

# Data Enrichment



Infomatik



## PROJEKTIDEE

In vielen Firmen, die B2B-Dienstleistungen anbieten, gibt es einen hohen Verwaltungsaufwand, wenn es um Kundendaten geht. Dieses Tool bietet eine Möglichkeit, diesen mühsamen Prozess, der Kundendatenermittlung und -aktualisierung vollständig zu automatisieren. Durch ein KI-gestütztes System werden die Kundendaten von der Firmenwebsite ausgelesen und in die Firmendatenbank geschrieben.



## FUNKTIONSWEISE

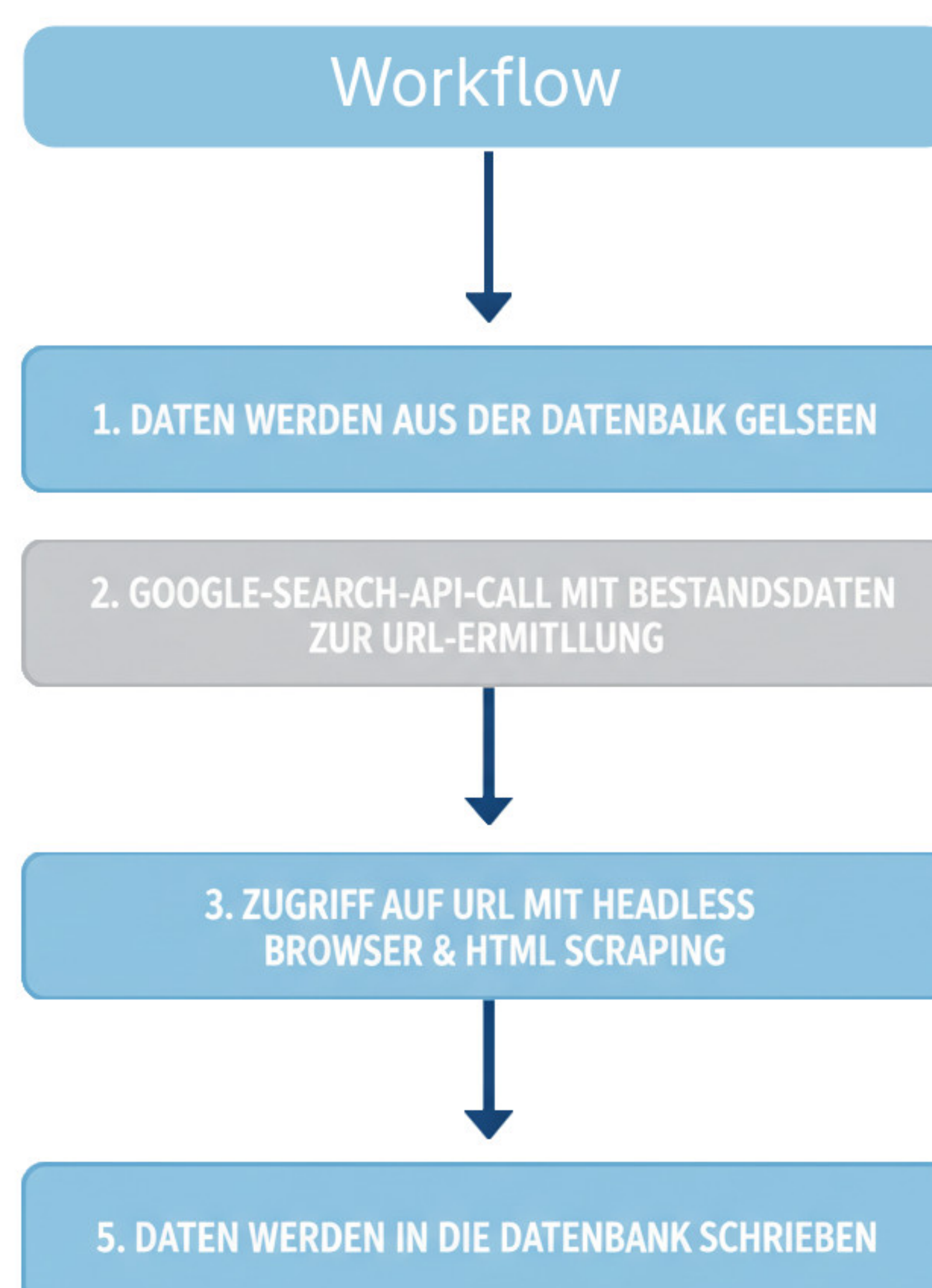
Ablauf für einen Datensatz:

- Daten werden aus der Datenbank gelesen.
- Mithilfe der Bestandsdaten wird ein Google-Search-API-Call ausgeführt, um die URL der Firmenwebsite zu ermitteln.
- Auf die URL wird mithilfe eines Headless Browsers zugegriffen, um das HTML der Website zu scrapen.
- Durch einen API-Call an Gemini werden die Daten aus dem HTML extrahiert.
- Die Daten werden in die Datenbank geschrieben.



## DATEN UND FAKTEN

Programmiersprache: Python  
Bibliotheken: u.a. pydantic, pytest, google.genai & playwright  
Code-Zeilen: ca. 3.500 Zeilen  
KI-Model: Gemini  
Headless Browser: Playwright  
Tests: Unit-Tests



Projektteam:  
Nick Luggin



max valier  
TFO BOZEN