



Michelle Kunz



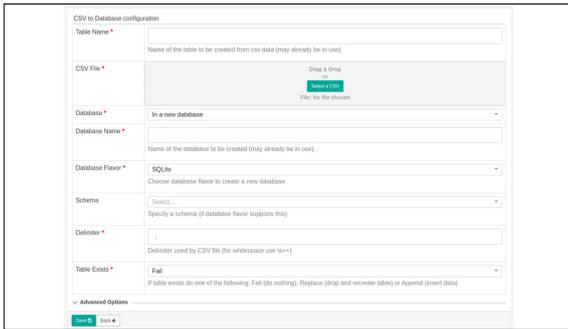
Renato Fabio Venzin



Sascha Gschwind

Diplomanden	Michelle Kunz, Renato Fabio Venzin, Sascha Gschwind
Examinator	Prof. Stefan F. Keller
Experte	Claude Eisenhut, Eisenhut Informatik AG, Burgdorf, BE
Themengebiet	Software

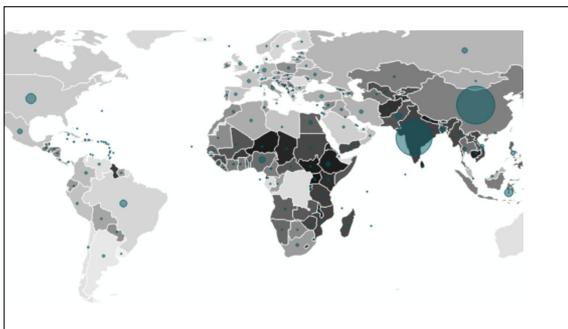
Apache Superset-Erweiterung mit CSV-Importer und Geokodierung



Neu designte Eingabemaske in Superset für den Import eigener CSV-Dateien (CSV-Importer)
Eigene Darstellung



Zusätzlich implementierte Eingabemaske zur Geokodierung, d.h. zur Transformation von Adressen zu Koordinaten
Eigene Darstellung



Einwohnerzahlen aller Länder als Ortsdiagrammkarte (Quelle: World Bank Data)
Eigene Darstellung

Ausgangslage: Damit Unternehmen wettbewerbsfähig bleiben, werden Daten gesammelt und analysiert. Anschliessend werden die Resultate inklusive Visualisierungen publiziert. Dafür werden sogenannte Business-Intelligence-Tools (BI-Tools) eingesetzt. Visualisierungen können dazu verwendet werden, Zusammenhänge aufzudecken und zu verstehen. Zusätzlich dienen Visualisierungen als Basis für Entscheidungen. Apache Superset ist ein solches BI-Tool. Es ist eine der wenigen Applikationen dieser Kategorie, die Open Source ist. Das macht Superset zur kostengünstigen Alternative, vermindert vendor-lock-in und macht es erweiterbar.

Ziel der Arbeit: Diese Arbeit erweitert die Software Apache Superset mit zwei Hauptzielen: Das erste Ziel ist, das Hochladen (upload) eigener Daten benutzerfreundlicher und moderner zu gestalten (sogenannter "CSV-Importer"). Das zweite Ziel ist die Erweiterung von Superset mit einer Geokodierungs-Funktion. Unter Geokodierung versteht man den Vorgang, aus Adressen Koordinaten zu gewinnen. Bisher mussten die Benutzer ihre Daten vorgängig über ein separates Tool selbst aufbereiten. Mit Hilfe dieser Funktion können Adressen direkt in Superset in Koordinaten transformiert werden. Damit können Benutzer nun noch mehr aus den Business-Daten holen, in dem sie zusätzlich Kartengrafiken erstellen und publizieren können.

Des Weiteren soll im Kontext dieser Arbeit Know-How über diese komplexe Software gesammelt und dokumentiert werden. Dazu gehören u. a. der Projektaufbau von Superset, die Konfiguration und die internen Strukturen. Das Know-How soll dem HSR-Institut für Software auch dazu dienen, eine eigene Instanz von Superset betreiben und erweitern zu können. Zusätzlich dienen Teile der Dokumentation dem Institut als Unterrichtsunterlagen für Weiterbildungskurse, um interessierten Personen Superset näher zu bringen.

Ergebnis: Die oben erwähnten Ziele wurden erreicht. Sowohl der CSV-Importer als auch die Geokodierung wurden mit dem JavaScript-Framework React modernisiert. Die damit durchgeführte Pionierarbeit dient den Maintainern von Superset als Vorlage und wird ihnen helfen, weitere User-Interfaces (inkl. Eingabemasken) auf die neuesten technologischen Standards anzuheben.

Der neu designte und erweiterte CSV-Importer wirkt übersichtlicher und verständlicher. Nebst dem Hochladen der Daten in eine vordefinierte Datenbank, kann nun zwischen einer neuen SQLite- oder PostgreSQL-Datenbank ausgewählt werden. Diese neue Datenbank wird im Hintergrund mit den gegebenen Softwarebibliotheken (SQLAlchemy, Pandas) konfiguriert und erstellt. Für die Geokodierung wurde eine eigene Eingabemaske implementiert, die im Backend auf externe Geokodierungs-Webservices von MapTiler und Google zugreift. Der Benutzer kann ein neues Koordinaten-Attribut anfügen oder ein bestehendes Attribut überschreiben lassen. Es ist auch möglich, dass nur diejenigen Adressen geokodiert werden, zu denen noch keine Koordinaten gefunden wurden. Der Vorgang der Geokodierung kann auch abgebrochen werden. Die erstellten Anleitungen entsprechen der aktuellen Version und dem Setup von Superset und verbessern dessen bisher vernachlässigte Dokumentation.