

Where the words are: a visual interactive exploration of plants names

Therón, Roberto

theron@usal.es
Universidad de Salamanca, Spanien

Dorn, Amelie

amelie.dorn@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Seltmann, Melanie

melanie.seltmann@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Benito, Alejandro

abenito@usal.es
Universidad de Salamanca, Spanien

Wandl-Vogt, Eveline

eveline.wandl-vogt@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Gabriel Losada Gómez, Antonio

alosada@usal.es
Universidad de Salamanca, Spanien

Wo die Wörter sind: eine visuell-interaktive Erforschung von Pflanzennamen

In den Digital Humanities werden häufig Visualisierungsmethoden eingesetzt, um bestimmte Trends, Beziehungen oder Inhalte innerhalb oder zwischen verschiedenen Datensätzen hervorzuheben. Oft werden gut etablierte und weit verbreitete Arten graphischer Darstellung von Daten herangezogen, wie Verbert (2015) gezeigt hat. Der Einsatz innovativer Visualisierungsmethoden für die Datenerforschung und den Datenzugriff ist jedoch bei Humanities-Projekten, die sich mit nicht-numerischen Daten beschäftigen, noch relativ selten. In diesem Beitrag stellen wir ein Visualisierungstool vor, das im Rahmen des DH-Projekts *exploreAT! – exploring Austria's culture through the language glass* entwickelt wird, und erläutern dessen Anwendung am Beispiel der Pflanzennamen-Sammlung für das Wörterbuch der bairischen Dialekte in Österreich.

exploreAT! (vgl. Wandl-Vogt et al, 2015) bietet unterschiedliche Einblicke in die vielfältige Beschaffenheit der deutschen Sprache in Österreich, durch exploratives Erforschen mittels einer Synthese von digitalen Infrastrukturen, Lexikographie, visueller Analyse und

Citizen Science. Das Projekt basiert auf einer Sammlung von Daten zu den bairischen Dialekten in Österreich aus dem frühen 20. Jahrhundert aus der Region der ehemaligen österreichisch-ungarischen Monarchie. Die Datenerhebung erfolgte ursprünglich mittels Fragebögen, die eine Vielzahl von Themen aus dem Alltag abdecken. Die gesammelten Daten bestehen aus rund 200.000 Stichwörtern in geschätzten 4 Millionen Datensätzen. Teile davon wurden als fünfbändiges Wörterbuch mit etwa 50.000 Stichwörtern (WBÖ), und Teile als Datenbank (DBÖ) ausgegeben. Innerhalb des Projekts gibt es vier spezifische, aber miteinander verbundene Arbeitsbereiche: kulturelle Lexikographie, semantisch-technologieorientierte Forschungsinfrastrukturen, visuelle Analyse und Bürgerwissenschaften. Des Weiteren werden use-cases für spezifische Themen wie Pflanzennamen, Farben oder Lebensmitteln entwickelt. TEI / XML Schnittstellen werden eingesetzt, um die Organisation von Metadaten, Konzepten und linguistischen Daten zu verbessern. Darüber hinaus ist vorgesehen, weitere Zugangspunkte zur Arbeit mit LOD zu schaffen, ontologische Ressourcen zu nutzen und damit die Visualisierung von konzeptionellen und semantischen Informationen zu gewährleisten.

Mit Hilfe des vorgestellten visuellen Analysetools werden weitere Einblicke in die komplexe Struktur dieser Dialektdaten gegeben, wobei ein intuitiver und leicht zugänglicher Ansatz vorgesehen ist. In diesem Beitrag nehmen wir Pflanzennamen als exemplarischen Fall für die visuelle Exploration, Analyse und Darstellung von Datenstrukturen.

Der Prototyp dieses Tools basiert auf einer Treemap-Visualisierungsmethode (vgl. Shneiderman, 1992), da diese eine kompakte Art und Weise für die Übertragung von Hierarchien ermöglicht. Der Zweck des Tools besteht darin, ein Mittel zur interaktiven Erforschung der verfügbaren Daten bereitzustellen, so dass der Benutzer Verständnis dafür gewinnt, wie das Wissen, das sich auf ein bestimmtes Wort (oder eine Zeichenkette) bezieht, in der Datenbank "gespeichert" ist, wobei die jeweiligen Lemmata mit der Benutzerabfrage zusammenhängen. Abgesehen von der Darstellung des resultierenden Sets von Lemmata, bauen wir eine Hierarchie je nach Kontext der Lemmata (in diesem Fall sind wir daran interessiert, die Lemmata in Bezug auf verschiedene Pflanzenarten zu gruppieren). Deshalb verwenden wir die beiden wichtigsten visuellen Merkmale von Treemaps: a) das Treemap-Layout (basierend auf einem Satz von verschachtelten Rechtecken, wobei jedes Rechteck einen Zweig der Hierarchie darstellt, der dann mit kleineren Rechtecken, welche Unterzweige darstellen, gekachelt wird) und b) die Fläche jedes Rechtecks (die proportional zur Größe der Daten ist).

