



Bachelor-/ Masterarbeit in Bioinformatik: Information Retrieval of GPS-dependent Data

Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de

Zusammenfassung

Sozio-ökonomische, geographische und ökologische Prozesse – wie z.B. Stadtentwicklung, Krankheitsverbreitung und die Zusammensetzung von Ökosystemen - sind oft von den Eigenschaften ihrer Umgebung beeinflusst [1]. Jedoch sind diese Zusammenhänge oft schwer aufzuklären, da deren Komplexität recht hoch ist und die dazu notwendigen Daten Forschenden z.T. nicht zur Verfügung stehen. Diese Informationen liegen zwar in öffentlichen Datenbanken und sind somit frei nutzbar, jedoch finden sie oft nicht den Weg in die Analysen von umweltbeeinflussten Prozessen. Ein zentrales Problem hierbei ist, dass die Daten nicht aggregiert, sondern auf viele unterschiedliche Datenbanken verteilt vorliegen.

Um dieses Problem zu lösen, haben wir das Tool SEDE-GPS entwickelt. Es sammelt aus verschiedenen öffentlichen Datenbanken gezielt Informationen zu vom User bereitgestellten GPS-Punkten und macht diese dann dem User zugänglich. In der momentanen Fassung aggregiert SEDE-GPS aus vier verschiedenen Datenquellen rund 18.000 features, die sozio-ökonomische, klimatische und geographische Eigenschaften abdecken. SEDE-GPS hat jedoch durch die Wahl der Quellen einige Limitationen, sammelt für Orte außerhalb von Europa z.B. nur eine viel kleinere Datenmenge als für Orte innerhalb Europas.

In dieser Bachelor- oder Masterarbeit soll das in Java programmierte Programm SEDE-GPS erweitert werden, sodass (1) weitere geographische Gebiete abgedeckt sind und (2) Datenbanken mit neuen Datentypen eingebunden werden (wie z.B. Daten zu Bodenbeschaffenheit, ökologischen Zuständen, Satellitenbilder). Die durch die neue Version von SEDE-GPS erzeugten Datensätze sollen dann ausgiebig charakterisiert werden. Wird dieses Projekt im Rahmen einer Masterarbeit bearbeitet, so soll anschließend die neue Version von SEDE-GPS genutzt werden, um wichtige Faktoren für einen natürlichen Prozess wie z.B. die Qualität von Ökosystemen zu identifizieren.

[1] Rossum, T.V., et al, Year-long metagenomic study of river microbiomes across land use and water quality. *Frontiers in Microbiology* 2015, 6.

[2] Sperlea T, Fuser S, Boenigk J, Heider D: SEDE-GPS: Socio-Economic Data Enrichment based on GPS information. Preprint 2018

Bitte melden Sie sich bei Interesse per Email bei:

Prof. Dr. Dominik Heider - dominik.heider@uni-marburg.de

Stand:

17.04.18