiges Werkzeug, in dem viele nützliche Daten zur Verfügung gestellt werden. Auf die Daten der Artikel kann automatisch zugegriffen werden. Eine automatische Produktion von Thematischen Karten funktioniert allerdings nur theoretisch. Praktisch wird es immer notwendig sein, die Daten im Detail zu sichten und die Konsistenz sowie ihre Formatierung zu prüfen.

Von besonderem Interesse ist es, Wikipedia selbst zu analysieren. In dem Daten aus den Strukturen abgeleitet werden, können mannigfaltige Informationen erzeugt werden. Hier besteht die Schwierigkeit jedoch darin, sich nützliche Methoden und Fragestellungen zu überlegen.

Eine systematische Durchforschung der Wikipediastrukturen nach sinnvollen Informationen gibt es noch nicht. Bisher wurden vor allem der Wiki-Artikel-Graph und der Wiki-Text-Graph analysiert. In kleinerem Maße auch der Interwikilink-Graph. Wikipedia hat allerdings noch deutlich mehr Strukturen, die man nutzen kann. Insbesondere die Medien, die ebenfalls Koordinaten besitzen, bieten ein großes Potential, um geographische Eigenschaften zu erforschen.

Dank

Der Autor dankt Jens Alder und Stefan Kühn für das Bereitstellen der Wikipedia-Koordinaten und für die Unterstützung für den Zugriff auf die Vorlagendatenbank. Ein weiterer Dank geht an Frauke Bittner für die Aufbereitung von den Flughafendaten aus der Vorlagendatenbank.

Literatur:

- Alder, T. WikiProjekt Vorlagenauswertung 2009 [cited 4 May 2009. Available from http://de.wikipedia.org/wiki/Wikipedia:WikiProjekt_Vorlagenauswertung.
- Bittner, F. 2010. Thematische Karten auf Basis von Wikipedia: Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover.
- Dahinden, T. 2009. Localization of uncertain and fuzzy-bordered areas by geocoded articles of a knowledge repository. Proceedings of the International Cartographic Conference, at Santiago de Chile.
- Dahinden, T. 2011. Estimation of the Locations of the Language-Versions of Wikipedia - a Case Study on Geographic Data Mining. In Advances in Cartography and GIScience. Selection from ICC 2011. Paris: Springer.
- Dolan, S. Six degrees of Wikipedia 2008 [cited 29 May 2008. Available from <http://www.netsoc.tcd.ie/~mu/wiki/>.
- Hecht, B. und Moxley, E. 2009. Terabytes of Tobler: Evaluating the First Law in a Massive, Domain-Neutral Representation of World Knowledge. Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009, Proceedings 5756:88-105.
- Hecht, B. und Raubal, M. 2008. GeoSR: Geographically Explore Semantic Relations in World Knowledge. AGILE Conf., 95-113.

- Hielscher, M. und Stöcklin, N. Wikibu. Zentrum für Bildungsinformatik, PHBern 2010 [cited 8. 10. 2010. Available from www.wikibu.ch.
- Jones, C., Purves, R., Clough, P.D. und Joho, H. 2008. Modelling Vague Places with Knowledge from the Web. International Journal of Geographical Information Science 22 (10):1045-1065.
- Kühn, S. und Alder, T. WP : GEO / Wikipedia-World. Wikipedia 2009 [cited 31 August 2009. Available from http://de.wikipedia.org/wiki/Wikipedia:WikiProjekt_Georeferenzierung/Wikipedia-World.
- Paelke, V., Dahinden, T., Eggert, D. und Mondzech, J. 2010. Location Based Context Awareness Through Tag-Cloud Visualization. Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science, at Hong Kong.
- Wikimedia_Foundation. Wikimedia downloads 2010 [cited 11 June 2010. Available from <http://download.wikimedia.org/>.
- Wikipedia. Vorlagen-Meister 2011 [cited 11 January 2011. Available from <http://de.wikipedia.org/wiki/Wikipedia:Helferlein/Vorlagen-Meister>.

Über den Verfasser: Dr.-Ing. Tobias *Dahinden* ist Wissenschaftlicher Mitarbeiter am Institut für Kartographie und Geoinformatik der Leibniz Universität Hannover. E-Mail: Tobias.Dahinden@ikg.uni-hannover.de ■

Abbildungsunterschriften:

Abb. 1: Kopf einer Wikipediaseite. Hervorgehoben: 1 interne und externe Links, 2 explizite Koordinate, 3 Infobox, 4 implizite Koordinaten, 5 Interwikilinks, 6 Links zu Wartungsseiten

Abb. 2: Fuss einer Wikipediaseite. Hervorgehoben: 7 Kategorien, 8 Naviblock

Abb. 3: Passagieraufkommen und Kapazität der Flughäfen in Europa gemäß der Deutschen Wikipedia

Abb. 4: Kantone der Schweiz sowie Kantonsgrenzen gemäß der Verteilung von Koordinaten verlinkter Artikel

Abb. 5: Schlagwortwolke der Wikipediaartikel rund um den Königswortherplatz (links) und im Vergleich dazu die Landmarken im selben Gebiet (rechts)

Abb. 6: Ausdehnung von Alemannisch (gelb) und Limburgisch (blau) gemäß der Analyse der Interwikilinks. Im Hintergrund liegt die Karte der Deutschen Mundarten von Brockhaus, 1894

The image shows a screenshot of the German Wikipedia article for "Leine (Aller)". The page is annotated with green circles and numbers 1 through 7. Annotations include:

- 1: Thuringen und Niedersachsen (Bundesländer)
- 2: Leine (Bund)
- 3: Leine (Bund)
- 4: Leine (Bund)
- 5: Leine (Bund)
- 6: Leine (Bund)
- 7: Leine (Bund)