Da in diesem Prototyp Pflanzennamen von größter Bedeutung sind, aber dem Benutzer, der eventuell mit den wissenschaftlichen Namen der Pflanzen nicht vertraut ist, wichtige Informationen verborgen bleiben könnten, entschieden wir uns für eine visuelle Art den Kontext (Pflanzen) der Lemmata, die in Zusammenhang mit

der Abfrage stehen, zu vermitteln: wir verwenden den Flickr-Webdienst, um Fotos abzurufen, die mit dem wissenschaftlichen Namen der Pflanze versehen sind (siehe Abbildung 1).

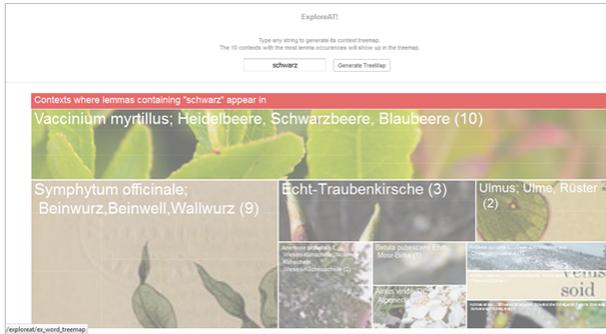


Abbildung 1: Beispiel für eine visuelle Darstellung von Pflanzennamen mit verschachtelten Rechtecken.

Als Ergebnis unseres visuellen Ansatzes kann der Benutzer die Verteilung der Lemmata in Abhängigkeit von den Pflanzen, auf die sie sich bezieht, verstehen (jedes Rechteck enthält das abgerufene Foto einer bestimmten Pflanze mit einem Bereich entsprechender Größe, die davon abhängt, wie viele Lemmata in Zusammenhang damit stehen). Der Benutzer kann dann auf das Rechteck seiner Wahl klicken, um tiefer zu gehen und alle relevanten Informationen für die mit dieser Pflanze zusammenhängenden Lemmata zu erhalten (siehe Abbildung 2).

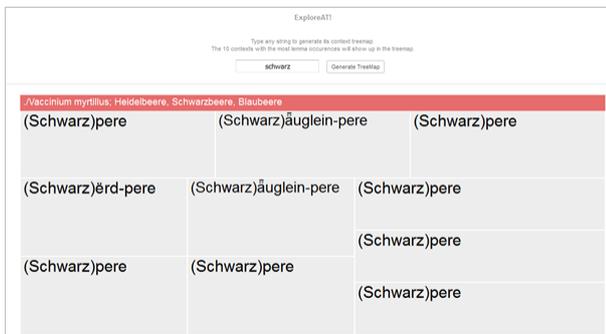


Abbildung 2: Beispiel für die Exploration von Pflanzennamen-Lemmata in einem bestimmten Rechteck, in diesem Fall Vaccinium myrtillus; Heidelbeere, Schwarzbeere, Blaubeere (siehe Abbildung 1).

Schließlich öffnen sich künftige Arbeitsfelder dank der Tatsache, dass dieser visuelle Ansatz auch noch gültig ist, wenn wir die Lemmata nach anderen Kriterien (d.h. nach einer mehrstufigen Hierarchie) gruppieren. Zum Beispiel könnte man zuerst die Lemmata nach Pflanze gruppieren; dann könnte man für eine bestimmte Pflanze die dazugehörigen Lemmata nach Zeit gruppieren, die wiederum nach Regionen gruppiert werden. Mit

diesen verschiedenen Arten der Gruppierung können diverse andere Daten mit einer ähnlichen strukturellen Beschaffenheit in derselben Weise visualisiert und analysiert werden. Dies würde unser Tool vielseitig und auch offen für andere Daten, nicht nur Pflanzennamen, machen.

Bibliographie

Verbert, Karen (2015): „On the Use of Visualization for the Digital Humanities“ in: *DH2015: Global Digital Humanities*.

Wandl-Vogt, Eveline / Kieslinger, Barbara / O'Connor, Alexander / Theron, Roberto (2015): „exploreAT! Perspektiven einer Transformation am Beispiel eines lexikographischen Jahrhundertprojekts“, in: *DHd 2015: Von Daten zu Erkenntnissen*.

Shneiderman, Ben (1992): „Tree visualization with treemaps: 2-d space-filling approach“, in: *ACM Transaction on Graphics (TOG)* 11 (1): 92-99.