

Digitales Lexikon des modernen Bayerischen

Thema:

Digitales Lexikon des modernen Bayerischen

Art:

MA

Betreuer:

Manuel Burghardt

Student:

Daniel Granvogel

Professor:

Christian Wolff

Status:

in Bearbeitung

Stichworte:

Digital Humanities, Lexikographie, Dialekt

angelegt:

2016-09-16

Beginn:

2016-08-01

Antrittsvortrag:

2016-10-24

Abschlussvortrag:

2017-05-08

Hintergrund

Seit 1913 arbeitet die bayerische Akademie der Wissenschaften an einem Wörterbuch des Bayerischen mit dem Ziel den Wortschatz der bairischen Dialekte in Bayern in ihrer ganzen Vielfalt zu dokumentieren. Als Grundlage für die Belege dienten hierbei unter anderem Fragebögen, Kundfahrten, Auswertung von verschiedenster Fachliteratur und die Analyse von älteren Quellen, teilweise zurückgehend bis ins Jahr 800. Aus diesen Daten konnte eine Wörterbuchkartei mit etwas 4 Millionen Einzelbelegen zu etwa 30.000 Stichwörtern erstellt werden. Seit circa 1958 ist die Materialsammlung abgeschlossen. Einen anderen Ansatz wie man Dialekt-Daten erfassen und aufbereiten kann zeigte die Studie von Burghardt, M. and Granvogel, D. and Wolff, C. „Creating a Lexicon of Bavarian Dialect by Means of Facebook Language Data and Crowdsourcing“. Hier wurde auf Basis von Facebook-Textdaten und einer Crowdsourcing-Komponente ein kleines Lexikon des Bayerischen erstellt und so nachgewiesen, dass man die Daten aus Social-Media sehr gut für sprachwissenschaftliche Fragestellungen verwenden kann.

Zielsetzung der Arbeit

Die Masterarbeit setzt sich in mehreren Punkten von der Bachelorarbeit ab und erweitert diese. So wird zum einen der Verarbeitungsprozess der Rohdaten aus Facebook verbessert und so gestaltet, dass er von Sprachwissenschaftlern und Dialektologen auf andere Rohdatensätze angewendet werden kann. Auch wird die Crowdsourcing-Komponente von Grund auf neu gestaltet werden. Zum

einen werden die in der vorangegangenen Studie angemerkten Verbesserungsvorschläge eingearbeitet und zum anderen wird die Komponente mit dem Mobile-First-Ansatz gestaltet, so dass es eine gut funktionierende mobile Version der Crowdsourcing-Komponente gibt. Dadurch soll die Teilnahme an der Übersetzung noch gesteigert werden.

Auch soll der Aspekt des Gamifications viel stärker in die Anwendung integriert werden. Ein weiterer Punkt, der die Masterarbeit von der Bachelorarbeit unterscheidet wird die Anzahl der herangezogenen Facebook-Gruppen sein, in der Fallstudie wurden nur Rohdaten aus einer einzigen Gruppe extrahiert, in dieser Arbeit sollen mehrere Gruppen aus verschiedenen Regionen Bayerns herangezogen werden.

Konkrete Aufgaben

1. „Toolkit-Box“ für die Beschaffung und Aufbereitung der Rohdaten erstellen
2. Korpus erstellen
3. Daten für Crowdsourcing-Komponente daraus extrahieren
4. WebApp erstellen (Crowdsourcing-Komponente)
5. Ein „public“ und ein „academic“ Korpus bereitstellen, das public-Korpus auf der Webseite für die Öffentlichkeit verfügbar machen
6. Übersetzung durchführen
7. Ergebnisse auswerten / analysieren

Erwartete Vorkenntnisse

HTML/CSS/JavaScript/Python/Angular

Weiterführende Quellen

Burghardt, M., Granvogl, D., & Wolff, C. (2016). Creating a Lexicon of Bavarian Dialect by Means of Facebook Language Data and Crowdsourcing. In Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC).

From:
<https://wiki.mi.ur.de/> - MI Wiki

Permanent link:
https://wiki.mi.ur.de/arbeiten/digitales_lexikon_des_modernen_bayerischen?rev=1490289547

Last update: **23.03.2017 17:19**