 The article text includes:

- Die Leine ist ein 201 km langer, linker Nebenfluss des südlicher Nebenflusses der Aller.
- Das Einzugsgebiet der Leine von 6.532 km² umfasst 42 % des Flussgebiets der Aller. Der Fluss ist Namensgeber des Landesgebiets, das einen wesentlichen Teil des Niedersächsischen Berglands darstellt.
- Die Leine entspringt im thüringischen Eichsfeld in Leinefeld, der südwestlichen der beiden Karsthäuser von Leinefeld-Walden.
- Die Anzahl ihrer Quellen unterscheidet sich je nach Literatur und wird auch nach Ergänzbarkeit, denn es werden später bis heute ergossenen, schon nach etwa 100 m Profiltiefe kommt die Leine fluss abwärts (geringer) Zufluss Line auf...

niedrige Auflösung – verkleinert

1

The image shows a screenshot of the German Wikipedia article for "Leine (Aller)" with a focus on the "Bund" section. Annotations include:

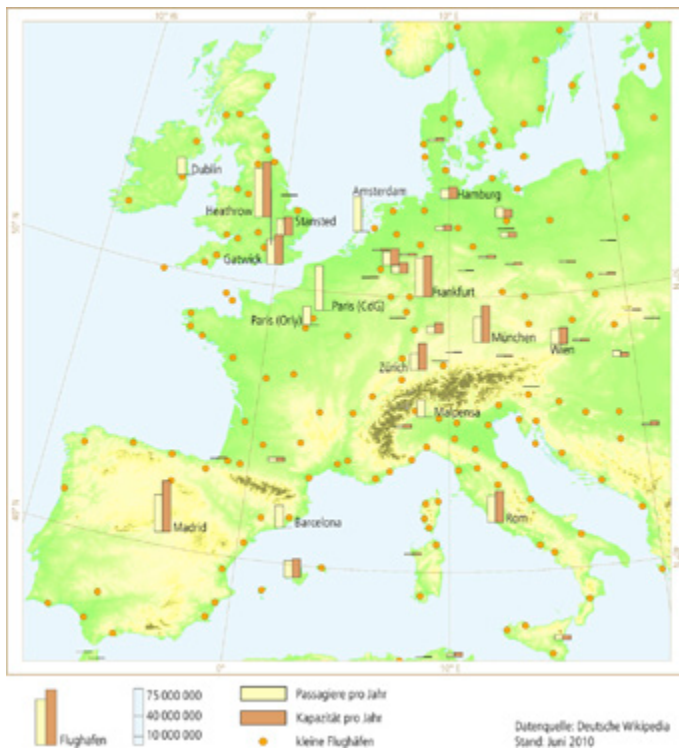
- 8: Kategorien: Flusssystem Leine | Fluss in Europa | Fluss in Niedersachsen | Fluss in Thuringen | Bundeswasserstraße
- 9: Bund

 The article text includes:

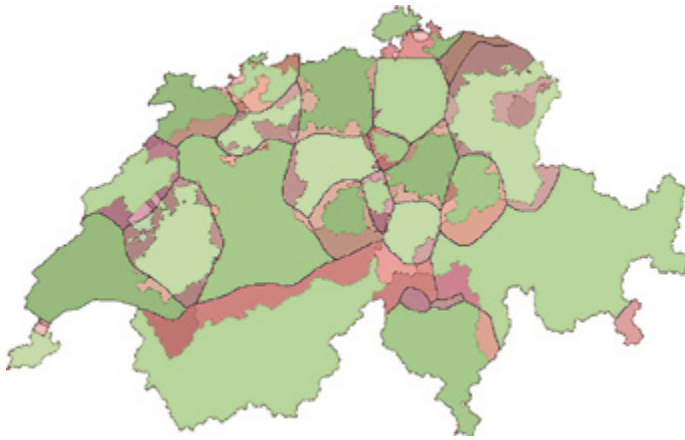
- Die Leine entspringt im thüringischen Eichsfeld in Leinefeld, der südwestlichen der beiden Karsthäuser von Leinefeld-Walden.
- Die Anzahl ihrer Quellen unterscheidet sich je nach Literatur und wird auch nach Ergänzbarkeit, denn es werden später bis heute ergossenen, schon nach etwa 100 m Profiltiefe kommt die Leine fluss abwärts (geringer) Zufluss Line auf...

niedrige Auflösung – verkleinert

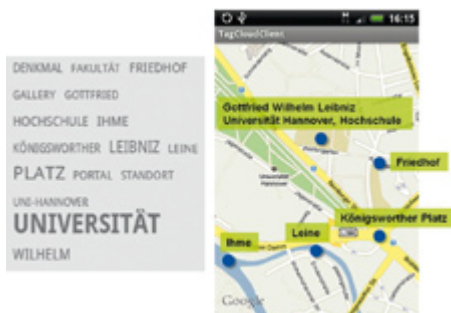
2



KN 4/2011
3474
10

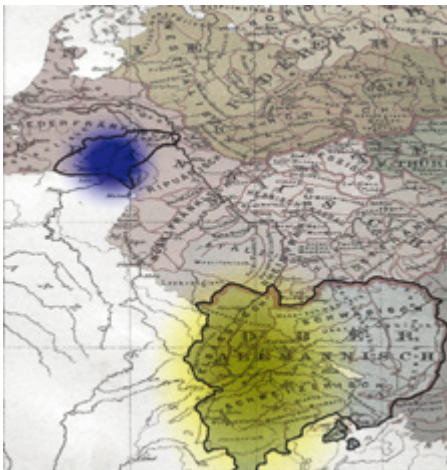


4



niedrige
Auflösung –
verkleinert

5



6